

## Supporting Information for ‘Cultural transmission and the evolution of human behaviour: A general approach based on the Price equation’

Claire El Mouden, Jean-Baptiste André, Olivier Morin & Daniel Nettle

### 1. Basic analytical approach

#### 1.1 Deriving the cultural Price equation

To derive a Price equation for the general cultural case, we match every individual  $j$  in a descendant generation with a set  $A_j$  of individuals in the previous generation who have had some kind of cultural influence on him for the trait under consideration. For each individual  $i$  in  $A_j$  we assign a weight  $Y_{ij}$  representing the degree of influence  $i$  has had on  $j$  in the domain of trait  $z$ . Note that since all of  $j$ 's cultural ancestors must, by definition, between them account for all of the cultural influence he receives, then:

$$\forall j, \sum_{i \in A_j} Y_{ij} = 1 \quad (S1)$$

Now for each individual  $i$  in the ancestral generation, we can also consider the set of individuals  $D_i$  in the descendant generation who are influenced by him in the domain of trait  $z$ . This leads naturally to a definition of the cultural fitness of individual  $i$  in that domain. We henceforth denote cultural fitness  $c$  to differentiate it from relative genetic fitness,  $v$ .

$$c_i = \sum_{j \in D_i} Y_{ij} \quad (S2)$$

Note that in a population of constant size the mean cultural fitness is always 1, and so no normalization of cultural fitness by the population mean cultural fitness is required. This is because each ancestor-descendant link, whatever its weight, has one ancestor at one end and one descendant at the other, implying, for a population of  $N$  individuals:

$$\sum_{i=1}^N \sum_{j \in D_i} Y_{ij} = \sum_{j=1}^N \sum_{i \in A_j} Y_{ij} = N \quad (S3)$$

Now we come to expressing the expected evolutionary change in trait  $z$  from one generation to the next. This will be given by  $\Delta \bar{z} = \bar{z}' - \bar{z}$ , where  $\bar{z}'$  is the mean value of  $z$  in the descendant generation, and  $\bar{z}$  is the mean value of  $z$  in the ancestral generation.  $\bar{z}'$  will be given by the population mean of the weighted sum of cultural influences on each individual,

plus the population mean of the extent to which each descendant individual  $j$  spontaneously departs from his cultural influences, which we can denote  $\Delta z_j$ . Thus:

$$\bar{z}' = \frac{1}{N} \sum_{j=1}^N \sum_{i \in A_j} Y_{ij} z_i + \frac{1}{N} \sum_{j=1}^N \Delta z_j \quad (\text{S4})$$

However, because of (S3),  $\sum_{j=1}^N \sum_{i \in A_j} Y_{ij} z_i$  can also be written as  $\sum_{i=1}^N \sum_{j \in D_i} Y_{ij} z_i$ . This means that (S4) can also be expressed as:

$$\bar{z}' = \frac{1}{N} \sum_{i=1}^N \sum_{j \in D_i} Y_{ij} z_i + \frac{1}{N} \sum_{j=1}^N \Delta z_j \quad (\text{S5})$$

As for  $\bar{z}$ , this is simply given by  $\frac{1}{N} \sum_{i=1}^N z_i$ . Thus, we have:

$$\Delta \bar{z} = \bar{z}' - \bar{z} = \frac{1}{N} \sum_{i=1}^N \sum_{j \in D_i} Y_{ij} z_i + \frac{1}{N} \sum_{j=1}^N \Delta z_j - \frac{1}{N} \sum_{i=1}^N z_i \quad (\text{S6})$$

Given the definition of cultural fitness  $c$  given in (S2), the first term on the right-hand side can be rewritten as  $E(c_i z_i)$ , whilst the last can be rewritten as  $E(z_i)$ , and the middle one as  $E(\Delta z_j)$ . Given that  $1 = E(c_i)$ , equation (S6) can be expressed as:

$$\Delta \bar{z} = E(c_i z_i) - E(c_i)E(z_i) + E(\Delta z_j) \quad (\text{S7})$$

Applying the definition of covariance, we obtain a Price equation in the following, familiar form:

$$\Delta \bar{z} = \text{cov}(c_i, z) + E_c(\Delta z) \quad (\text{S8})$$

Thus, even though the cultural case involves variable numbers of ancestors, and continuously variable levels of influence of each ancestor on each descendant, the Price equation holds in its standard form, albeit with cultural fitness instead of genetic fitness. Note that we do not require a third term on the right-hand side of the Price equation to capture variation in number of ancestors per descendant, as does the approach of Kerr and Godfrey-Smith (2009) designed to generalize the Price equation in a similar way that we do here. This is because values of the trait that lead to their being fewer ancestors per descendant necessarily involve those ancestors having higher influence weights  $Y$  on

average. Thus, our approach captures potential covariance between trait-values and the numbers of ancestors per descendant without the need for a separate third term.

As we describe in the main text (section 2), the cultural Price equation (S8) can also be decomposed into a form containing an unweighted expectation and a covariance, as in (S9).

$$\Delta\bar{z} = cov(c, z) + E(\Delta z) + cov(c, \Delta z) \quad (S9)$$

### 1.2 Decomposing cultural fitness

This section provides a method for decomposing the various components of both cultural and genetic fitness, and exploring how genetic and cultural fitness are aligned in a given system. We can decompose both genetic and cultural fitness into their components by assuming that an individual's fitness (in either sense) can be predicted by a series of factors  $f_1, f_2 \dots f_n$ , each of which has a different weight, which may be positive, negative or zero. We can thence rewrite the cultural fitness of individual  $i$  as:

$$c_i = c_0 + \beta(c, f_1)f_{1i} + \beta(c, f_2)f_{2i} + \dots + \beta(c, f_n)f_{ni} + \varepsilon_i \quad (S10)$$

Here, the  $\beta$  represent partial regression coefficients of cultural fitness on each of the factors  $f$  in turn, holding all the others constant, and  $\varepsilon$  represents a residual. Equation (S10) leads to an expansion of the selection term of equation (S8):

$$\begin{aligned} cov(c, z) &= cov(c_0 + \beta(c, f_1)f_{1i} + \beta(c, f_2)f_{2i} + \dots + \beta(c, f_n)f_{ni} + \varepsilon_i, z) \\ &= \beta(c, f_1)cov(f_1, z) + \beta(c, f_2)cov(f_2, z) + \dots + \beta(c, f_n)cov(f_n, z) \end{aligned} \quad (S11)$$

In essence, (S11) states that the relationship of a trait-value to cultural fitness will depend upon the sum of the trait's covariances with the component factors of cultural fitness, each weighted by the contribution of that factor to overall cultural fitness in that domain and population at that time. An identical decomposition can be done for genetic fitness with the same set of factors  $f$ , but  $v$  in place of  $c$ . Thus, the relationship between cultural and genetic fitness is given by the relationship between the sets of weights  $\sum_{k=1}^n \beta(c, f_k)$  and  $\sum_{k=1}^n \beta(v, f_k)$ .

It follows that, in principle, to evaluate how biological and cultural fitness overlap for any quantifiable trait (e.g. religiosity), we would identify all the possible factors that could affect either the cultural or genetic fitness of the people doing it (e.g. longevity, mating success, social status, social connectedness, social visibility) and then write two forms of (S11) to describe the impact of each factor on biological and cultural fitness. We could then compare these two equations and evaluate how similar the sets of  $\beta$ s are for cultural and genetic fitness.

## 2. Can culture select for behaviours that are genetically deleterious?

### 2.1 Conditions for cultural evolution to increase or decrease genetic fitness

In this section, we address the question of whether and when cultural transmission increases genetic fitness. A simple way to do this is to ask what we should expect the change in genetic fitness to be if genetic fitness itself were transmitted genetically versus culturally. More exactly, it is all the behaviours that contribute to fitness that could be transmitted genetically or culturally, but the result is the same. We are also considering only direct selective effects. There will often also be indirect effects of natural selection on population mean fitness via environmental change, as for example when population growth leads to greater competition (Frank & Slatkin, 1992). These are not central to our argument, which concerns the direct effects of cultural vs. genetic selection.

The expected change in genetic fitness due to selection under genetic transmission would be:

$$\Delta_g \bar{w} = \text{cov}(w, w) = \text{var}(w) \quad (\text{S12})$$

Here,  $w$  is absolute, rather than relative, genetic fitness, related to relative genetic fitness by  $v = \frac{w}{\bar{w}}$ . Equation S12 is Fisher's fundamental theorem (Fisher, 1930, Price, 1972). By contrast, the change in genetic fitness due to selection under cultural transmission would be:

$$\Delta_c \bar{w} = \text{cov}(c, w) = \beta(c, w) \text{var}(w) \quad (\text{S13})$$

Here  $\beta(c, w)$  is the overall regression coefficient of cultural on genetic fitness, and it stems from the alignment of the components of the two fitnesses (see section 1.2). It follows from (S13) that:

$$\Delta_c \bar{w} = \beta(c, w) \Delta_g \bar{w} \quad (\text{S14})$$

That is, the change in mean genetic fitness brought about by cultural selection is determined by the product of the coefficient of the regression relationship between cultural and biological fitness, and the change in genetic fitness that would occur under genetic transmission. Where  $\beta(c, w)$  is positive, cultural selection increases genetic fitness, and as  $\beta(c, w)$  approaches 1, cultural selection does exactly the same thing as genetic selection, and at the same rate. However, where cultural and biological fitness become negatively associated, then cultural selection will tend to reduce the mean genetic fitness of the population. In other words, cultural selection will tend to reduce genetic fitness when  $\beta(c, w)$  comes to be negative. However, we have thus far only considered the change due to selection, and this is not the only component of evolutionary change. A fuller expression for the expected change in genetic fitness due to cultural evolution is given by:

$$\Delta_c \bar{w} = \beta(c, w) \Delta_g \bar{w} + E_c(\Delta_c w) \quad (\text{S15})$$

Thus, for a behaviour that reduces genetic fitness to spread culturally requires not just that  $\beta(c, w) < 0$  is met, but in fact that:

$$\beta(c, w) < -E_c(\Delta_c w) \quad (\text{S16})$$

In other words, for cultural evolution to operate in a contrary direction to biological imperatives, cultural fitness has not just to be negatively related to genetic fitness, but has to be strong enough to overwhelm any evolved psychological biases in individuals making them averse to learning or adopting behaviours injurious to their reproductive success.

## *2.2 Effect of genetic selection on the covariation between genetic and cultural fitness*

Whenever cultural and genetic fitness are not perfectly correlated under the social learning strategies currently extant in the population, then genetic selection favours any mutant learning strategy that increases their correlation. To see why, assume that the current learning strategy employed by population members produces on average  $\beta(c, w) = \tau$ , where  $\tau < 1$ . Now imagine a mutant strategy that causes people to pay attention to different cultural models, and produces an average value of  $\beta(c, w) = \tau + \delta$ , where  $\delta$  is a small positive value. The genetic fitness of the mutant learner is  $w_0 + (\tau + \delta) \Delta_g w$ , whereas that of the wild type is  $w_0 + \tau \Delta_g w$ . As the former is necessarily greater than the latter,

natural selection on genes will always favour tuning the organism's use of social learning in such a way as to bring cultural fitness into closer and closer alignment with genetic fitness. Indeed, it is straightforward to see that inclusive fitness is maximized when both fitnesses are perfectly correlated ( $\beta(c, w) = 1$ ).

Genetic selection can tune the alignment between genetic and cultural evolution in two ways. First, it can introduce motivational or cognitive biases within the individual, making the Price Equation's transmission component relatively more important than the cultural selection component in such a way as to neutralize maladaptive cultural trends. Second, it can tune the components of cultural fitness by altering individuals' social learning strategies, making them prone to imitate certain types of content (what Boyd and Richerson (1985) call 'direct bias'), or people with certain types of attribute ('indirect bias'). Thus, it is plausible that genetic evolution closely shapes social learning strategies such as to make *cultural* transmission promote *genetic* fitness.

### 3. Cultural evolution of altruism

The evolution of altruism through cultural selection is usually framed as a multi-level selection process, where the altruistic behaviour is negatively associated with fitness at the individual level, but positively associated with fitness at the group level (Henrich, 2004). The Price equation has long been used to capture the evolutionary dynamics of such multi-level selection situations (Price, 1972, Okasha, 2006). It does so by partitioning the change due to selection term  $cov(c, z)$  into two subcomponents, so that (using an unorthodox notation for simplicity):

$$cov(c, z) = cov_{group}(\bar{c}, \bar{z}) + E[cov_{individual}(c, z)] \quad (S17)$$

Here, the first term is the covariance between the level of the trait in the group ( $\bar{z}$ ), and the cultural fitness of that group ( $\bar{c}$ ), that is, its ability to persist and influence other groups, whereas the second term is the expectation across all groups of the covariance within the group between the level of the trait and the cultural fitness of the individual. By rewriting (S17) as:

$$cov(c, z) = \beta_{group}(\bar{c}, \bar{z}) \cdot var_{group}(\bar{z}) + E[\beta_{individual}(c, z) \cdot var_{individual}(z)] \quad (S18)$$

- we see that what matters for predicting which traits are selected for is the relative strengths of cultural selection on the trait at the individual and the group level, and the relative sizes of the trait variances within and between groups.

The very same argument - that cultural selection could favour altruism more broadly than genetic selection would - can be made in a logically equivalent way with no reference to group selection, drawing instead on the notion of cultural kin selection (Allison, 1992). Cultural evolution theorists are well aware of this possibility (see e.g. Henrich, 2004, Boyd & Richerson, 2010), but it has been relatively little discussed, perhaps because of a residual tendency in the human behavioural ecology community to see group selection and inclusive fitness as different evolutionary forces, rather than different but equivalent notations for the same force. Indeed, it is common to find papers arguing that cultural group selection is an alternative to kin selection as an explanation for human altruism (Fehr & Fischbacher, 2003). In fact, cultural group selection can be seen as a form of kin selection, the kinship being cultural rather than genetic.

To re-express the cultural group selection argument as cultural kin selection, let us assume that individuals interact socially with neighbours in ways that affect their cultural fitness, and let the trait of interest  $z$ , now describe the level of altruistic values,  $a$ . The cultural fitness of a focal individual is then given by (13).

$$c_f = c_0 + \beta(c, a_f)a_f + \beta(c, a_n)a_n + \varepsilon_f \quad (\text{S19})$$

Here,  $c_0$  is the background cultural fitness in the population,  $a_f$  is the focal's level of altruistic values,  $a_n$  is the level of altruistic values of the focal's neighbours, and  $\varepsilon$  is a residual. The  $\beta$ s are the partial regression coefficients of cultural fitness on the level of the focal's own altruistic values, holding neighbours' values constant  $\beta(c, a_f)$ , and on the neighbours' altruistic values, holding the focal's values constant  $\beta(c, a_n)$ . The cultural change in  $a$  due to cultural selection can now be written, following Queller (1992), as:

$$\Delta \bar{a}_{cs} = \text{cov}(c, a) = \text{var}(a)[\beta(c, a_f) + \beta(a_n, a_f) \cdot \beta(c, a_n)] \quad (\text{S20})$$

Expression (S20) gives rise directly to a general form of Hamilton's rule (Hamilton, 1964). An altruistic value  $a$  will be culturally selected for exactly when:

$$\begin{aligned}
cov(c, a) &> 0 \\
\beta(c, a_f) + \beta(a_n, a_f) \cdot \beta(c, a_n) &> 0 \\
\beta(a_n, a_f) \cdot \beta(c, a_n) &> -\beta(c, a_f)
\end{aligned} \tag{S21}$$

Here,  $\beta(c, a_n)$  is the personal cultural fitness benefit of having neighbours with the value,  $-\beta(c, a_f)$  is the personal cultural fitness cost of having the value, and  $\beta(a_n, a_f)$  is cultural relatedness under the regression definition of relatedness (the extent to which the focal's social values predict those of his neighbours: Hamilton (1972)). Thus, the condition for cultural selection to favour helping others in the surrounding group increase their cultural influence depends only on the costs of such helping to one's own cultural influence, the benefits of the help to the cultural influence of its recipients, and, crucially, the extent to which the population is assorted such that individuals of like values interact. The 'cultural group selection' and 'cultural kin selection' conditions expressed in (S18) and (S21) respectively are equivalent decompositions of (S8).

Expressions (S18) and (S21) are the conditions for cultural natural selection to favour *cultural* altruism. A behaviour that is culturally altruistic is not necessarily genetically altruistic, and there is no reason to believe that behaviours that are good for the survival of human groups are in general culturally altruistic. However, when people discuss the evolution of altruism through cultural group/kin selection, what they have in mind appears to be behaviour that provides *genetic* benefits to human groups, even though the selection mechanism is cultural. To argue that a trait evolving through cultural kin/group selection provides a genetic benefit to group-mates requires the further assumption that genetic and cultural fitness are perfectly coupled in that population under the currently extant learning rules, so that all cultural costs are also genetic costs, and all genetic costs are also cultural costs. In the rest of this section, we consider the case where such a perfect coupling is in place.

In such a population, cultural natural selection favours being culturally altruistic to the point where:

$$\beta(a_n, a_f) > \frac{-\beta(c, a_f)}{\beta(c, a_n)} \tag{S22}$$

This is a rearrangement of cultural Hamilton's rule (S21). Now, in the same population, genetic selection favours a trait for a genetically altruistic behavior when:

$$\beta(g_n, g_f) > \frac{-\beta(v, a_f)}{\beta(v, a_n)} \quad (S23)$$

This is genetic Hamilton's rule,  $v$  represents biological fitness, and the left-hand side represents genetic relatedness. Now, by assuming that the trait of interest is one that provides both cultural and genetic fitness benefits, which results from assuming that cultural and genetic fitness are perfectly linked ( $c = v$ ) as discussed above, we obtain an expression for the point to which genetic selection favours behavior that is both culturally and genetically altruistic:

$$\beta(g_n, g_f) > \frac{-\beta(c, a_f)}{\beta(c, a_n)} \quad (S24)$$

Now, comparing (S24) to (S22), altruistic behaviour will be selected for under cultural but not genetic transmission in the interval:

$$\beta(g_n, g_f) < \frac{-\beta(c, a_f)}{\beta(c, a_n)} < \beta(a_n, a_f) \quad (S25)$$

This is the additional range of altruism culturally selected for which is not genetically selected for, and thus if cultural relatedness is higher than genetic relatedness, then cultural transmission does indeed favour more widely altruistic behaviour than genetic transmission. However, it is obvious that genetic selection *always* disfavours altruism in this interval, precisely because (S24) is not met. Thus, to the extent that cultural relatedness makes us take on real fitness costs for our 'cultural' brothers and sisters, natural selection on genes is always opposed to this, and there is antagonistic selection on genes to shut dependence on culture down. Therefore, as André and Morin (2011) have also argued, purely cultural assortment can never lead to the *genetic* evolution of propensities to be biologically altruistic.

So far we have only considered the impact of cultural *selection* on altruism, and neglected the transmission component. A full statement of the condition for altruism to evolve is obtained by substituting (S20) into (S8):

$$\text{var}(a)[\beta(c, a_f) + \beta(a_n, a_f) \cdot \beta(c, a_n)] + E_c(\Delta a) > 0 \quad (\text{S26})$$

Expression (S26) allows us to see that even where condition (S22) is met, altruistic behaviour may not be able to evolve. In fact, altruistic behaviour requires that the stricter condition (S27) be met.

$$\text{var}(a)[\beta(c, a_f) + \beta(a_n, a_f) \cdot \beta(c, a_n)] > -E_c(\Delta a) \quad (\text{S27})$$

Given the conflict of interest between genes and culture described above, and given that humans have a long evolutionary history of being surrounded by individuals whose genetic relatedness is relatively low but who are prone to influencing their behaviour, genetic selection should have produced evolved biases or motivational mechanisms to resist doing things in the genetic interests of those other individuals rather than oneself. The consequence would be that genetically altruistic values may be resisted and distorted by individual psychology (substantial  $E_c(\Delta a)$  term). These kinds of transmission biases have generally been ignored in previous discussions of the evolution of altruism through cultural assortment, but they would be a predictable outcome if cultural assortment tended to be stronger than genetic assortment over evolutionary time.

## References

- Allison, P. 1992. The cultural evolution of beneficent norms. *Social Forces* **71**: 279-301.
- Boyd, R. & Richerson, P. J. 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Boyd, R. & Richerson, P. J. 2010. Transmission coupling mechanisms: cultural group selection. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**: 3787-3795.
- Fehr, E. & Fischbacher, E. 2003. The nature of human altruism. *Nature* **425**: 785.
- Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- Frank, S. A. & Slatkin, M. 1992. Fisher's fundamental theorem of natural selection. *Trends in Ecology & Evolution* **7**: 92-95.
- Hamilton, W. D. 1964. The genetical evolution of social behaviour, I and II. *Journal of Theoretical Biology* **7**: 1-16, 17-52.
- Hamilton, W. D. 1972. Altruism and related phenomena, mainly in the social insects. *Annual Review of Ecology and Systematics* **3**: 192-232.
- Henrich, J. 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization* **53**: 3-35.
- Kerr, B. & Godfrey-Smith, P. 2009. Generalization of the Price equation for evolutionary change. *Evolution* **63**: 531-536.
- Okasha, S. 2006. *Evolution and the Levels of Selection*. Oxford University Press, Oxford.
- Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics* **35**: 485-&.
- Queller, D. C. 1992. Quantitative genetics, inclusive fitness, and group selection. *American Naturalist* **139**: 540-558.