

---

## APPENDIX A

# Mathematical models of transmission

---

In this appendix I will present details of mathematical models of cultural and genetic transmission. In Section A.1 I outline B&R’s treatment of cultural transmission and the factors influencing cultural evolution. In Section A.2 I outline a simple mathematical model of genetic transmission and biological evolution by natural selection. Finally, in Section A.3 I describe the mathematical details of B&R’s model of the dual transmission of cultural traits and a genetically-encoded direct bias.

### A.1 Models of cultural transmission

This Section covers B&R’s basic models of unbiased cultural transmission (Section A.1.1), and their treatment of the various pressures acting on cultural transmission (Section A.1.2).

#### A.1.1 Basic cultural transmission models

##### A.1.1.1 Transmission of dichotomous traits

B&R provide a simple model of the cultural transmission of a dichotomous trait, where individuals are either characterised as having cultural trait  $c$  or  $d$ .  $p$  is the proportion of individuals in the population with cultural variant  $c$ , and  $1 - p$  is the proportion of individuals with variant  $d$ .  $p'$  is the proportion of individuals in the population with cultural variant  $c$  after cultural transmission.

Each individual acquires their cultural variant based on their observations of the cultural variants of  $n$  cultural parents, or models. The probability that a naive individual acquires variant  $c$  based on the behaviour of  $n$  models is therefore:

$$Prob(individual = c | X_1, \dots, X_n) = \sum_{i=1}^n A_i X_i$$

where  $X_i = 1$  if the  $i$ th model possesses variant  $c$  and  $X_i = 0$  if the  $i$ th model possesses variant  $d$ ,  $A_i$  is the probability that the naive individual acquires the variant of the  $i$ th model and  $\sum_i A_i = 1$ .  $A_i$  therefore gives the importance of the  $i$ th cultural parent in the enculturation process.

Given this equation we can now calculate  $p'$ , the proportion of individuals with variant  $c$  after cultural transmission. For this we require the probability that a given set of models  $(X_1, \dots, X_n)$  is formed,  $Prob(X_1, \dots, X_n)$ . This leads to:

$$p' = \sum_{x_1=0}^1 \dots \sum_{x_n=0}^1 Prob(c|x_1, \dots, x_n) Prob(X_1 = x_1, \dots, X_n = x_n)$$

In other words, the proportion of individuals with variant  $c$  is equal to the probability that an individual will acquire variant  $c$  based upon exposure to a specific set of  $n$  models, multiplied by the probability of the formation of that set of models and summed over all sets of models.  $p'$  therefore depends on the probability of forming sets of models. If we assume that the probability of any cultural parent possessing variant  $c$  is equal to  $p$  (i.e. cultural parents are drawn at random from the population) then it can be shown that cultural transmission leaves the frequency of cultural variants in the population unchanged i.e.

$$p' = p$$

Therefore, if the original population exhibits variation for some cultural trait, cultural transmission itself will not reduce that variation or alter the distribution of variants, assuming random selection of cultural parents — cultural transmission alone will not lead to cultural evolution or cultural adaptation.

#### A.1.1.2 Transmission of continuous traits

In the continuous trait model each individual is characterised by a single number,  $X$ , representing the value of their culturally-acquired character. In this case a population cannot be characterised by the proportion of one cultural variant, as in the dichotomous character model, but must rather be modelled as a distribution over values of  $X$ ,  $P(X)$ .

Making the simplifying assumption that  $P(X)$  can be approximated by a normal distribution allows a population to be characterised by the mean value of  $X$  in the population,  $\bar{X}$  and the variance of  $X$  in the population,  $V$ .

B&R consider how a blending inheritance model would alter the mean and the variance of a cultural characteristic in a population. As in the dichotomous model, each naive individual is exposed to the behaviour of  $n$  models, with the cultural variant of the  $i$ th model being  $X_i$ . Based on these observations, the naive individual makes an estimate of the  $i$ th model's cultural rule,  $Z_i$ , where:

$$Z_i = X_i + e_i$$

where  $e_i$  is a random variable drawn from the normal distribution with mean 0 and variance  $E_i$ , representing errors in the naive individual's estimate of the model's cultural character. As the name suggests, in a blending inheritance model the enculturated individual's cultural variant,  $X_0$ , is simply the average of their estimates of their  $n$  models' variants:

$$X_0 = \sum_{i=1}^n A_i Z_i$$

where, as in the dichotomous model,  $A_i$  is the importance of the  $i$ th model.

By a similar method to that used for the dichotomous case, it is possible to calculate the mean value of  $X$  in the population after cultural transmission,  $\bar{X}'$  and variance of  $X$  in the population after transmission,  $V'$ . B&R show that, assuming non-selective formation of sets of models:

$$\bar{X}' = \bar{X}$$

In other words, blending inheritance does not change the population mean of the cultural variant, as with the dichotomous model. However, the variance of the population does not necessarily remain unchanged. Assuming non-selective formation of sets of models, equal weighting for all models ( $A_i = 1/n$  for all  $i$ ) and no correlation between errors in a given set of models ( $Cov(e_i, e_j) = 0$  for all  $i$  and  $j$ ):

$$V' = (1/n)(V + \bar{E})$$

where  $\bar{E}$  is the average value of  $E_i$  (recall that  $E_i$  gives the variance of a normal distribution, and the errors made by the learner when estimating the trait of the  $i$ th cultural parent come from this distribution) for the set of  $n$  models. In other words, there are two forces acting on the population. Assuming no errors in transmission ( $\bar{E} = 0$ ), blending transmission tends to reduce the variance in the population, with variance being reduced faster for larger numbers of cultural parents. The counteracting force, dependent on the average error introduced during transmission ( $\bar{E}$ ), tends to increase the variance in the population, with errors of larger variance increasing the variance in the population more.

### A.1.2 Pressures acting on cultural transmission

B&R provide mathematical accounts of how three pressure acting on transmission can result in cultural change and cultural evolution. These are:

1. Natural selection of cultural variants, resulting from selective removal of enculturated individuals.
2. Guided variation, resulting from individual learning by enculturated individuals.
3. Biased transmission, resulting from the strategy of learners during cultural transmission. The forces of biased transmission can be further subdivided into three forms:
  - (a) Directly biased transmission, resulting from a preference for learners to acquire one cultural variant over another.
  - (b) Indirectly biased transmission, resulting from a preference for learners to acquire cultural traits which are associated with other cultural traits.
  - (c) Frequency-dependent transmission, resulting from a disproportionate preference for learners to acquire the most (or least) frequent cultural trait in the population.

In Sections A.1.2.1 to A.1.2.5 B&R's models for these pressures are reviewed. In the interests of clarity, a separate section is devoted to each of the three subtypes of biased transmission.

#### A.1.2.1 Natural selection of cultural variants

B&R model the natural selection of cultural variants by assuming that there are a set of  $n$  distinct social roles (e.g. mother, father, uncle, priest, teacher). Each naive individual acquires their cultural characteristic based on observation of a subset of these roles  $\tau_j$ .

As before, the weight of social role  $k$  is  $A_k$ . The weight of social role  $k$  with respect to a subset of social roles  $\tau_j$ ,  $A_{kj}$  is:

$$A_{kj} = \begin{cases} A_k & \text{if } k \text{ belongs to } \tau_j \\ 0 & \text{otherwise} \end{cases}$$

Working within the dichotomous traits model, the probability that a naive individual acquires variant  $c$  based on the behaviour of the set of individuals with phenotypic values  $X_1, \dots, X_n$  and the set of cultural parents with roles  $\tau_j$  is therefore:

$$\text{Prob}(\text{individual} = c | \tau_j, X_1, \dots, X_n) = \frac{\sum_{k=1}^{k=n} A_{kj} X_k}{\sum_{k=1}^{k=n} A_{kj}}$$

This equation normalises the weight of the cultural parent with role  $k$  by the weights of all roles present in the set of roles  $\tau_j$  and is clearly related to the earlier equation for the cultural transmission of dichotomous traits.

In order to model natural selection we must assume that the probability that an individual attains a particular social role  $k$  depends on the cultural variant that that individual possesses. Let  $\Omega_{ck}$  be the probability that an individual with cultural variant  $c$  attains social role  $k$  and, similarly,  $\Omega_{dk}$  be the probability that an individual with cultural variant  $d$  attains social role  $k$ . As before, we will assume that the frequency of variant  $c$  in the population is  $p$ . The frequency of individuals with variant  $c$  attaining social role  $k$ ,  $p'_k$  is therefore:

$$p'_k = \frac{\Omega_{ck} p}{\Omega_{ck} p + \Omega_{dk} (1 - p)}$$

Working under the assumption that sets of roles  $\tau_j$  are formed at random, and following a similar procedure to that outlined for the basic cultural transmission rule, B&R show that the frequency of cultural variant  $c$  in the population after differential attainment of social roles and cultural transmission by the linear rule,  $p''$ , is:

$$p'' = \sum_{k=1}^n \overline{A}_k p'_k$$

where  $\bar{A}_k$  gives the importance of parents in the  $k$ th social role averaged over all possible sets of cultural parents  $\tau_j$  according to the frequency with which those sets occur:

$$\bar{A}_k = \sum_j Prob(\tau_j) \left( \frac{A_{kj}}{\sum_l A_{lj}} \right)$$

This equation can be combine with the equation for  $p'_k$  given above. If  $\sigma_k$  is the selective advantage of variant  $c$  with respect to role  $k$  ( $\sigma_k = (\Omega_{ck}/\Omega_{dk}) - 1$ ) and assuming that selection is weak, the equation becomes:

$$p'' = p + p(1-p) \left( \sum_{k=1}^n \bar{A}_k \sigma_k \right)$$

where the sum is the selection advantage of variant  $c$  in role  $k$  averaged over all social roles and weighted by the importance of each role ( $\bar{A}_k$ ). Roles which offer a high selective advantage ( $\sigma_k$ ) will have a strong influence, even if that social role is not weighted particularly highly in contribution to cultural transmission (i.e.  $\bar{A}_k$  is not particularly high relative to  $\bar{A}_{l \neq k}$ ). Variant  $c$  will increase if this quantity is positive and decrease if it is negative — if variant  $c$  offers a selective advantage when averaged over social roles then it will increase in frequency in the population. In other words, if possessing variant  $c$  makes an individual more likely to occupy a role which allows them to enculturate others and transmit that variant, then  $c$  will increase in frequency in the population.

#### A.1.2.2 Guided variation

A model of guided variation requires a model of individual learning. B&R assume that an individual can be characterised by a number  $X$ , the initial value of their phenotype prior to individual learning, and a number  $Y$ , the value of their phenotype after individual learning. This is therefore a continuous trait model of cultural characteristics. The goal of learning is determined by the environment, which is characterised by a number  $H$ . The aim of learning is essentially to move  $Y$  towards  $H$ .  $L$  is a parameter determining the reliance of an individual on individual learning, with high  $L$  indicating a high reliance on individual learning. Errors made during the learning process are represented by a normally distributed random variable  $\epsilon$  with mean 0 and variance  $V_e$ . It can be shown that:

$$Y = aX + (1-a)(H + \epsilon)$$

where  $a = V_e/(V_e + L)$  is a parameter that gives the importance of individual learning —  $a \approx 1$  ( $L \ll V_e$ ) corresponding to a tendency to rely on the initial value of the phenotype  $X$  and  $a \approx 0$  ( $L \gg V_e$ ) corresponding to a tendency to ignore the initial value of the phenotype and move towards the value preferred by the environment,  $H$ .

How does this type of individual learning change the mean value and the variance of a population's cultural characteristic? Prior to individual learning the mean value of the trait in the population is given by  $\bar{X}$  and the variance is given by  $V$ . The mean value after individual learning,  $\bar{Y}$  is:

$$\bar{Y} = a\bar{X} + (1 - a)H$$

where  $a$  is as before. As in the individual case, in the population case the mean value for the trait will tend towards the value favoured by the environment if  $a < 0.5$ . The variance of the population after individual learning,  $U$ , is:

$$U = a^2V + (1 - a)^2V_e$$

Individual learning both decreases the variance of the trait in the population through movement towards the environmentally-determined goal ( $U = a^2V \dots$ ) and increases it due to errors introduced by individual learning ( $U = \dots (1 - a)^2V_e$ ).

If we assume that the culturally-acquired value for the phenotype forms the initial value of the phenotype which can subsequently be altered by individual learning, this model of individual learning can be simply added to the blending model of cultural transmission. Assuming non-selective formation of sets of models, equal weighting for all models ( $A_i = A_j = 1/n$ ) and no correlation between errors in a given set of models ( $Cov(e_i, e_j) = 0$ ) the mean value of  $X$  in the next generation,  $\bar{X}'$ , is:

$$\bar{X}' = a\bar{X} + (1 - a)H$$

i.e. when individual learning is powerful ( $a \approx 0$ ) the population moves towards the value of the phenotype favoured by the environment, due to the transmission of cultural traits favoured by individual learning, and when individual learning is weak ( $a \approx 1$ ) the mean

value of the population's cultural trait remains unchanged by individual learning. The variance after transmission,  $V'$ , is:

$$V' = (1/n)(a^2 V + (1 - a)^2 V_e + \bar{E})$$

i.e. blending both reduces  $((1/n))$  and increases  $(\dots + \bar{E})$  variance, and individual learning both reduces  $(a^2 V)$  and increases  $((1 - a)^2 V_e)$  variance.

#### A.1.2.3 Directly-biased transmission

Direct bias can be simply modelled using the model of the transmission of dichotomous characters given earlier. As before, the probability that an individual acquires cultural variant  $c$  given the set of cultural parents  $X_1, \dots, X_n$  is:

$$\text{Prob}(c|X_1, \dots, X_n) = \sum_{i=1}^n A_i X_i$$

In the unbiased case, the value of a particular  $A_i$  is independent of  $X_i$  — the cultural variant used by a model does not affect the importance of that model to the naive individual. However, in the biased case,  $A_i$  depends on the intrinsic importance of the  $i$ th model, given by  $\alpha_i$ , and the biasing function,  $\beta(X_i)$ :

$$A_i = \frac{\alpha_i(1 + \beta(X_i))}{\sum_{j=1}^n \alpha_j(1 + \beta(X_j))}$$

where the biasing function is:

$$\beta(X_i) = \begin{cases} B & \text{if } X_i = 1 \\ -B & \text{if } X_i = 0 \end{cases}$$

$B$  gives the strength of the bias in favour of cultural variant  $c$ . Assuming  $B > 0$  (variant  $c$  is favoured over variant  $d$ ), if the  $i$ th model has cultural variant  $c$  then the intrinsic weight of that model will be increased by a factor  $1 + B$ , whereas if the  $i$ th model has variant  $d$  then the intrinsic weight of that model will be decreased by  $1 - B$ . Note that this model of a biasing function is in principle arbitrary with respect to the functionality of the cultural trait, with an arbitrary preference in favour of one variant over the other determined by  $B$ .

However,  $B$  could be linked to the expected fitness payoff of the variants, in which case the bias would be non-arbitrary and in favour of the cultural variant which is expected to yield the greatest fitness payoff.

B&R consider the case where each individual is exposed to two models, with intrinsic weights  $\alpha_1$  and  $\alpha_2$ . This can be interpreted as either the case where each individual has two cultural parents, or the case where each individual has multiple cultural parents but is encultured in a serial fashion, observing  $X_2$  for each parent in turn, comparing it to their own value, ( $X_1$ ) and deciding on which of the two possibilities ( $X_1$  or  $X_2$ ) to adopt. If  $p$  is the frequency of variant  $c$  in the population prior to such an episode of cultural transmission, its frequency after cultural transmission,  $p'$ , will be:

$$p' = p + p(1 - p) \left( \frac{4B\alpha_1\alpha_2}{1 - B^2(\alpha_1 - \alpha_2)^2} \right)$$

Assuming that both cultural parents have equal weight ( $\alpha_1 = \alpha_2 = 0.5$ ), this reduces to:

$$p' = p + p(1 - p)B$$

In other words, directly biased transmission will increase the frequency of the favoured variant in the population. The rate of increase depends on the strength of the bias ( $B$ ) and the variance in the population ( $p(1 - p)$ ).

#### A.1.2.4 Indirectly-biased transmission

A model of indirect bias requires a model of the transmission of multiple cultural traits. B&R develop a model of the blending transmission of two quantitative cultural traits which is based on the basic transmission model for single continuous traits (outlined in Section A.1.1.2). The  $j$ th individual is characterised by a two-place vector  $X_j = (X_{1j}, X_{2j})$ . As before, a naive individual observing individual  $j$  forms an estimate of that individual's cultural variants,  $Z_j = (Z_{1j}, Z_{2j})$  such that:

$$Z_{1j} = X_{1j} + e_{1j}$$

$$Z_{2j} = X_{2j} + e_{2j}$$

As before,  $e_{1j}$  and  $e_{2j}$  are random variables drawn from normal distributions with mean 0, variances  $E_{1j}$  and  $E_{2j}$  respectively and covariance  $E_{12j}$ .

As for the earlier definition of the blending rule, naive individuals observe and estimate the cultural character of  $n$  models and then form their own cultural character,  $X_0 = (X_{10}, X_{20})$ , by averaging over observed models:

$$X_{10} = \sum_{j=1}^n A_{1j} Z_{1j}$$

$$X_{20} = \sum_{j=1}^n A_{2j} Z_{2j}$$

where  $A_{ij}$  is the importance of the  $j$ th model in transmitting cultural characteristic  $i$  ( $i = 1$  or 2). As with the single-trait blending model, we need to know how this type of transmission will affect the mean value of trait  $i$  in the population,  $\bar{X}_i$ , and the variance of trait  $i$ ,  $V_i$ . By similar methods, it can be shown that the mean and variance after transmission ( $\bar{X}'_i$  and  $V'_i$ ) are given by:

$$\bar{X}'_i = \bar{X}_i$$

$$V'_i = \sum_{j=1}^n A_{ij}^2 (V_i + \bar{E}_i)$$

where  $\bar{E}_i$  is the weighted average of errors introduced during transmission:

$$\bar{E}_i = \frac{\sum_{j=1}^n A_{ij}^2 E_{ij}}{\sum_{j=1}^n A_{ij}^2}$$

If we assume that all cultural parents for a given trait have equal weight (i.e.  $A_{ij} = A_{ik} = 1/n$ ) then this reduces to:

$$V'_i = (1/n) (V_i + \bar{E}_i)$$

i.e. as before, blending inheritance leaves the mean in the population unchanged and both decreases and increases variance, depending on the number of cultural parents and variance of the errors introduced.

Cultural transmission will also affect the covariance between the values of traits 1 and 2,  $C_{12}$ . For the simplified case where each cultural parent has equal weight ( $A_{ij} = A_{ik} = 1/n$ , which implies  $A_{1j} = A_{2j}$  i.e. the models are equally important in transmitting both traits, rather than some models being important in the transmission of one trait and other models being important in the transmission of other traits), the covariance of the values after transmission,  $C'_{12}$ , is given by:

$$C'_{12} = (1/n) \left( C_{12} + \frac{\sum_{j=1}^n E_{12j}}{n} \right)$$

i.e. as with variance, co-variance is reduced by the factor  $(1/n)$  and increased by correlated errors, measured by the degree of correlation between errors averaged over all models  $(\sum_{j=1}^n E_{12j}/n)$ . For more complex cases where the different traits are influenced by different sets of models (i.e.  $A_{1j} \neq A_{2j}$ ) the covariance between the traits tends to decrease.

Given this blending model of the transmission of two quantitative characters, it is possible to model indirect bias. We will consider trait 1 to be the indicator trait and trait 2 to be the indirectly biased trait, so that an individual can be characterised by a two-place vector  $X = (X_I, X_D)$  ( $I$  for Indicator trait,  $D$  for derived trait). As described above, individuals acquire their trait based on the weighted average of their estimate of the variants of their cultural parents i.e.  $X_{i0} = \sum_{j=1}^n A_{ij} Z_{ij}$  where  $i = I$  or  $D$ .

As discussed above, the indicator trait is a directly-biased trait — some values for the indicator trait are intrinsically preferred.  $A_{Ij}$  is therefore a function of the intrinsic influence of parent  $j$  with respect to trait  $I$ ,  $\alpha_{Ij}$ , and the estimated value of model  $j$ 's trait  $I$ ,  $Z_{Ij}$ :

$$A_{Ij} = \frac{\alpha_{Ij} (1 + \beta(Z_{Ij}))}{\sum_{k=1}^n \alpha_{Ik} (1 + \beta(Z_{Ik}))}$$

where  $\beta(x)$  is a direct bias function. This equation should be familiar due to its similarity to the equation from the direct bias section.

The importance of the  $j$ th cultural parent with respect to the indirectly biased trait,  $A_{Dj}$ , will be a function of that parent's intrinsic importance,  $\alpha_{Dj}$ , and the estimate of the  $j$ th model's value for the *indicator* trait,  $Z_{Ij}$  (rather than the estimate of the  $j$ th model's value for the indirectly biased trait  $Z_{Dj}$ ):

$$A_{Dj} = \frac{\alpha_{Dj} (1 + \theta(Z_{Ij}))}{\sum_{k=1}^n \alpha_{Dk} (1 + \theta(Z_{Ik}))}$$

where  $\theta(x)$  is the indirect bias function, of a similar form to the direct bias function.

Assuming the non-selective formation of sets of models, weak biasing functions and equal intrinsic weightings to all cultural parents ( $\alpha_{ij} = \alpha_{ik} = 1/n$ ), the mean values of the traits after transmission,  $\bar{X}'_i$  can be calculated given the mean values of the traits prior to transmission,  $\bar{X}_i$ :

$$\bar{X}'_I = \bar{X}_I + (1/n) Cov(Z_I, \beta(Z_I))$$

$$\bar{X}'_D = \bar{X}_D + (1/n) Cov(Z_D, \theta(Z_I))$$

$Cov(Z_i, f(Z_j))$  is the covariance of the trait  $Z_i$  and the bias function  $f$  applied to some trait  $Z_j$ . If increases in  $Z_i$  tend to result in increases in  $f(Z_j)$  then  $Cov(Z_i, f(Z_j)) > 0$ . On the other hand, if increases in  $Z_i$  tend to result in *decreases* in  $f(Z_j)$  then  $Cov(Z_i, f(Z_j)) < 0$ .  $Cov(Z_I, \beta(Z_I))$  therefore gives the strength and direction of the direct bias — if  $Cov(Z_I, \beta(Z_I)) < 0$  then the mean value of  $X_I$  must be above the value favoured by the direct bias and the mean value will decrease through transmission by an amount proportional to the magnitude of  $Cov(Z_I, \beta(Z_I))$ . Similarly, if  $Cov(Z_I, \beta(Z_I)) > 0$  then the mean value of  $X_I$  will increase.

$Cov(Z_D, \theta(Z_I))$  gives the strength and direction of the indirect bias, and depends on whether values of  $Z_D$  and  $Z_I$  are correlated. Consider the case where  $Z_D$  and  $Z_I$  are positively correlated. Higher values of  $Z_D$  will be associated with higher values of  $Z_I$ . If the current value of  $Z_I$  associated with the current value of  $Z_D$  is below the optimum value given by  $\theta(Z_I)$  then increases in  $Z_D$  will result in increases in  $\theta(Z_I)$  and therefore  $Cov(Z_D, \theta(Z_I)) > 0$ . Similarly, if the current value of  $Z_I$  associated with the current value of  $Z_D$  is above the optimal value then increases in  $Z_D$  will result in decreases in  $\theta(Z_I)$  and therefore  $Cov(Z_D, \theta(Z_I)) < 0$ . In either case, the mean of the population's value for  $X_D$  will tend towards the value associated with the value of  $X_I$  which maximises the indirect bias function  $\theta$  — “variants of the indirectly biased trait that are positively correlated with the admired variants of the indicator trait will increase in frequency” (B&R p254). Similarly, variants of the indirectly biased trait which are

negatively correlated with the admired variants of the indicator trait will decrease in frequency.

#### A.1.2.5 Frequency-dependent bias

In Section A.1.1.1 a model was described which gave the probability of acquiring cultural variant  $c$  on the basis of  $n$  models for the unbiased dichotomous case. The frequency-dependent bias case is very similar:

$$Prob(individual = c | X_1, \dots, X_n) = \sum_{i=1}^n A_i X_i + D \left( \sum_{i=1}^n A_i X_i \right)$$

Assuming that each model has equal importance this becomes

$$Prob(individual = c | j) = j/n + D(j)$$

where  $j$  is the number of parents with cultural variant  $c$ :

$$j = \sum_{i=1}^n X_i$$

and  $D(j)$  is the frequency-dependent bias function. When  $D(j) = 0$  for all  $j$  there is no frequency-dependent bias and the model reduces to the unbiased case. If  $D(j) > 0$  for  $j > n/2$  and  $D(j) < 0$  for  $j < n/2$  then transmission is biased in favour of conformity — the probability of acquiring the majority trait ( $j > n/2$ ) indicates that the trait is possessed by more than half the set of  $n$  models) is increased by a factor  $D(j)$ , and the probability of acquiring the minority trait ( $j < n/2$ ) is decreased by the factor  $D(j)$ . Conversely, if  $D(j) < 0$  for  $j > n/2$  and  $D(j) > 0$  for  $j < n/2$  then transmission is biased in favour of non-conformity — the probability of acquiring the majority variant is decreased and the probability of acquiring the minority variant is increased.

Assuming non-selective formation of sets of parents and some value  $k$  such that  $k > n/2$  and  $k$  is minimised (i.e. the lowest value of  $k$  such that  $k$  represents more than half the number of models  $n$ ), it can be shown that the proportion of individuals with variant  $c$ ,  $p'$ , after cultural transmission is:

$$p' = p + \sum_{j=k}^n D(j) \binom{n}{j} [p^j (1-p)^{n-j} - p^{n-j} (1-p)^j]$$

This rather complex equation deserves some explanation. There are  $\binom{n}{j}$  ways to pick  $j$  individuals from a population of size  $n$ . The probability that one of these will exhibit exactly  $j$  individuals with variant  $c$  and  $n - j$  individuals with variant  $d$  is  $p^j (1-p)^{n-j}$  and the probability that one of these will exhibit exactly  $j$  individuals with variant  $d$  and  $n - j$  individuals with variant  $c$  is  $p^{n-j} (1-p)^j$ . There are therefore  $\sum_{j=k}^n \binom{n}{j} (p^j (1-p)^{n-j})$  ways to pick sets of models from the population such that more than half of the models have variant  $c$ , and  $\sum_{j=k}^n \binom{n}{j} (p^{n-j} (1-p)^j)$  ways to pick sets of models such that more than half have variant  $d$ . The proportion of  $c$  in the population therefore increases according to the frequency-dependent bias function  $D(j)$  applied to the difference between the probability of picking sets of models such that the majority are of type  $c$  and the probability of picking sets of models such that the majority are of type  $d$ .

In the case of conformist transmission,  $D(j) > 0$  for  $j > k$  and  $D(j) < 0$  for  $j < k$ . Therefore, the frequency of variant  $c$ ,  $p$ , will increase whenever  $p > 0.5$  (if  $c$  is the more frequent variant in the population then it will increase in frequency) and decrease when  $p < 0.5$  (if  $c$  is the less frequent variant in the population then it will decrease in frequency). The rate of change of  $p$  is at its lowest as  $p$  approaches 1 or 0 (the two saturation points) or 0.5 (the point where the population is perfectly split between the two variants). Conformist transmission results in the spread of the most common cultural variant.

## A.2 Genetic transmission and natural selection

The simplest models of natural selection acting on genetic transmission<sup>1</sup> deal with the changes in frequency of alleles of a single gene in asexually-reproducing haploid populations — each individual has a single gene drawn from a set of  $n$  alleles and each individual inherits the allele of their single parent. In sexually-reproducing diploid populations the equations are complicated by the fact that each individual has two alleles for each gene and receives one allele from each of their two parents.

---

<sup>1</sup>The mathematical model given here is based on the models given in B&R, Hartl & Clark (1997) and Futuyma (1998).

Ontogeny is typically treated in a very simplistic manner in mathematical models of population genetics. In the haploid organism, single gene case there are  $n$  distinct alleles and therefore  $n$  distinct genotypes  $G_1 \dots G_n$ . It is typically assumed that there are  $n$  distinct phenotypes  $F_1 \dots F_n$  and ontogeny maps genotype  $G_i$  onto phenotype  $F_i$ . Selection then acts on the phenotype, but since there is a one-to-one correspondence between genotypes and phenotypes we can talk of selection acting on genotypes and effectively ignore ontogeny.

Suppose that each individual with phenotype  $F_i$  survives with probability  $s_i$ . If  $N_{G_i}$  is the number of individuals with genotype  $G_i$  (and therefore phenotype  $F_i$ ) prior to selection then the number of individuals with genotype  $G_i$  after selection,  $N'_{G_i}$  is:

$$N'_{G_i} = s_i N_{G_i}$$

If we assume that every surviving individual with genotype  $G_i$  leaves, on average,  $o_i$  offspring then the number of individuals with genotype  $G_i$  the next generation,  $N''_{G_i}$ , is given by:

$$N''_{G_i} = o_i N'_{G_i} = s_i o_i N_{G_i} = f_i N_{G_i}$$

where  $f_i$  gives the overall fitness of genotype  $G_i$ , the probability that individuals with that genotype will survive to reproductive age multiplied by the average number of offspring produced.

Now consider a population with two genotypes  $G_a$  and  $G_b$  with fitness  $f_a$  and  $f_b$  respectively. Evolution by natural selection takes place in such a population where the two genotypes do not reproduce at equal rates —  $f_a \neq f_b$ . Typically we are not interested in the absolute numbers of the two genotypes, but the proportion of the population with genotype  $G_a$  and the proportion of the population with genotype  $G_b$ . We will define these as  $P_{G_a} = \frac{N_{G_a}}{N}$  and  $P_{G_b} = \frac{N_{G_b}}{N}$ , where  $N$  is the overall population size ( $N = N_{G_a} + N_{G_b}$ ). We can then calculate the proportion of genotype  $G_a$  at the next generation, which I will denote by  $P'_{G_a}$ :

$$P'_{G_a} = \frac{f_a N_{G_a}}{f_a N_{G_a} + f_b N_{G_b}} = \frac{f_a P_{G_a} N}{f_a P_{G_a} N + f_b P_{G_b} N} = \frac{f_a P_{G_a}}{f_a P_{G_a} + f_b P_{G_b}}$$

The proportion of genotype  $G_a$  at the next generation therefore depends on the proportions of genotypes  $G_a$  and  $G_b$  and their fitness. We can calculate how  $P_{G_a}$  will change over time:

$$\Delta P_{G_a} = \frac{f_a P_{G_a}}{f_a P_{G_a} + f_b P_{G_b}} - P_{G_a} = \frac{P_{G_a} P_{G_b} (f_a - f_b)}{f_a P_{G_a} + f_b P_{G_b}}$$

If  $f_a > f_b$  then genotype  $G_a$  will increase in frequency, and if  $f_a < f_b$  then it will decrease in frequency. The rate of change is at a maximum when  $P_{G_a} P_{G_b}$  is at a maximum, which occurs when  $P_{G_a} = P_{G_b} = 0.5$  — in other words, natural selection depends on genetic diversity, and the rate of evolution is higher when the population exhibits more diversity.

### A.3 Dual transmission and direct bias

Within the dual transmission model, B&R consider the circumstances under which a biological capacity for individual learning and biased and unbiased cultural transmission will be favoured by natural selection. For the purpose of this thesis it is sufficient to review their model of the genetic evolution of direct bias. Recall from Section A.1.2.3 above that direct bias on cultural transmission will increase the frequency of the favoured variant in a population, with the rate of increase depending on the strength of the direct bias, given by the biasing function  $\beta$ , and the cultural variance in the population. B&R expand this model, following their general technique outlined above, to consider the case where  $\beta$  is determined genetically — an individual's genotype determines their preference for cultural variants.

B&R assume that there are two cultural variants,  $c$  and  $d$ , and two genetic variants in the population,  $e$  and  $f$ . Genotypes  $e$  and  $f$  define biasing functions  $\beta_e$  and  $\beta_f$  such that:

$$\begin{aligned}\beta_e(X_i) &= 0 \\ \beta_f(X_i) &= \begin{cases} B & \text{if } X_i = 1 \\ -B & \text{if } X_i = 0 \end{cases}\end{aligned}$$

Recall that  $X_i = 1$  if the learner's  $i$ th cultural parent possesses variant  $c$ , and  $X_i = 0$  otherwise.  $e$  is therefore the unbiased allele and  $f$  is the biased allele, where the bias is in favour of cultural variant  $c$  if  $B > 0$ . We can now calculate the probability that an

individual with genotype  $G$  acquires cultural variant  $c$  given that it is exposed to cultural parents with the cultural variants  $X_1, \dots, X_n$ . This is given by:

$$Prob(c|X_1, \dots, X_n, G) = \frac{\sum_{i=1}^n \alpha_i X_i (1 + \beta_G(X_i))}{\sum_{i=1}^n \alpha_i (1 + \beta_G(X_i))}$$

As before,  $\alpha_i$  gives the intrinsic importance of the  $i$ th cultural parent. This is essentially identical to the equation for directly biased transmission given in Section A.1.2.3, with the addition of a specified genotype  $G$  which gives the particular biasing function  $\beta_G$  to be used.

B&R first assume that genetic parents are selected at random from the pool of possible parents, where the frequency of genotype  $G$  in that pool is  $q_G$ . The frequency of genotype  $G$  among offspring,  $q'_G$ , therefore remains unchanged — there is no natural selection acting on genetic transmission.  $p$  gives the frequency of cultural variant  $c$  in the parent population. Assuming that individuals have just two, equally-weighted, cultural parents, the frequency of individuals with genotype  $G$  and cultural variant  $X$  after cultural transmission,  $F'_{GX}$ , is therefore given by:

$$\begin{aligned} F'_{ec} &= q_e p \\ F'_{fc} &= q_f (p + p(1 - p)B) \\ F'_{ed} &= q_e (1 - p) \\ F'_{fd} &= q_f (1 - p - p(1 - p)B) \end{aligned}$$

As we would expect, individuals with the unbiased allele  $e$  have the same frequency of the two cultural variants as was present in the parent population — individuals with allele  $e$  and variant  $c$  occur with frequency given by the product of the frequency of genotype  $e$  and cultural variant  $c$  ( $p$ ), and individuals with allele  $e$  and cultural variant  $d$  occur with frequency given by the product of the frequency of genotype  $e$  and cultural variant  $d$  ( $1 - p$ ). Among individuals with the biased allele  $f$  cultural variant  $c$  increases in frequency according to the strength of the bias and the cultural variance in the parent population, and variant  $d$  decreases by a similar factor.

B&R then go on to add natural selection to the model. Natural selection weeds individuals out after cultural transmission and prior to breeding, with the probability that an

individual with genotype  $G$  and cultural variant  $X$  survives to breeding age being given by  $W_{GX}$ .  $W_{GX}$  depends on the selective advantage of cultural variant  $c$ ,  $s$ , and the cost of biased transmission,  $z$ :

$$\begin{aligned} W_{ec} &= 1 + s \\ W_{fc} &= 1 + s - z \\ W_{ed} &= 1 \\ W_{fd} &= 1 - z \end{aligned}$$

Individuals with cultural variant  $c$  gain the fitness payoff  $s$ . Individuals with the biased genotype  $f$  suffer the cost of that bias,  $z$ . We can now calculate the expected frequency of individuals with genotype  $G$  and cultural variant  $X$  after selection,  $F''_{GX}$ , according to the equations given above for dealing with natural selection:

$$F''_{GX} = \frac{W_{GX} F'_{GX}}{W_{ec} F'_{ec} + W_{fc} F'_{fc} + W_{ed} F'_{ed} + W_{fd} F'_{fd}}$$

B&R then go on to make several simplifying assumptions. They assume that cultural variant  $c$  is always favoured by selection ( $s > 0$ ) and that bias has no cost or a positive cost ( $z \geq 0$ ) but that these factors are weak ( $z, s \ll 1$ ). Given these assumptions, B&R work through a rather complex set of equations, keeping track of  $q_f$ , the frequency of the biased genotype (henceforth  $q$ ) and  $p$ , the frequency of cultural variant  $c$ .  $q''$  and  $p''$  give the frequencies of these two characters in the next generation.

Assume for a moment that  $p$ , the proportion of individuals with cultural variant  $c$ , is fixed at some arbitrary value. What happens to  $q$ , the frequency of individuals with the biased genotype?

$$q'' = q + vq(1 - q)$$

where  $v$  gives the “selection differential” of the biased allele and is given by:

$$v = B(sp(1 - p)) - z$$

If  $v$  is positive the biased allele will increase in frequency. First consider the case where  $z = 0$  — the biased genotype has no associated cost. If  $p = 0$  or  $p = 1$  then  $v = 0$  and the biased allele does not change frequency — if the population exhibits no cultural variation then the biased allele has no fitness advantage over the unbiased allele and does not change in frequency. If the population exhibits cultural variation then  $v > 0$  and the biased allele will increase in frequency. Now consider the case where  $z > 0$  — the biased genotype has a cost. If the population exhibits cultural variance ( $0 < p < 1$ ) then the sign of  $v$  will depend on the relative values of  $B$ ,  $s$ ,  $p$  and  $z$ . If the population exhibits no cultural variation ( $p = 0$  or  $p = 1$ ) then  $v$  will be negative and the biased genotype will decrease in frequency — the biased allele will suffer a fitness penalty due to its cost and no fitness benefit over the unbiased allele due to the lack of cultural variation. To summarise, in a population which is completely converged culturally (on either variant) the frequency of the biased variant should either remain constant (if biased learning is costless relative to unbiased learning), or decrease (if biased learning has a cost).

What can we predict about the frequency of cultural variant  $c$ , given by  $p$ ? Variant  $c$  is always favoured by selection, and by biased transmission when  $q > 0$ . Therefore variant  $c$  will increase in frequency until the population reaches equilibrium at  $p = 1$ . As discussed in the previous paragraph, at this equilibrium state the biased genotype either has no advantage over the unbiased genotype or is at a disadvantage (where  $z > 0$ ). Therefore, at equilibrium we should expect selection to either be neutral with respect to bias, or to see only the unbiased allele — directly biased transmission pushes the population to converge on the favoured cultural variant, at which point selection pressure on the population's genotypes either stops, or acts to reduce the frequency of the biased allele which drove cultural convergence in the first place.

