JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12296

Cultural transmission and the evolution of human behaviour: a general approach based on the Price equation

C. EL MOUDEN*, J.-B. ANDRÉ†, O. MORIN‡ & D. NETTLE§

*Department of Zoology & Nuffield College, University of Oxford, Oxford, UK
†Ecologie & Evolution, CNRS UMR 7625, Institut de Biologie de l'Ecole Normale Superieure, Paris, France
‡Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria
§Centre for Behaviour and Evolution & Institute of Neuroscience, Newcastle University, Newcastle, UK

Keywords:

cultural evolution;
cultural fitness;
cultural group selection;
dual-inheritance theory;
human behavioural ecology;
kin selection;
Price equation.

Abstract

Transmitted culture can be viewed as an inheritance system somewhat independent of genes that is subject to processes of descent with modification in its own right. Although many authors have conceptualized cultural change as a Darwinian process, there is no generally agreed formal framework for defining key concepts such as natural selection, fitness, relatedness and altruism for the cultural case. Here, we present and explore such a framework using the Price equation. Assuming an isolated, independently measurable culturally transmitted trait, we show that cultural natural selection maximizes cultural fitness, a distinct quantity from genetic fitness, and also that cultural relatedness and cultural altruism are not reducible to or necessarily related to their genetic counterparts. We show that antagonistic coevolution will occur between genes and culture whenever cultural fitness is not perfectly aligned with genetic fitness, as genetic selection will shape psychological mechanisms to avoid susceptibility to cultural traits that bear a genetic fitness cost. We discuss the difficulties with conceptualizing cultural change using the framework of evolutionary theory, the degree to which cultural evolution is autonomous from genetic evolution, and the extent to which cultural change should be seen as a Darwinian process. We argue that the nonselection components of evolutionary change are much more important for culture than for genes, and that this and other important differences from the genetic case mean that different approaches and emphases are needed for cultural than genetic processes.

Introduction

Over several decades, a number of influential scholars have argued that cultural change in humans can be modelled as an evolutionary process (see Mesoudi *et al.*, 2004, 2006 for overviews). This 'cultural evolution' literature in fact encompasses a number of different approaches using different assumptions (Godfrey-Smith, 2012). One particularly influential approach is the body of 'dual-inheritance' theory stemming from such works as Cavalli-Sforza & Feldman (1981) and Boyd & Richerson (1985). This approach is based on the insight that each human inherits not just a genotype, but also a set of socially learned information. A key claim of the theory is that this body of socially learned information (i.e. culture) evolves partly through a Darwinian selection process of its own. Dual-inheritance theorists argue that the outcome of evolution in systems featuring dual inheritance can, under certain circumstances, differ from what the outcome would be under genetic inheritance alone. For example, it has been claimed that cultural evolution favours altruism under conditions where it would be not be selected for genetically (Boyd & Richerson, 1985; Henrich, 2004; Bell *et al.*, 2009).

However, the dual-inheritance literature has developed via specific models whose results rely upon

Correspondence: Daniel Nettle, Centre for Behaviour and Evolution & Institute of Neuroscience, Newcastle University, Henry Wellcome Building, Framlington Place, Newcastle NE2 4HH, UK;
Tel.: +44 191 222 8993; fax: +44 191 208 5227;
e-mail: daniel.nettle@ncl.ac.uk

particular and often quite strong assumptions (regarding e.g. population structure and available learning strategies) and idealizations (such as natural selection on genes being negligibly weak compared with cultural selection). As a consequence, different theorists can make different models and arrive at diametrically opposed predictions about the same phenomena. For example, one set of models finds that helping behaviour can evolve more easily under cultural than genetic transmission (Boyd & Richerson, 1985; Boyd *et al.*, 2011), whereas other models using different assumptions predict the opposite result is possible, that sometimes helping is less likely to evolve under cultural than genetic selection (Lehmann *et al.*, 2008; Molleman *et al.*, 2013). Without a more general level of theory, there is no obvious way of adjudicating between these different views. There is also debate about what terms such as adaptive, fitness, relatedness, altruism and so forth mean in a system with both cultural and genetic components, and even whether Darwinian evolution is an appropriate basic metaphor for cultural change (see Sperber, 1985; Atran, 2001; Claidière & André, 2012; Claidière *et al.*, 2013). Thus, it would be useful for progress in dual-inheritance theory if there were a general formal framework onto which specific cases and models could be mapped, just as there is for genetic evolution.

In biology, the general framework to describe evolution is provided by the Price equation and work arising from it (Price, 1970, 1972; Queller, 1992, 2011; Frank, 1997; Grafen, 2002; Taylor, 2009; Gardner *et al.*, 2011). The Price equation describes the change in any system where there is evolution by natural selection (Price, 1970, 1972; Gardner, 2008). It relies on no assumptions about the nature of the evolving entities or selective forces, the population structure or the system of inheritance. The generality of the Price equation makes it attractive for modelling aspects of cultural evolution such as multilevel selection (Anderson, 2004; Henrich, 2004; Boyd & Richerson, 2010), and it has been extended from purely genetic evolution to situations where there are multiple types of heritable influence (Collins & Gardner, 2009; Helantera & Uller, 2010; Day & Bonduriansky, 2011). However, the potential utility of the Price equation for clarifying what the attributes of cultural evolution are at the individual level, and how cultural processes interact with genetic evolution, has not yet been fully exploited.

Here, we use the Price equation to explore evolutionary change in a system with both genetic and cultural inheritance. In line with previous dual-inheritance models, the evolutionary units we shall consider (the 'Darwinian individuals' in the sense of Godfrey-Smith (2012)) are human beings. This is not the only possible approach to cultural evolution; for example, the cultural artefacts or performances themselves could be considered as the evolutionary units (see e.g. Rogers & Ehrlich, 2008). However, our main purpose here is to explore the evolutionary consequences of a 'dual inheritance', and the only way to do this is to define the same evolutionary unit – the individual person – for both sources of heritable influence. For this reason, we do not consider these other approaches to culture any further in the current paper.

The first section of the paper presents a cultural version of the Price equation and uses this to define key concepts such as cultural fitness. Section 'Interplay between cultural and genetic evolution' examines the impact of cultural transmission on genetic fitness, and in particular when cultural evolution can favour traits that are deleterious to genetic fitness. Section 'Cultural transmission and the evolution of altruism' examines in the detail the evolution of altruism in a system with both genetic and cultural transmission. The 'Discussion' section discusses the implications of our formal analysis for how human culture should be understood. Specifically, we ask to what extent cultural change should be seen as a Darwinian process, and to what extent it is autonomous from genetic evolution. Throughout, our formal analysis is restricted to the Supporting Information (Document S1), with the main text providing a narrative description of the results.

A Price equation approach to cultural evolution

The Price equation is a simple mathematical statement about change. It describes how any measurable trait z (e.g. body size, bride price, degree of altruism) changes from one generation to the next.

The Price Equation for genetic evolution

When used to study genetic evolution, the simplest form of the Price equation (1) partitions the total evolutionary change into two components:

$$\Delta\bar{z} = cov(v, z) + E_v(\Delta z) \qquad (1)$$

where $\Delta\bar{z}$ represents the change in the population average value of a trait z from one biological generation to the next. The first term on the right-hand side of (1) represents the change due to *natural selection.* The expression *cov(v,z)* denotes the covariance between individuals' values of z and their relative genetic reproductive success, or relative fitness, v; that is, their number of descendants in the next generation divided by the population mean number of descendants. For example, if individuals with larger bodies tend to have more surviving offspring, then the population average body size will increase, whereas if they have fewer surviving offspring, the population average of the trait will decrease. If there is no association between body size and surviving offspring, the population average body size will not change due to natural selection.

The second term on the right-hand side of (1) describes the change due to *transmission*. Formally, the expression $E_v(\Delta z)$ represents the fitness-weighted average across the population of the change in trait value between an ancestor and its descendant. If offspring are identical copies of their parents, then this term will be zero. If, for example due to improving nutrition, all offspring tended to have larger bodies than their parents, then this term would be nonzero, as the population average body size is increasing through time in the absence of selection.

We note two points about the Price equation. First, eqn (1) shows that there will always be change due to natural selection as long as there is any covariance between individuals' fitness and their values of the trait z. The only situations in which there is no change due to selection is where either all individuals have the same trait value, all individuals have the same fitness, or there is no association between an individual's trait value and their fitness. Second, because the population mean of z will change in each generation towards the higher-fitness value of z, the direct effect of natural selection is to increase genetic fitness. To be precise, natural selection increases the *inclusive* genetic fitness of individuals, rather than personal reproductive success, where inclusive fitness describes an individual's impact on the reproductive success of all its relatives (and itself), weighted by the genetic relatedness of each (Grafen, 2006; Gardner, 2008).

The Price equation for cultural evolution

We now turn to cultural evolution. Here, the ancestor-descendant relationships represent cultural influence rather than genetic parenthood. Person A is a cultural ancestor of person B if the value of z person B has was influenced by the value of z person A had. Cultural generations are therefore defined by instances of social learning; they are not the same as biological ones. They can potentially be much shorter than biological generations, but also potentially much longer in a culture with writing or other forms of lasting recording. Note that individuals can only directly culturally influence those in the next generation, because the next generation is *defined* in terms of direct cultural influence. A consequence of this is that a person's set of cultural ancestors for any particular trait can be very heterogeneous in biological age, containing, for example, peers, members of the previous biological generation and some long-dead individuals who have left recordings, writings or artefacts. The pace of generations may also be different for different cultural traits, depending on how often those traits are learned. Moreover, individual humans can be both ancestor and descendant of each other (though not within the same cultural generation) if they transmit an idea to someone else and then learn a derived form of it back at a later time point (for discussion of implications of this, see Strimling *et al.*, 2009). We return to the complexity in delineating the set of cultural ancestors and the limitations that imposes on our approach, in the 'Discussion' section.

In the standard biological derivation of the Price equation, given the nature of genetic inheritance, the number of ancestors per descendant is taken to be the same for all members of the population, and all ancestors of a given descendant are taken to have an equal 'share' in the phenotype of that descendant (Price, 1970; p. 521). It is because of these assumptions that genetic fitness is simply a function of the number of descendants a focal has, as the share an ancestor has in each of them is fixed. By contrast, in the cultural case, the number of ancestors of any given descendant could vary from individual to individual, possibly in ways that depend on the trait value of the ancestors. Furthermore, it is natural to assume that some cultural ancestors of a given descendant are more influential than others. Therefore, cultural fitness is a measure of cultural influence, reflecting both the number of people who learn from an individual, and the degree to which their traits are influenced when they do learn. This definition of cultural fitness is similar to previous ones (Griffiths *et al.*, 2008), although our formulation (see Section 1 in Document S1) shows that it is not just the *number* of people in subsequent generations who are influenced by the focal individual that is maximized, but also the *degree* of influence on each of those people. Thus, cultural fitness captures the overall influence that a cultural trait in an ancestral generation has on cultural traits in a descendant generation. In Section 'Cultural transmission and the evolution of altruism' (Section 3 in Document S1), we discuss how cultural fitness, just like genetic fitness, can be subdivided into either within- and between-group fitnesses, or into direct and indirect fitness effects, to study evolution from a group or individual perspective, respectively.

We use this definition of cultural fitness to derive a cultural version of the Price equation, by allowing the number of ancestors of each descendant to be variable rather than fixed, and assigning each ancestor-descendant link with a weight, rather than merely existing or not. Despite this extra complexity, we find that – for our assumed scenario of an isolable, measurable cultural trait influenced by that same trait in some set of cultural ancestors – the equation in its familiar form still holds (S8). Thus, the *cultural Price equation* can be expressed as:

$$\Delta\bar{z} = cov(c, z) + E_c(\Delta z) \qquad (2)$$

where c is cultural, rather than genetic fitness and other terms have the same meaning as in equation (1) (see Table 1 for formal definitions of terms). The first term on the right-hand side measures the change due to *cultural natural selection*. If individuals with a high value of z are more culturally influential than those

Table 1 Informal and formal definitions of key evolutionary concepts as applied to culture. Formal expressions use the notation established in the Document S1.

Concept	Informal definition	Formal expression
Cultural fitness	The total quantitative influence of the cultural phenotype of an individual in an ancestral generation on the cultural phenotypes of individuals in the descendant generation, for the trait under consideration	c
Cultural natural selection	The generation-to-generation change in the population mean of cultural traits, arising from covariance between traits and cultural fitness.	$cov(c,z)$
Cultural altruism	Behaviour that reduces the cultural fitness of the actor, whilst increasing the cultural fitness of those with whom the actor interacts	$\beta(c,z_f)<0$ *and* $\beta(c,z_n)>0$
Cultural relatedness	The regression coefficient of the relationship between the cultural values of an actor and those of the people with whom he interacts, for a particular culturally transmitted trait	$\beta(z_n,z_f)$
Cultural Hamilton's rule	The condition that a behaviour that reduces the actor's lifetime cultural influence can only be culturally selected for if the cost to him is less than the product of the cultural benefit to his interaction partners and their cultural relatedness to him	$\beta(c,z_f)+\beta(z_n,z_f)\cdot\beta(c,z_n)>0$

with a low value, the population mean of the trait will increase due to selection, and *vice versa*. It follows that just as genetic natural selection maximizes genetic fitness, cultural natural selection maximizes cultural fitness. Note that an individual's cultural fitness, c, and their genetic fitness, v, are distinct quantities, the one representing the degree to which others have socially learned from them, and the other representing the number of surviving children they have had relative to others in the population. Thus, the cultural evolutionary consequences of cultural selection are not the same as the genetic evolutionary consequences, and key concepts such as adaptation and maladaptation within cultural evolution should not be conflated with their genetic counterparts (Mesoudi *et al.*, 2004).

Comparing cultural and genetic fitness

Although cultural and genetic fitness are logically distinct, it is possible that empirically they will correlate. Many factors will influence each type of fitness, and some factors may be common to both. For example, dying in infancy means you can neither have children nor influence many people. In Document S1 (Section 1.2), we show formally how to define the correlation between cultural and genetic fitness and identify the factors contributing to each type of fitness.

Although the concepts of genetic and cultural fitness appear similar, the cultural case has complexities the genetic one lacks. In genetic inheritance, individuals have the same parents for all of their traits. This common inheritance means that, leaving aside exceptional cases such as segregation distortion, all an individual's genes will have the same fitness value. As a result, biologists can empirically measure genetic fitness by measuring individual fitness. By contrast, in cultural inheritance, individuals can learn one trait from one set of cultural ancestors and another trait from a completely different set. This means that an individual has many fitnesses, each corresponding to a different cultural trait. For example, Beethoven had enormous influence in the domain of music, and none at all in the domain of cookery. Therefore, although Beethoven's overall genetic fitness can be measured, it would be impossible to assign him a single value for cultural fitness. We return to this and other differences between the genetic and cultural cases in the 'Discussion' section.

Transmission effects in cultural evolution

The second term on the right-hand side of equation (2) captures the transmission component of evolutionary change. In genetic evolution, this component is usually assumed to be zero, because fair meiosis and random mutation ensure that there is no systematic tendency for offspring to differ from their parents in one particular direction. As a result, biologists often make predictions assuming that natural selection is the only force at work. By contrast, for cultural change, the transmission term cannot be ignored as it may represent a large portion of overall cultural evolutionary change (Claidière *et al.*, 2013). This term captures any tendency of human minds to transform representations or practices in the course of teaching, learning, remembering and implementing them. To understand the different types of transmission effects, it is useful to decompose this transmission component further (following Okasha, 2006, p. 26). This yields:

$$E_c(\Delta z) = E(\Delta z) + cov(c, \Delta z) \quad (3)$$

From (3), we can see that there are two key sources of potential cultural evolutionary change other than cultural natural selection. The first occurs if there is a uniform tendency of humans to systematically transform traits through learning or other cognitive processes ($E(\Delta z)$). This is what Heywood (2005) calls a *constitutive transmission bias*. It is the change in the trait that would occur in each generation even if there was no variation in cultural fitness. There is extensive evidence from experimental psychology that such constitutive biases (sometimes referred to as content biases or cognitive attractors) exist and exert a strong influence on the outcome of iterated social learning (Griffiths *et al.*, 2008).

The second way that there will be a nonselective influence on cultural change is if the amount by which individuals modify their cultural inheritance depends on their level of cultural influence (i.e. the $cov(c,\Delta z)$ term is not zero). This will often be the case. For example, people might adhere uncritically to inherited practices as long as they are successful and influential, but when their success and influence is low, seek to modify those practices in particular directions. This might reflect, for example, their reasoning about the causes of social success. This is what Heywood (2005) refers to as *induced transmission bias*, that is, a component of evolutionary change that arises from the fact that some individuals do better than others, exerts a directional influence on trait values, and yet is not natural selection. Again, this is likely to be a much more important force in cultural change than it is in genetic evolution, where mutation is generally independent of fitness.

Interplay between cultural and genetic evolution

A popular view in the dual-inheritance literature is that cultural evolution was central to human biological success as it made the scale and distribution of today's human population possible. This assumes that *cultural* evolution served to increase *genetic* fitness. For this to be true, cultural fitness for the trait under consideration must be positively correlated with genetic fitness (Section 2.1 in Document S1); that is, the individuals leaving most offspring must also on average have the highest cultural influence. This correlation may hold as, for example, markers of social status tend to be positively correlated with reproductive success across different kinds of society (Irons, 1979; Nettle & Pollet, 2008). However, social status is not necessarily a good proxy for cultural fitness, since people could accord status to a particular individual without adopting that individual's traits. The association between cultural and genetic fitness is unclear and needs further empirical investigation.

Conversely, the correlation between cultural and genetic fitness could be negative. In these cases, cultural natural selection would act to decrease genetic fitness. A putative example is low fertility in developed countries. Richerson & Boyd (2005) argue that in wealthy societies, high social standing, and therefore cultural influence, goes to members of professions that require long investment in education. If this is correct, any behaviour associated with reaching such a profession would be likely to spread under cultural natural selection despite the fact they have fewer children. However, we show in Document S1 (Section 2.1) that the circumstances where cultural evolution reduces genetic fitness are evolutionarily unstable. First, the transmission component of cultural evolution is likely to be substantial. This transmission is driven by human minds that are genetically adapted to avoid behaviours that are detrimental to their biological interests. For example, an individual's experience of nausea, which is a biological adaptation to prevent poisoning, is sufficient to overcome cultural pressure to eat unpleasant food. This would introduce a dramatic transmission bias against the cultural evolution of toxin consumption that produced no biological benefit.

Second, genetic natural selection continually acts to bring cultural and genetic fitness into alignment (Section 2.2 in Document S1). This is because, where a trait is transmitted culturally, as the correlation between genetic and cultural fitness becomes more positive, genetic fitness always increases. Indeed, genetic equilibrium is only possible when cultural and genetic fitness are perfectly positively correlated. Genetic evolution can tighten the coupling between cultural and genetic fitness by adjusting the social learning strategies that underpin cultural transmission. There is evidence that genetic evolution has shaped such strategies: people are discriminating about when and from whom they learn socially, in ways that appear genetically adaptive (Rendell *et al.*, 2011; Morgan *et al.*, 2012). Therefore, it seems reasonable to assume, at the very least for traits that have significant genetic fitness consequences, that cultural and genetic fitness will be positively correlated.

Our analyses do not imply that cultural evolution only favours traits that increase genetic fitness. First, many types of cultural trait may be neutral or near neutral from a genetic point of view; in such cases, the two evolutionary systems will have no impact on each other. Second, given the pace of environmental change and the fact that cultural evolution is often much faster than genetic evolution (due in part to the potential for very short cultural generations), human populations may be far from a genetic evolutionary equilibrium, so genetic selection on social learning strategies will generally be ongoing. Third, even in a system at equilibrium, we would expect some genetically deleterious traits to persist simply due to constraints on what kinds of

learning mechanisms can evolve. There may be biological, cognitive and informational constraints on the achievable precision of social learning strategies, and there will be a limit after which the marginal benefit of increasing the precision and specificity of social learning strategies will not exceed the cost.

Cultural transmission and the evolution of altruism

One of the more distinctive claims made in the cultural evolution literature is that cultural transmission facilitates the evolution of altruism in cases where genetic selection alone would not. This has been dubbed 'cultural group selection' (Fehr & Fischbacher, 2003; Henrich, 2004; Boyd & Richerson, 2010). The essence of the cultural group-selection argument is that selection for altruism at the group level overcomes selection against it at the individual level because, under cultural transmission, more of the variation resides at the group level than that is true for genetic variation (Bell *et al.*, 2009). In Document S1, we lay out, following Henrich (2004), how the selection term of the cultural Price equation can be decomposed into within-group and between-group components (Section 3 in Document S1).

Group selection can be equivalently described in terms of selection at the individual level (see Section 3 in Document S1; Allison, 1992; Henrich, 2004; Bell *et al.*, 2009; Boyd & Richerson, 2010). For altruism to be favoured by cultural natural selection, a cultural version of Hamilton's rule (Hamilton, 1964, 1972) must be satisfied; that is, the cost to the actor must be less than the benefit to his neighbours scaled by the coefficient of cultural relatedness. Cultural relatedness is a measure of how much more similar two individuals are in their cultural trait values than two randomly chosen members of the population would be (Hamilton, 1972). It should be noted that just as individuals have separate cultural fitnesses for every trait, so they have separate relatednesses, reflecting the fact that each trait can be acquired via a different pattern of social learning.

Although the cultural group selection formulation explains the evolution of altruism in terms of selection between groups outweighing selection within groups, the cultural inclusive fitness formulation describes the same evolutionary process in terms of the altruist increasing the total representation of his cultural traits in the next generation through the sum of his direct and indirect fitness. Note, the definitions of altruism are not identical under the two approaches; a behaviour may be altruistic from a group-selection perspective, yet directly beneficial from a Hamiltonian perspective if helping the group also directly increases the fitness of the focal individual (West *et al.*, 2011).

If our neighbours share our values to a greater extent than they share our genes (i.e. cultural relatedness is higher than genetic relatedness), there is greater cultural selection for cultural altruism than there is genetic selection for genetic altruism.

However, as we have shown that cultural fitness is distinct from genetic fitness, it follows that cultural altruism and genetic altruism are not the same thing. Cultural altruism is any act that reduces the actor's cultural influence while increasing the cultural influence of another individual. An example of cultural altruism would be philosopher A deciding to work all day as a farmer, at cost to his philosophical output, in order to provide food so that philosopher B could write books and articles. Philosopher A's direct cultural influence in the domain of philosophy would be reduced, but as long as philosopher B shared A's views to a sufficient degree, the net effect of A's activity might be to aid the propagation of his philosophical views indirectly via B's increased cultural fitness.

Clearly, as the above example illustrates, cultural altruism is not what people have in mind when they discuss the evolution of altruism via cultural transmission. Rather, they have in mind acts that are *genetically* altruistic, such as laying down one's life for others or redistributing resources, where cultural evolution facilitates their spread. Importantly, in many cases, a genetically altruistic trait will not be culturally altruistic. For example, dying defending a stranger from attack may be genetically altruistic, but if the dead hero's deeds or attributes were emulated, it would be culturally directly beneficial, and not culturally altruistic at all.

Our analysis of genetic altruism under cultural transmission in Document S1 (Section 3) leads to a number of conclusions. Where genetic altruism is maladaptive, that is, it does not satisfy genetic Hamilton's rule, but is directly culturally beneficial, as in the hero example just discussed, then cultural relatedness is irrelevant; cultural selection directly favours the genetic altruism, whereas genetic selection, obviously, acts against it. If cultural evolution leads to people performing genetically altruistic acts in cases where it is maladaptive, then we expect genetic selection on minds to respond by evolving psychological biases or resistances to avoid being swayed in maladaptive directions by those around us.

In the case where a behaviour is both genetically and culturally altruistic, then the different degrees of genetic and cultural relatedness become relevant. Here, cultural selection would favour altruism in a wider range of circumstances than genetic selection (eqns S22 and S23 in Document S1). There are three points to make about this. First, empirical studies to date suggest that cultural relatedness for various traits within ethnic groups, although higher than genetic relatedness, is still not very high in absolute terms (equivalently, more cultural variation is at the within-group than the between-group level; Bell *et al.*, 2009; Rzeszutek *et al.*,

2012; Henrich *et al.*, 2012; Ross *et al.*, 2013). Moreover, for some cultural traits, cultural relatedness is better predicted by geographical distance than ethnic group identity (e.g. the composition of a folk tale, Ross *et al.*, 2013). This heavily constrains the type of culturally altruistic acts that would be favoured by cultural selection. Second, because of the way culture is inherited, high cultural relatedness on one set of traits such as language, folk knowledge or verbal attitudes does not imply high cultural relatedness for behaviourally costly traits such as the propensity to perform altruistic acts. As we have stressed, in the cultural case, there is a separate relatedness for every trait, so the overall coefficient of relatedness that can be defined for genetic inheritance does not have a precise counterpart in culture.

Third, where a behaviour is both genetically and culturally altruistic and cultural relatedness is higher than genetic relatedness, there is antagonistic co-evolution between genes and culture, with genetic evolution favouring adaptations that resist paying the genetic costs of altruism to help genetically unrelated cultural brothers and sisters. Hence, as is the case for traits that are culturally directly beneficial and genetically altruistic, genetic natural selection should be expected to produce psychological resistances to either prevent the individual adopting these cultural messages or to transform in a way that neutralizes their deleterious genetic fitness consequences. Whether – and how quickly – gene win out in these antagonistic co-evolutionary situations is not clear and will depend on such factors as the relative rates of change and the size of the fitness effects. As the timescale of cultural change is so much quicker than genetic change, cultural evolution may be able to persistently keep a certain level of genetically maladaptive traits in a population despite genetic counter-selection. However, as André & Morin (2011) point out, cultural evolution will never result in genetic selection for altruism, as is sometimes claimed, unless *genetic* Hamilton's rule is also satisfied.

Discussion

We have presented a formal analysis of evolutionary change in a simple culturally transmitted trait using the Price equation as a framework. This led to definitions of key terms such as fitness, altruism and relatedness for the cultural case (Table 1), allowed us to identify the maxim and of cultural natural selection, and examine the interplay between cultural and genetic evolution. In this section, we discuss implications of the results for the conceptualization of human culture. In the subsections 'Is cultural evolution autonomous?' and 'Is cultural evolution Darwinian?', we assume that our formal analysis captures the key features of cultural change, and explore, on this assumption, what our results suggest about whether cultural evolution is an autonomous process, and whether it is a Darwinian processes. In the subsection 'Difficulties modelling cultural change with the Price equation', we examine our starting assumptions more critically and question the extent to which the key features of culture can in fact be captured using the cultural Price equation as we have set it out.

Is cultural evolution autonomous?

Does our analysis suggest cultural evolution represents an *autonomous* system? In other words, once cultural transmission is in place, does cultural evolution generally operate in an ancillary role, handmaiden of genetic adaptation, or does it break free of the influence of genetic evolution completely? Our framework suggests that the scope of cultural evolution is limited by natural selection on genes in two ways. First, an important part of cultural evolution is the transmission component – which reflects the action of minds that have been shaped by natural selection to process information in ways that enhance genetic fitness. Even small prior biases in human cognition, if iterated over many generations of learning, communicating and remembering, will have a huge impact on the content of cultural representations (Sperber, 1996; Claidière & Sperber, 2007; Claidière *et al.*, 2013). Therefore, we expect cultural change to be driven recurrently towards the (generally genetic fitness-increasing) goals of human minds. Second, although cultural fitness is a distinct quantity, if it is not aligned with genetic fitness, then there is genetic selection to change the learning rules that underpin cultural transmission, making minds more discriminating. For these reasons, cultural evolution cannot become completely autonomous. In this, we echo Lumsden & Wilson's (1981) famous conclusion that 'genetic natural selection operates in such a way as to keep culture on a leash' (p. 13).

Our conclusion is that traits spread by culture that significantly impact survival and reproduction should be expected to be, in general, genetically beneficial. Traits spread by culture that are genetically costly represent, from the point of view of adaptationist analysis of human behaviour, either evolutionary disequilibrium or constraints on the optimality of the mechanisms that can evolve. Given the rapid rate of cultural change compared with genetic change, it may well be that culture is capable of maintaining a level of genetically costly traits in a population, with genetic natural selection constantly lagging behind, but the extent of this remains to be determined. The one situation where cultural evolution is largely autonomous is where traits are genetically neutral. For example, which sounds groups of people use to denote body parts or numbers is irrelevant to genetic fitness as long as they all use the same ones. It is in such scenarios where it is most fruitful to see cultural evolution as

able to range without strong constraint from genetic evolution.

Is cultural evolution Darwinian?

A separate issue to whether cultural evolution constitutes an autonomous process is whether it is best seen as a specifically *Darwinian* process. Opinion is sharply divided on this issue (Boyd & Richerson, 1985; Sperber, 1985; Mesoudi *et al.*, 2004; Claidière & Sperber, 2007; Claidière & André, 2012). The Price equation is useful in this regard as its partition of evolutionary change into two components, selection and transmission, allows us to identify a continuum of 'Darwinian-ness' for evolving systems. At one end of the continuum, the selection component of evolutionary change dominates, and the transmission component is negligibly small. This means that fidelity of inheritance is very high, and fitness differences due to phenotypic properties are pre-eminently important in explaining change. These are key features that characterize the most paradigmatically Darwinian processes (Godfrey-Smith, 2009). In straightforward genetic cases, because selection is the predominant component of evolutionary change and transmission relatively unimportant, selective advantage coupled to blind variation is the predominant explanatory principle for the features that we see.

At the other end of the continuum is the case where the transmission component is large and the selection component negligible. Such a case would be where individuals were exposed to ideas, but deliberately or subconsciously modified them through their cognition to such an extent that the trait they exhibit bore little resemblance to the traits of the previous generation. Changes in trait frequency would then be best explained with reference to the (genetically evolved) transformative properties of the human mind, not cultural natural selection. Such a situation does not possess the usual paradigmatic features of a Darwinian process (Godfrey-Smith, 2009; although see Claidière *et al.*, 2013; for a broader conceptualization of Darwinianness that does include such cases). Thus, the disagreement between those who advocate a Darwinian or a non-Darwinian approach to cultural evolution comes down in large part to different views about the relative importance of selection versus transmission in cultural change (Sperber, 1996; Sperber & Hirschfeld, 2004).

It is clear that cultural evolution does not sit as far to the selection end of the continuum as genetic evolution does (Claidière *et al.*, 2013). Thus, the question is more one of degree; Is selection important enough relative to transmission to merit the name Darwinian? It seems likely that the balance of selection and transmission will be different for different traits. The balance is consequential because it will affect the dynamics of the system. For example, if the selection component dominates, we may expect to see many of the features of genetic evolution recur in culture: vestigial characteristics, path dependence and local optima being important. By contrast, if transmission dominates, then 'cognitive attractors' (the consequences of specific transmission biases) will explain recurrently observed patterns of culture (Claidière & Sperber, 2007). Here, there has been empirical progress, with researchers able to demonstrate experimentally that transmission biases exist and lead to the kind of outcomes recurrently observed across cultures (see Xu *et al.*, 2013, for a particularly clear recent example). This suggests that transmission effects represent an important component of cultural evolutionary change.

However, the 'Darwinian' debate is resolved, cultural change differs from genetic evolution in that both its selection and transmission properties are themselves outcomes of a Darwinian process with greater primacy – the genetic evolution of the human mind – and so cultural evolutionary explanations always tend to raise a deeper genetic evolution question: why would human beings be so constituted as to learn like that? (see Nettle, 2009; Dickins & Rahman, 2012).

Difficulties modelling cultural change with the Price equation

Our modelling framework characterized the change in a cultural trait using the framework of selective change in biology provided by the Price equation. In forcing the cultural case into this framework, a number of important differences from biological cases become evident, above and beyond the fact already mentioned that transmission is likely to be much more important relative to selection for culture than for genes.

We were forced to recognize that an individual does not have a single cultural fitness, but arbitrarily many, each corresponding to a different domain of socially learned behaviour. Correspondingly, any two individuals have a different cultural relatedness for each domain. By contrast, in the standard organismal case, the effect of common ancestry is to make genetic relatedness more or less the same across much of the genome. It is this genetic shared interest that makes possible complex social adaptations such as those for altruism (West & Gardner, 2013). For culture, then, we would ideally need to model evolution using a vector of cultural fitnesses and a vector of relatednesses, one fitness and one relatedness for each domain. This not only creates much greater complexity than any biological case, but raises profound problems of how the domains are to be identified and delimited from one another.

These issues also relate to the difficulties arising from the definition of a cultural trait, z. The Price equation approach is only valid if trait values can be associated with a single measure of fitness. In genetic evolution, uniform inheritance means all genes on a genome have

the same expected fitness (Price, 1970; Grafen, 1985; Gardner *et al.*, 2011). However, in culture, different traits have different patterns of inheritance; hence, different fitness values and different relatedness with other individuals. Therefore, the evolution of any idea or skill that draws upon multiple independently segregating traits cannot easily be studied as a whole except by approximation. It is hard to imagine how to isolate constituent parts of a complex idea or skill that each had a uniform inheritance and empirically quantify how they contributed to the final result, but this would be necessary if we wanted to understand their cultural evolution in detail. In genetic evolution, common fitness across the genome allows adaptations underpinned by complex heritable information to evolve. By contrast, in culture, as traits have different inheritance patterns, even if natural selection can operate, maximization of inclusive fitness may be antagonistic between traits. Thus, cultural natural selection will be less effective as generating the appearance of design that is the hallmark of Darwinian adaptation, particularly for social traits.

This leads to the related problem of how to identify generations and instances of inheritance. In biology, it is usually fairly straightforward to identify when reproduction has occurred and tell which is the ancestral individual and which the descendant. For the cultural case, we were forced to allow the set of ancestors to be variable in number, but also heterogeneous in time, so that both a peer and a long-dead authority who left behind writings could both be in the same ancestral generation relative to a particular learner. Clearly, this would cause extreme difficulty studying the dynamics of cultural change both for theoreticians and for empiricists. In addition, a generation is just the time it takes for someone to change their phenotype through social learning, allowing for many generations within a human lifetime, and the possibility of individuals featuring at multiple different points in their own pedigrees. This allows for temporal dynamics quite different from those of genetic evolution (Strimling *et al.*, 2009).

This links to perhaps the greatest difficulty with trying to study cultural change using evolutionary theory, which is the nature of trait inheritance. In biology, it is sensible to isolate a trait such as body size, exhaustively catalogue the parents of an individual and measure heritability using the covariance between the trait in the parents and the same trait in the offspring. Consider by contrast a cultural case like musical style. A composer's style might be influenced by the music of his predecessors, and the degree of similarity might be measurable. However, his musical style might also be influenced by the violent imagery of poets, political world events or even the changing economic landscape in his city. Thus, poets, politicians and local businessmen might become his musical cultural parents, without ever making any music at all. Not only would this make it very difficult to catalogue the complete set of his musical cultural parents, but heritability, defined in terms of phenotypic covariance between parents and offspring, could not be computed in any simple way, because in many cases, it would be a *different* trait in the ancestors that influenced the trait in the descendant.

An alternative approach avoiding some of this unwelcome complexity would be that of Helantera & Uller (2010), who take ancestor–descendant links to exist only between genetic parents and offspring. Horizontal and oblique transmission of cultural traits is then treated within the transmission term of the Price equation. This maintains the simplicity of what a generation is, how heritability can be measured, and when an ancestor–descendant link exists. However, to define ancestor–descendant links in this way only seems justified if biological parents are especially important in the transmission of culture relative to other community members. If social learning is not predominantly from biological parents, then it is not desirable to define them as the only parents for cultural traits. Moreover, the Helantera & Uller (2010) approach does not allow the study of change within one human lifetime, as it assumes that replication happens once per biological generation.

Our formal analysis shows that it is theoretically possible to define precise expressions for terms such as fitness and trait for a simplified cultural case. However, the great success of evolutionary theory in biology is that it is possible for biologists to study adaptations without knowing the mechanistic details of how genetic inheritance works. This is because the modes of reproduction are simple and DNA replication exerts little by way of transmission effects, so the simplifying assumptions of these models contain are, in general, reasonable reflections of reality (Grafen, 1984; Claidière & André, 2012). Unfortunately, this approach cannot apply to cultural evolution. To make testable models of cultural evolution, it will be necessary to know about the specific mechanistic details of each evolutionary scenario, since fidelity in inheritance is variable, transmission biases are important and domain-specific, the modes of reproduction multiple and shifting, and inheritance multifarious. Hence, although cultural evolutionary models are definable, the prospect of developing a few simple models that offer very general predictions seems poor. Lest this seem a depressing conclusion, there are areas where progress can be made, such as the empirical study of transmission biases and the psychology underlying them, and epidemiological models of particular cases where the dynamics of transmission can easily be characterized.

Conclusion

The formal framework presented in this paper highlights the variety of challenges for modelling cultural evolution, many of which have already been debated verbally within the cultural evolution literature (see

Mesoudi *et al.*, 2006; for discussion). Despite our large caveats in Section 'Difficulties modelling cultural change with the Price equation', we nonetheless see the framework as making a positive contribution if only to help promote conceptual and semantic clarity. In the cultural evolution literature, for example, transmission effects from iterated learning are sometimes labelled verbally as 'selective pressures on culture', when nothing like natural selection is involved. Papers often conflate cultural adaptation (cultural patterns becoming better able to transmit themselves) with genetically adaptive consequences (humans being able to increase their numbers or colonize novel environments), or cultural altruism (helping others become culturally influential at a cost to one's own cultural influence) with genetic altruism (reducing one's reproductive success to the benefit of the reproductive success of others). Formalization highlights the need to keep these concepts distinct. It also helps clarify, at least in principle, which of the putative consequences of cultural transmission are genetically adaptive, which are culturally adaptive, and in which cases cultural and genetic selection operate antagonistically, although we recognize that profound – and perhaps inherently unresolvable – difficulties remain in creating a general evolutionary theory of cultural change.

References

Allison, P. 1992. The cultural evolution of beneficent norms. *Soc. Forces* **71**: 279–301.

Anderson, E.S. 2004. Population thinking, Price's equation and the analysis of economic evolution. *Evol. Inst. Econ Rev.* **1**: 127–148.

André, J.B. & Morin, O. 2011. Questioning the cultural evolution of altruism. *J. Evol. Biol.* **24**: 2531–2542.

Atran, S. 2001. The trouble with memes – inference versus imitation in cultural creation. *Hum. Nat.* **12**: 351–381.

Bell, A.V., Richerson, P.J. & McElreath, R. 2009. Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc. Natl. Acad. Sci.* **106**: 17671–17674.

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. *Philos. Trans. R. Soc. B-Biol. Sci.* **365**: 3787–3795.

Boyd, R., Richerson, P.J. & Henrich, J. 2011. Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behav. Ecol. Sociobiol.* **65**: 431–444.

Cavalli-Sforza, L.L. & Feldman, M.W. 1981. *Cultural Transmission and Evolution*. Princeton University Press, Princeton.

Claidière, N. & André, J.B. 2012. The transmission of genes and culture: a questionable analogy. *Evol. Biol.* **39**: 12–24.

Claidière, N. & Sperber, D. 2007. The role of attraction in cultural evolution. *J. Cogn. Cult.* **7**: 89–111.

Claidière, N., Scott-Phillips, T. & Sperber, D. 2013. How Darwinian is culture? *Philos. Trans. R. Soc. B-Biol. Sci.* in press.

Collins, S. & Gardner, A. 2009. Integrating physiological, ecological and evolutionary change: a Price equation approach. *Ecol. Lett.* **12**: 744–757.

Day, T. & Bonduriansky, R. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. *Am. Nat.* **178**: E18–E36.

Dickins, T.E. & Rahman, Q. 2012. The extended evolutionary synthesis and the role of soft inheritance in evolution. *Proc. Biol. Sci.* **279**: 2913–2921.

Fehr, E. & Fischbacher, E. 2003. The nature of human altruism. *Nature* **425**: 785.

Frank, S.A. 1997. The Price Equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**: 1712–1729.

Gardner, A. 2008. The Price equation. *Curr. Biol.* **18**: R198–R202.

Gardner, A., West, S.A. & Wild, G. 2011. The genetical theory of kin selection. *J. Evol. Biol.* **24**: 1020–1043.

Godfrey-Smith, P. 2009. *Darwinian Populations and Natural Selection*. Oxford University Press, Oxford.

Godfrey-Smith, P. 2012. Darwinism and cultural change. *Philos. Trans. R. Soc. B-Biol. Sci.* **367**: 2160–2170.

Grafen, A. 1984. Natural selection, kin selection and group selection. In *Chapter 3 of Behavioural Ecology*, 2nd edn (J.R. Krebs & N.B. Davies, eds), pp. 62–84. Blackwell Scientific Publications, Oxford.

Grafen, A. 1985. A geometric view of relatedness. *Oxford Surv. Evol. Biol.* **2**: 28–90.

Grafen, A. 2002. A first formal link between the Price equation and an optimization program. *J. Theor. Biol.* **217**: 75–91.

Grafen, A. 2006. Optimization of inclusive fitness. *J. Theor. Biol.* **238**: 541–563.

Griffiths, T.L., Kalish, M.L. & Lewandowsky, S. 2008. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philos. Trans. R. Soc. B-Biol. Sci.* **363**: 3503–3514.

Hamilton, W.D. 1964. The genetical evolution of social behaviour, I and II. *J. Theor. Biol.* **7**(1–16): 17–52.

Hamilton, W.D. 1972. Altruism and related phenomena, mainly in the social insects. *Annu. Rev. Ecol. Syst.* **3**: 192–232.

Helantera, H. & Uller, T. 2010. The Price equation and extended inheritance. *Philos. Theory Biol.* **2**: e101.

Henrich, J. 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. *J. Econ. Behav. Organ.* **53**: 3–35.

Henrich, J., Boyd, R., McElreath, R., Gurven, M., Richerson, P.J., Ensminger, J. *et al.* 2012. Culture does account for variation in game behavior. *Proc. Natl. Acad. Sci.* **109**: E32–E33.

Heywood, J.S. 2005. An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection. *Evolution* **59**: 2287–2298.

Irons, W. (1979) Cultural and biological success. In: *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective* (N. A. Chagnon & W. Irons, eds). pp. 257–272. Duxbury, North Sciutate.

Lehmann, L., Feldman, M.W. & Foster, K.R. 2008. Cultural transmission can inhibit the evolution of altruistic helping. *Am. Nat.* **172**: 12–24.

Lumsden, C.J. & Wilson, E.O. 1981. *Genes, Mind, and Culture: The Coevolutionary Process*. Harvard University Press, Cambridge, MA.

Mesoudi, A., Whiten, A. & Laland, K.N. 2004. Is human cultural evolution Darwinian? Evidence reviewed from the perspective of The Origin of Species. *Evolution* **58**: 1–11.
Mesoudi, A., Whiten, A. & Laland, K. N. 2006. Towards a unified science of cultural evolution. *Behav. Brain Sci.* **29**: 329–346.
Molleman, L., Quiñones, A.E. & Weissing, F.J. 2013. Cultural evolution of cooperation: the interplay between forms of social learning and group selection. *Evol. Hum. Behav.* **34**: 342–349.
Morgan, T.J.H., Rendell, L.E., Ehn, M., Hoppitt, W. & Laland, K.N. 2012. The evolutionary basis of human social learning. *Proc. Biol. Sci.* **279**: 653–662.
Nettle, D. 2009. Beyond nature versus culture: cultural variation as an evolved characteristic. *J. Roy. Anthropol. Inst.* **15**: 223–240.
Nettle, D. & Pollet, T.V. 2008. Natural selection on male wealth in humans. *Am. Nat.* **172**: 658–666.
Okasha, S. 2006. *Evolution and the Levels of Selection*. Oxford University Press, Oxford.
Price, G.R. 1970. Selection and covariance. *Nature* **227**: 520–521.
Price, G.R. 1972. Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**: 485–490.
Queller, D.C. 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* **139**: 540–558.
Queller, D.C. 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. *Proc. Natl. Acad. Sci.* **108**: 10792–10799.
Rendell, L., Fogarty, L., Hoppitt, W.J.E., Morgan, T.J.H., Webster, M.M. & Laland, K.N. 2011. Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends Cogn. Sci.* **15**: 68–76.
Richerson, P.J. & Boyd, R. 2005. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago University Press, Chicago.
Rogers, D.S. & Ehrlich, P.R. 2008. Natural selection and cultural rates of change. *Proc. Natl. Acad. Sci.* **105**: 3416–3420.
Ross, R. M., Greenhill, S. J. & Atkinson, Q. D. 2013. Population structure and cultural geography of a folktale in Europe. *Proc. Biol. Sci.* **280**: 20123065.
Rzeszutek, T., Savage, P.E. & Brown, S. 2012. The structure of cross-cultural musical diversity. *Proc. Biol. Sci.* **279**: 1606–1612.
Sperber, D. 1985. Anthropology and psychology: towards an epidemiology of representations. *Man* **20**: 73–89.
Sperber, D. 1996. *Explaining Culture: A Naturalistic Approach*. Blackwell, Oxford.
Sperber, D. & Hirschfeld, L.A. 2004. The cognitive foundations of cultural stability and diversity. *Trends Cogn. Sci.* **8**: 40–46.
Strimling, P., Enquist, M. & Eriksson, K. 2009. Repeated learning makes cultural evolution unique. *Proc. Natl. Acad. Sci.* **106**: 13870–13874.
Taylor, P. 2009. Decompositions of Price's formula in an inhomogeneous population structure. *J. Evol. Biol.* **22**: 201–213.
West, S.A. & Gardner, A. 2013. Adaptation and Inclusive Fitness. *Curr. Biol.* **23**: R577–R584.
West, S.A., El Mouden, C. & Gardner, A. 2011. 16 common misconceptions about the evolution of cooperation in humans. *Evol. Hum. Behav.* **32**: 231–262.
Xu, J., Dowman, M. & Griffiths, T. L. 2013. Cultural transmission results in convergence towards colour term universals. *Proc. Biol. Sci.* **280**: 20123073.

Supporting information

Additional Supporting Information may be found in the online version of this article:
Document S1 Formal analysis and results.

Received 17 April 2013; revised 7 November 2013; accepted 9 November 2013

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 Υ_{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:

$$\forall_j, \sum_{i \in A_j} \Upsilon_{ij} = 1 \tag{S1}$$

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.*

$$c_i = \sum_{j \in D_i} \Upsilon_{ij} \tag{S2}$$

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:

$$\sum_{i=1}^{N} \sum_{j \in D_i} \Upsilon_{ij} = \sum_{j=1}^{N} \sum_{i \in A_j} \Upsilon_{ij} = N \tag{S3}$$

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 Δz_j. Thus:

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

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

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

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

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

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:

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

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 ϒ 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).

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

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 $f_1, f_2 \ldots f_n$, 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:

$$c_i = c_0 + \beta(c, f_1)f_{1_i} + \beta(c, f_2)f_{2_i} + \cdots + \beta(c, f_n)f_{3_i} + \varepsilon_i \quad \text{(S10)}$$

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

$$cov(c, z) = cov(c_0 + \beta(c, f_1)f_{1_i} + \beta(c, f_2)f_{2_i} + \cdots + \beta(c, f_n)f_{n_i} + \varepsilon_i, z)$$
$$= \beta(c, f_1)cov(f_1, z) + \beta(c, f_2)cov(f_2, z) + \cdots + \beta(c, f_n)cov(f_n, z) \quad \text{(S11)}$$

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 βs 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:

$$\Delta_g \overline{w} = cov(w, w) = var(w) \qquad \text{(S12)}$$

Here, *w* is absolute, rather than relative, genetic fitness, related to relative genetic fitness by $v = \frac{w}{\overline{w}}$. 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:

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

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:

$$\Delta_c \overline{w} = \beta(c, w) \Delta_g \overline{w} \qquad \text{(S14)}$$

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:

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

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 δ 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 though 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):

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

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:

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

- 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).

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

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 ε is a residual. The βs are the partial regression coefficients of cultural fitness on the level of the focal's own altruistic values, holding neighbours' values constant $\beta\left(c, a_f\right)$, 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:

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

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:

$$cov(c,a) > 0$$
$$\beta(c,a_f) + \beta(a_n,a_f) \cdot \beta(c,a_n) > 0$$
$$\beta(a_n,a_f) \cdot \beta(c,a_n) > -\beta(c,a_f) \quad \text{(S21)}$$

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:

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

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:

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

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:

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

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

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

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):

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

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.

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

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 beneficient 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.