

Cultural Attractor Theory and Explanation

Andrew Buskell*

Cultural attractor theory (CAT) is a highly visible and audacious approach to studying human cultural evolution. However, the explanatory aims and some central explanatory concepts of CAT remain unclear. Here I remedy these problems. I provide a reconstruction of CAT that recasts it as a theory of forces. I then demonstrate how this reinterpretation of CAT has the resources to generate both cultural distribution and evolvability explanations. I conclude by examining the potential benefits and drawbacks of this reconstruction.

Keywords

Cultural Evolution • Cultural Attractor Theory • Explanation

1 Introduction

Over the past forty years, researchers have constructed sophisticated techniques and tools for understanding the structure and dynamics of cultural change (e.g. Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Sperber 1996; Odling-Smee et al. 2003). Cultural evolutionary research is now what Hacking (1995) would call ‘cutting-edge’—with research institutions, international meetings, and a dedicated journal in the works. Cultural Attractor Theory (CAT) is one approach within this field of cultural evolution research. Introduced by anthropologist Dan Sperber (1996), CAT is now a multi-disciplinary approach with on-going work in primatology (Claidière, Smith, et al. 2014), psychology (Miton et al. 2015), sociology (Morin 2013, 2015), and theoretical modelling (Claidière and Sperber 2007; Claidière, Scott-Philips, et al. 2014).

Discussions of CAT have tended to focus on its account of learning. This is for good reason. CAT’s account of learning is a vivid contrast to the idealised depictions of learning at work in many cultural evolutionary models (e.g. Boyd and Richerson 1985; Henrich and Boyd 2002). While these models take learning to be replicative—where learners copy from models—CAT takes learning to be *reconstructive*, relying on inferential capacities (Sperber 1996). But such discussions rarely take into account the aspirations of those working within CAT. Indeed, CAT claims to be “a generalization of [cultural evolutionary] approaches, [one] that provides novel

*Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge, UK, CB2 3RH, ab2086@cam.ac.uk

tools” (Scott-Phillips and Sperber 2015, 45). We are similarly told that CAT may even explain the processes that “cause cultural items to tendentially evolve towards greater design” (Claidière, Scott-Phillips, et al. 2014). And to be sure, CAT has built up a number of interesting conceptual tools and models (Sperber 1996; Claidière and Sperber 2007; Claidière, Scott-Phillips, et al. 2014; Morin 2015).

However, despite its aspirations, CAT is often unclear about its explanatory aims and central explanatory concepts. Further, commentators have highlighted how CAT employs competing characterisations of its key explanatory concepts (Acerbi and Mesoudi 2015; Buskell 2017). Here I remedy both these problems. Below I offer a reconstruction of CAT that takes it to be a theory of forces. On my reconstruction, a *force of attraction* provides the resources to make progress on two explanatory projects: *distribution explanations* that account for the frequency of cultural variants within a population, and *cultural evolvability explanations* that account for why cultural variants of a particular form were more likely to evolve.

This reconstructive project is normative. Though I note areas of disagreement with the CAT literature, an accurate portrayal of cultural epidemiological viewpoints is not my aim. Instead I attempt to reconcile CAT with the explanatory agenda of cultural evolutionary theory. The reconstruction on offer is thus how cultural epidemiologists *should* employ CAT when engaged in the broad enterprise of modelling and explaining cultural evolution.

This reconstruction has numerous positive features. It highlights methods and strategies that cultural epidemiologists can put to work, articulates how one should understand the ‘cultural attractor’ concept, and shows how CAT can be separated from Sperber’s (1996) modular account of cognition. Yet there are also some negative consequences. Most notably, the reconstruction on offer is much humbler about the explanatory scope of CAT. Accepting it might mean backing off on audacious claims about the generality and power of CAT.

2 Cultural Attractor Theory

2.1 Reconstructive Learning

Work in cultural evolution pursues several explanatory aims. These include accounts of human cognitive evolution (Boyd and Richerson 1985; Henrich 2016); the dynamics of cultural transmission (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Odling-Smee et al. 2003; Henrich 2004; Powell et al. 2009); how cultural traditions change over time (Sperber 1996; Morin 2015); and how such traditions themselves may spread and go extinct across distinct populations (Mace and Holden 2005; Currie and Mace 2014).

Within the cultural evolutionary literature, CAT is distinguished by its account of human learning. As I have said above, the dominant account of learning used in cultural evolutionary models is *replicative* (e.g. Boyd and Richerson 1985; Odling-Smee et al. 2003; cf. Henrich and Boyd 2002; Henrich 2004). This replicative picture is one where learners reliably acquire whatever behaviours, skills, or pieces of know-how (what I’ll call ‘cultural variants’) are currently of interest. Though this learning process might be error-prone and noisy—particularly when what is being learned is complex or when teachers are uncooperative—multiple exposures to learning opportunities can lead to the reliable replication of cultural variants over time.

Yet as CAT theorists note, the replicative picture is metaphorical, and should not be taken too seriously (Sperber 2000). Individuals cannot actually copy things like beliefs or skills, because learners do not have access to beliefs or skills; all that learners have access to are the ways that beliefs or skills are manifest in behaviour. So copying is a misleading picture of the nuts-and-bolts of how learning actually occurs. A more realistic picture takes learning to be

reconstructive (Sperber 1996, 2000). On a reconstructive picture, when individuals learn from one another, they construct a variant of their own on the basis of observable evidence. This is what I have elsewhere called *reconstructive learning* (Buskell 2017).

To be sure, there are many possible causes of the states of affairs that learners observe. When an individual is attempting to learn a particular behaviour, say, how to carry out long-division, then they will have to piece together evidence from a wide variety of sources: textbooks, examples, and whatever help comes from teaching or instruction. Integrating all these sources of evidence involves the learner's inferential capacities. Learning is *inferential* rather than imitative or replicative.

Unsurprisingly, these inferences can be sensitive to a range of local factors. Learning is not only influenced by background beliefs, but also variations in an individual's affective, cognitive, and axiological state (Sperber 1996). One area where such differences are acute is in politics. Here, local influences can cause the same piece of information to be reconstructed in wildly different ways. Consider a news bulletin announcing the privatisation of public resources by the ruling political party. If I have antecedent beliefs about the effectiveness of the ruling political party, I might understand or integrate such news positively. On the other hand, if I have antecedent beliefs that the ruling political party is acting in their own self-interest, I am likely to view such privatisation more sceptically.

A clear articulation of such inferential *biases* on reconstructive learning is given by Claidière and Sperber who write:

When an individual acquires a new cultural item (e.g. a skill, a belief, or a norm), she never just copies the variant or variants she observes; rather, drawing on the information transmitted and her own background knowledge, inferential abilities, and interests, she constructs a variant of her own. This variant is likely to depart from the variants on which it is based. (2007, 91–92)

Reconstructive learning is “influenced by past inputs” (Sperber 1996, 118)—past inputs that might lead one to be a partisan politico, for instance—and these past influences can generate biases in learning. Inferential biases, as I will use the term, consist in the way that beliefs and values, as well as one's cognitive and affective state, impact and influence the inferences made by individuals during learning.

The causal factors that generate such inferential biases are what cultural epidemiologists call *factors of attraction*. In other work (Buskell 2017), I have shown that cultural epidemiologists appeal to distinct kinds of factors of attraction: factors that bias inferential capacities during learning (inferential biases); factors that motivate the expression of certain cultural variants (motivational factors); and factors that impact the accessibility and availability of information and evidence (ecological factors).

Below I suggest that the central causal-explanatory factors CAT should appeal to are those that bias inferential capacities. As I go on to show, such inferential capacities can be construed as a force impinging upon an otherwise unbiased population of learners. This construal of CAT shows how it can contribute to standard cultural evolutionary theory, and in doing so, improve the realism of their models.

2.2 Explanatory Payoffs

I suggest that CAT can gain explanatory traction on phenomena by adopting what Lewens (2015) calls the *kinetic* approach to culture. The kinetic approach is aggregative, modelling the downstream outcome of interactions among homogeneously construed individuals. As I

hinted at above, in the real world, individuals' inferential capacities are influenced by a number of factors of attraction—some held widely across many human populations, others more locally. Yet to generate tractable models, CAT researchers assume homogeneous populations (or sets of sub-populations) where individuals share sets of inferential biases. Assuming such homogeneity, cultural epidemiologists can model and predict the way in which biases generate populational changes.

The explanatory concept used by cultural epidemiologists to describe and ultimately explain downstream population changes is what they call *cultural attractors*. Here is Sperber's characterisation of the concept:

[An] attractor, as I have characterized it, is an abstract, statistical construct, like a mutation rate or a transformation probability. To say that there is an attractor is just to say that, in a given space of possibilities, transformation probabilities form a certain pattern: they tend to be biased so as to favour transformations in the direction of some specific point, and therefore cluster at and around that point. (1996, 111–112)

Cultural attractors are characterisations of systematic transformation biases within a population that influence how individuals learn. This much is straightforward. But what does Sperber mean when he calls such attractors 'abstract statistical constructs'? I suggest we take the analogy between transformations and mutations at face value.

In toy models of population genetics, a mutation rate can be modelled as a fixed frequency at which a particular allele mutates into another. Consider an asexual haploid population with two non-fatal alleles, A and a . Further, assume that A mutates into a at some rate μ every generation. Even under situations of strong selection, where the selection coefficient $s \gg \mu$, there will always be a proportion of the population μ that has the allele a .

An analogous situation holds for CAT. Just as researchers in population genetics model the rate of one allele mutating into another, so too, I suggest, are cultural epidemiologists modelling the rate of transformation from one cultural variant into another: cultural variant A is transformed into a distinct cultural variant a at a particular rate. More than this however, cultural epidemiologists hold that cultural variants have *multiple* transformation probabilities. The cultural variant A not only has a transformation rate to transform into a , but distinct rates for transforming into a' , or a'' . It is the *aggregate* of all various transformation rates that Sperber points to when he calls cultural attractors an abstract statistical construct. Cultural attractors, as I construe them, are an aggregate of transformation rates for a set of cultural variants.

With this characterisation on the table, we can ask how it is that such aggregate transformation rates can explain downstream population outcomes. Here we face some controversy. According to Robert Boyd and Peter Richerson, there are two different characterisations of transformations appealed to by cultural epidemiologists. On the one hand, cultural epidemiologists appeal to "systematic transformations [that] create a new nonselective force that moves the population towards nearby attractors" (Boyd and Richerson 2005, 84). On the other hand, so they allege, cultural epidemiologists write as if "the transformations that occur during social learning are unsystematic, so that people observing the same behaviour infer very different cultural variants; consequently, cultural replication is so noisy and inaccurate that weak selective forces would be swamped" (Boyd and Richerson 2005, 84). According to Boyd and Richerson, reconstructive learning is either directional and systematic or noisy and inaccurate.

While accuracy of this characterisation would no doubt be challenged by cultural epidemiologists, I think it is correct suggesting that CAT can be interpreted so as to make progress on two distinct explanatory endeavours. In what follows I'll suggest that these two distinct kinds of

explanatory endeavours recruit the concept of cultural attractors in different ways. First, cultural attractors can explain systematic changes in the frequency distribution of variants and explain insofar as they account for how the transformations made during learning lead to some variant(s) being acquired rather than others. This strategy provides what Godfrey-Smith (2009) calls distribution explanations.

But there is another way in which cultural attractors are recruited in explanations. This occurs when one focuses on how variants are “transformed or distorted to become more similar to a particular form” (Mesoudi and Whiten 2008, 3492). That is, when cultural attractors are employed as part of an explanation for why certain variants have the form that they do. Here, cultural attractors are put forward as explanations for why variants of a particular form are more likely to evolve than others. I call these latter kinds of explanations *cultural evolvability explanations*.

3 Cultural Attractors and Distribution Explanations

3.1 Cultural Attractors

As I’ve suggested, one of CAT’s explanatory aims is to provide *distribution explanations*. Broadly, distribution explanations are those that articulate why a distribution of variants at some initial time has the same or different distribution at some later time. So, for instance, a distribution explanation for different brands of ice-skates would explain why the proportion of skaters wearing ice-skate brands $[x, y, z]$ at distribution $[20\%, 55\%, 25\%]$ at time t_1 changes to a distribution of $[60\%, 30\%, 10\%]$ at time t_2 .

Unfortunately, evaluating CAT’s success in providing distribution explanations has been hampered by the multiple accounts of the cultural attractor concept. As discussed above, the cultural attractor concept is statistical. And though I have put forward my own interpretation of this concept above, it is important to note that the CAT literature contains several such interpretations. In other work (Buskell 2017), I’ve noted that there are four distinct characterisations of this cultural attractor concept. Here I build on this past work and show how the conceptual ambiguity of the cultural attractor concept has explanatory consequences.

We can group the distinct characterisations of the cultural attractor concept into two broad categories: kinetic characterisations and distributional characterisations. Kinetic characterisations follow the strategy discussed above, where one aggregates the effect of a homogeneously construed population of individuals. Distributional characterisations, on the other hand, identify attractors with particular distributional changes. Here I argue that CAT *should* employ kinetic characterisations—what I call TRANSFORMATION PATTERN¹—to the exclusion of the others.

TRANSFORMATION PATTERN should be familiar. It is the characterisation of cultural attractors I have reconstructed above. On this interpretation, cultural attractors are identified with aggregates of transformation rates (Sperber 1996, 2000; Sperber and Claidière 2008). Under simplifying assumptions, this aggregate can be identified with shared inferential biases that lead to systematic population-level effects.

CLUSTER POINT, INCREASING FREQUENCY, and EQUILIBRIUM POINT, on the other hand, identify attractors with distinct distribution outcomes. CLUSTER POINT (Sperber 1996, 108–112; Sperber and Claidière 2008, 92) identifies cultural attractors with distributions that change over time to cluster around a particular variant. INCREASING FREQUENCY (Claidière, Scott-Phillips, et al. 2014; Scott-Phillips and Sperber 2015) identifies cultural attractors with distributions where some variant increases in relative frequency over time. Finally, EQUILIBRIUM FREQUENCY

¹Following standard convention, I refer to the notion, or concept, ‘xyz’ by capitalising it to XYZ.

(Claidière, Scott-Phillips, et al. 2014) identifies cultural attractors with the distribution that occurs at some equilibrium point.

I will have more to say about the explanatory merits of these various characterisations of the cultural attractor concept in a later section ('Circularity and Pan-attractivism'). However, before moving on, it is important to emphasise a crucial point. Distributional characterisations of the cultural attractor concept (CLUSTER POINT, INCREASING FREQUENCY, EQUILIBRIUM POINT) identify attractors with subsets of distribution outcomes. For instance, CLUSTER POINT identifies cultural attractors with distributions that change to cluster around a single variant. But why should one restrict cultural attractors, such that they can only be used to explain a restricted subset of possible outcomes?

Though it might be tempting to assume that such a restriction allows greater scope for predicting or explaining distributions, I am not convinced. Identifying attractors with distribution outcomes seems to rob the concept of any descriptive or explanatory power: attractors just end up being whatever distribution actually occurs.

My hunch is that such distributional characterisations are best interpreted as slips of the tongue—a mistaken identification of attractors with outcomes rather than the causal processes that generate such outcomes. The slip between the two is easy to make, particularly if one assumes that aggregate transformation rates can be approximated by smooth functions, with increasing likelihoods of adopting a particular variant clustering around a particular attractor 'point' or 'peak' (e.g. Sperber 1996; Claidière and Sperber 2007). With such an assumption, transformation rates work to generate distributions of variants with equilibrium distributions that cluster around a central variant that increases in relative frequency.

But this neat and tidy picture is not a conceptual necessity. Transformation rates can be irregularly distributed across many cultural variants, and inferential biases may cause 'leaps' from variants across 'culture-space'. Once one notices that transformation rates can involve variable and jumpy reconstructions, the neat picture articulated above loses its appeal. It is also at this point where the contrast between the kinetic and distributional characterisations of the cultural attractor concept becomes clear: while kinetic characterisations explain how distributions are generated by aggregations of transformation rates, distributional characterisations are identifications of attractors with some mere subset of distributions.

If this analysis is right, then TRANSFORMATION PATTERN is not only the most inclusive of the cultural attractor characterisations, but also the most explanatory. It can capture the wide variety of possible population outcomes, including those with distributions captured by clustering, increases of relative frequency, and stable equilibria. As a result, I suggest that cultural epidemiologists should reject definitions of cultural attractors that identify the concept with an overly restricted subset of distribution outcomes. To the extent that CAT generates distribution explanations, cultural epidemiologists should understand cultural attractors as a description of shared, aggregate transformation rates.

3.2 *The Force of Attraction*

Evolutionary theory is often conceived of as a theory of forces (Sober 1993; Stephens 2004, 2010). Such forces drive changes in populational properties in line with the force's strength (or magnitude) and direction (Sober 1993; Stephens 2010). In evolutionary theory, these forces (at least selection, mutation, and migration) can be additive, having a stronger net effect when acting in concert. Or such forces may counteract one another, leaving the distribution of variants in line with whatever 'zero-force state' is operative for a particular population.

Understanding evolutionary theory as a theory of forces has some explanatory payoffs.² Notably, one can appeal to the direction and magnitude of combined forces to explain why a population could come to have certain distributions of variants. Here I will argue that cultural evolution can also be understood as a theory of forces, and that understanding it in this way demonstrates how cultural attractors can generate distribution explanations.

For Sober (1993) a theory of forces is only conceivable if one is able to determine how a system will act in the absence of forces acting upon it. For single-locus, two-allele organisms with sexual reproduction, this zero-force state is captured by the Hardy-Weinberg equations.³ The most important feature of such a zero-force state is that once “genotypes have gone to their Hardy-Weinberg equilibrium values, there they stay, in the absence of an evolutionary force” (Sober 1993, 34). Of course, the converse of the claim is not true—one cannot assume from the fact that a population is at Hardy-Weinberg equilibrium that there are no evolutionary forces at work. There may be forces whose net effect cancels each other out.

This has relevance for the discussion of the force of attraction and cultural evolutionary theory. For one requires an analogous articulation of a zero-force state within cultural evolution. This I find in what Peter Richerson, Robert Boyd, and Joseph Henrich call *unbiased learning* (Boyd and Richerson 1985; Henrich 2001). Unbiased learning occurs when individuals reliably acquire cultural variants from a random member of their population. This is usually cashed out in a generational model, where individuals of a new incoming generation learn from, and then replace, individuals in a previous generation.⁴ Under such conditions, the incoming generation reliably replicates the previous generation’s distribution of cultural variants.

Actually, the list of idealisations required to generate unbiased learning is somewhat more onerous than mere random copying in non-overlapping generations. To generate unbiased learning one has to assume that cultural variants are exclusive (that is, that one can only hold one variant at a time, akin to alleles at a locus in haploid organisms); that there is no assortative learning or biased sampling of particular individuals; that there is no population structure that would make learning non-random; that there are no frequency-dependent effects; and most importantly for current concerns, that individuals reliably acquire the variants of their models without distortion or deviation. When all of these idealisations are satisfied (and, perhaps, more that have yet to be identified), then subsequent generations will display the characteristic static distributions of unbiased learning.

The force of attraction can impinge on populations characterised by such unbiased learning, and in doing so, explain systematic changes in population-level distributions. To take one example, imagine a scenario where incoming members of a high-school ice-hockey team have to purchase their equipment (hockey sticks, helmets, ice-skates, etc.). Let us assume that this school is highly competitive so that only final-year students make the team (or, in other words, that there are non-overlapping generations). Further, let us assume that incoming members of the team learn from the outgoing members what brands of equipment to purchase. Ima-

²This having been said, understanding evolution as a theory of forces is contentious, and many payoffs of the approach have been called into question (Matthen and Ariew 2002, 2009; Walsh et al. 2002; Lewens 2004; Walsh 2007).

³Other background assumptions need to be satisfied as well, namely, that sexual recombination is fair (i.e. that there is no ‘meiotic drive’) and that mating is non-assortative. For more on the notion of background assumptions needed in a biological theory of forces, see Stephens (2004).

⁴Unbiased learning could also characterise a non-generational model where individuals at each time step sample from other individuals in a population. However, to characterise unbiased learning in the technical sense that I use it in this paper, such individuals would have to reliably change their variant to that of their model in a way unaffected by previous instances of learning.

gine that this situation is characterised by unbiased learning, where incoming team members are randomly paired up with outgoing members and end up purchasing the same equipment as their model. In such a scenario, year after year, successive generations of hockey teams will end up purchasing the same brands of equipment—the same kinds of hockey sticks, helmets, and ice-skates and in the same proportion.⁵

But if there is a situation where individuals infer that brands with red colouration give a competitive advantage—perhaps in line with the notorious ‘red effect’ on competition (Hill and Barton 2005; though cf. Fortunato and Clauset *preprint*)—then what was unbiased becomes biased. This can lead to a directional effect, perhaps over successive generations of hockey teams, where individuals purchase more and more red gear.

Models of the force of attraction given by Claidière and Sperber (2007) make this informal example more precise. In their work Claidière and Sperber examine how individuals come to adopt one of thirty-one smoking variants, corresponding to smoking between zero and thirty (whole) cigarettes per day. Two forces influence the number of cigarettes that individuals come to smoke: the force of attraction and the force of cultural selection. The force of attraction is simply the aggregate of transformation rates, as discussed above. The force of cultural selection, as Claidière and Sperber use the term, is one of discriminate sampling, where individuals preferentially learning from particular individuals. Without rehearsing the technical details of this model, Sperber and Claidière show that these two forces can interact in complex ways, generating a wide range of variant distributions. But in one special case, where the force of attraction acts in isolation from other forces, Sperber and Claidière (2007, 94–95) note that the variants become allocated in direct proportion to the strength of the force of attraction. And this is just what one would expect: in the absence of other forces, reconstructive learning ends up shaping population-level distributions in accordance with the strength and direction of the aggregated transformation rates.

Sperber and Claidière’s model was not intended as a comprehensive nor a general characterisation of CAT. Nonetheless, I take it to be a useful example in showing how aggregates of transformation rates can factor into distribution explanations. CAT can explain downstream distributions of cultural variants in a population by showing how such distributions result from the force of attraction.

Other benefits come along for free when one understands CAT in this way. To see why, consider the way that forces are employed to explain simple physical phenomena. To explain the trajectory of a billiard ball as it rolls across the table, one need only appeal to a force acting on the ball with a certain magnitude and direction. One does not need to provide a causal interpretation of this force; it might have been caused by a cue striking the ball, but might equally have been the result of a powerful gust of wind, or the final trick in an elaborate Rube Goldberg contraption. In fact, the cause of the force can be explanatorily irrelevant—all that might matter is why the ball went *this way* with *this speed* rather than *that way* with *that speed*. In precisely the same way, cultural epidemiologists can provide distribution explanations appealing to a force of attraction.

⁵As Sober (1993) notes with regards to the zero-force law in evolutionary biology, though deviation from Hardy-Weinberg equilibrium is evidence of a force impinging on a population, the fact that a population is in Hardy-Weinberg equilibrium is not evidence of a *lack* of impinging forces. It might be the case that the activity of two forces (selection and mutation, say) are counteracting one another. In the same way, though deviation from a static distribution of cultural variants is evidence of a force impinging on a population, the fact that a population has a static distribution of cultural variants is not evidence of a lack of impinging forces. I might be that two forces (selection and attraction, say) are counteracting one another.

This is a point that is not recognised by cultural epidemiologists despite their occasional use of ‘force talk’. Yet it is important. Positing a force of attraction means that CAT theorists can explain distribution outcomes without having to characterise the complex set of causal factors that generates such a force (*contra* Sperber 1996). This saves cultural epidemiologists from needing to articulate the nitty-gritty of how inferential biases combine together to jointly change distributions of cultural variants. (I will have more to say about this below.)⁶

4 Cultural Evolvability and Cultural Attraction

So far I have suggested that cultural epidemiologists can provide distribution explanations by appealing to a force of attraction. But cultural epidemiologists are engaged in another kind of explanatory project. Consider, for instance, a recent claim made by Thom Scott-Phillips and Dan Sperber:

Much of what gets taught in dance classes are classic steps and rhythms that are highly characteristic of a given dance and that have contributed to its cultural success. As such, the form these steps take is in large part the consequence of a range of underlying factors of attraction, which are variously cognitive, biological, or environmental in nature, and which include, in particular, physical affordances of the human body that make certain steps easier to teach and learn, the (highly locally situated in time and space) repertoire of dance steps already known to the learners, pan-human or culture-specific aesthetic preferences, and of course the rhythms being danced to (rhythm and dance being a glaring example of the way in which one type of cultural item may have an impact on the frequency of another). (2015, 45)

Or, take Sperber’s example of the story of Little Red Riding Hood:

In the logical space of possible versions of a tale some versions have a better form: that is, a form seen as being without either missing or superfluous parts, easier to remember, and more attractive. The factors that make for a good form may be rooted in part in universal human psychology and in part in a local cultural context. In remembering and verbalizing the story, tellers are attracted towards the better forms. (1996, 108)

Broadly speaking, cultural epidemiologists are concerned with explaining the *stability* of culture: the way that certain kinds of cultural forms tend to remain prevalent in a population for long stretches of time (Sperber 1996; Claidière and Sperber 2007; Morin 2015). Suggesting that dancing steps or stories can have a ‘better form’ is shorthand for saying that some cultural variants have features that make them more likely to arise and to be more stable over time.

Explanations of the stability in form are not primarily distributional (though, as I show below, these two kinds of explanation interact in complex ways). Instead they are concerned with characterising cultural lineages, and explaining how the *form* of such lineages might change

⁶Of course, at the same time, this does not rule out such work. There are explanations where the impact of specific causal facts are at stake. For instance, in some circumstances, what researchers are interested in are precisely what caused the ball to roll the way it did, given some set of background conditions. In similar ways, in some situations we are interested in precisely what caused individuals to acquire a particular cultural variant, given some set of background conditions.

and stabilise over time.⁷ These explanations are what I will call *cultural evolvability explanations*.⁸

Evolvability explanations, according to Rachael Brown (2014) advert to features of populations that determine “the probability of a particular evolutionary outcome in the future (for example, adaptedness, diversity)” (560). For Brown (2014, 554), the features that ground such outcomes are varied—they may result from developmental constraints, standing diversity, phenotypic plasticity, or states of the environment. Whatever the relevant features, they jointly constitute complex dispositional properties: likelihoods of exploring evolutionary possibilities. While Brown characterises such possibilities in terms of microphysical worlds, she notes that one can interpret such possibilities in terms of trajectories through phenotypic spaces, like morphological-space or culture-space.

Evolvability explanations have a general form: “it was, ‘selection aside,’ more probable that population x would evolve the characteristic or characteristics of interest than population y ” (Brown 2014, 560). Thus, putting processes of natural selection to one side, a population has a robust dispositional property such that it is more likely to evolve a character of interest than some other population (presumably lacking such a robust dispositional property).

A similar characterisation can be given for cultural evolvability explanations. These are contrastive explanations that account for the reasons why—selective processes aside—it is more probable that a population x would culturally evolve a variant or variants of interest than population y . Here I am reading ‘selective processes’ in line with the characterisation of the ‘force of cultural selection’ discussed above (Claidière and Sperber 2007). This is a process of informational sampling. More concretely, we can think of these as discriminate sampling processes (Millstein 2002) that determine the individual(s) from whom one learns (Reisman 2005; Ramsey and de Block 2017). Thus, cultural evolvability explanations have the general form: ‘it was, processes of *discriminate informational sampling aside*, more probable that population x would evolve the characteristic or characteristics of interest than population y .’

There are many possible processes aside from selection that can serve as difference-makers. Such difference-makers explain why *this* population rather than *that* population is more likely to culturally evolve a variant of a particular form. For cultural epidemiologists, factors of attraction are the relevant difference-makers. Different factors of attraction affect the inferential capacities of populations such that they end up more likely to have a cultural tradition, or variant, of a particular form.

Stability in form emerges as one possibility in such cultural evolvability explanations. Whether because of selective pressures or cognitive constraints, certain cultural forms may be more likely to evolve and less likely to change over time. But stability isn’t everything. There are a range of possible trajectories that cultural lineages can take, and not all need stall and stabilise around a particular form. CAT’s explanatory focus on the stability of lineages seems restrictive.

On the reconstruction offered here, CAT provides evolvability explanations. This involves

⁷For Sperber (2001, 2006; Sperber and Hirschfeld 2007), lineages of cultural variants can be conceived of long chains of reconstructions and productions—reconstructions of the content of cultural variants by means of reconstructive learning, and productions or expressions of that variant in behaviour—what he calls ‘cultural cognitive causal chains’. The details of Sperber’s account need not trouble us here; suffice to say that he thinks that culture is stable enough to sort it into lineages or traditions, but labile enough to exhibit variation and change over time (Sperber 2001; Sperber and Hirschfeld 2007).

⁸Of course, as a matter of actual fact, any particular lineages may exhibit stasis in form for long periods of time. Yet stasis may be the result of distinct causal process. The presence of strong selection pressure counteracting transformations is one possible explanation. A similar explanation, but one more amenable to CAT, is that a particular lineage is already at an optimal (understood here as the most likely to be transformed towards) variant. Both scenarios generate stasis, but appeal to distinct sets of forces.

broadening their explanatory agenda: instead of focusing on explanations of stability, it instead contributes to explaining all the dynamics of cultural evolution. The distinct contribution CAT makes to such evolvability explanations is in showing how such trajectories may be both changed and constrained given the inferential structure of individuals. To put it another way, aggregate transformation probabilities feature as part of the complex disposition of populations that determines why some cultural forms are more likely to occur than others.

5 Distribution and Evolvability Explanations

It seems clear that in some situations there will be a relationship between the distribution of variants in a population and the extent to which a population explores culture space. These are situations exemplified in an example from Sperber (1996):

Consider your views on President Clinton. They are likely to be very similar to the views of many, and to have been influenced by the views of some. However, it is unlikely that you formed your own views simply by copying, or by averaging other people's views. Rather, you used your own background knowledge and preferences to put into perspective information you were given about Clinton, and to arrive by a mixture of affective reactions and inferences at your present views. (106)

Sperber is describing how the acquisition of cultural variants is not merely a function of one's inputs. What he is emphasising is the connection between inferential biases, the distribution of variants in a population, and how these can lead to downstream changes in public opinion. Below, I show how this example can be explained using either distribution or evolvability explanations. I'll then suggest how these two explanations might overlap and intersect.

First, assume a population of individuals whose beliefs about President Clinton are relatively uniform. They have, roughly, the same set of inferential biases. On the basis of this, one can explain how opinion about President Clinton is likely to be distributed. It will be distributed in line with the force of attraction. But now consider that the example might also be offering an evolvability explanation. As we've seen, cultural evolvability explanations account for why populations are more likely to explore some regions of culture-space. If one is concerned with those areas of culture-space that generate support or opposition to President Clinton, one can frame the issue as follows: Why is it that a population is more likely to support Clinton than oppose Clinton?

Here we have to shift to a comparative framework, considering two sub-populations. Let us assume that both sub-populations are exposed to the same information; that is, they have exactly the same information on President Clinton. Let us also assume with Sperber that differences in background beliefs (about public policy, say) are the only relevant difference-makers between Clinton-supporters and Clinton-detractors. The means controlling for discriminate informational sampling—for instance, that neither of the populations are selective viewers of partisan news networks. Under such conditions (and others enumerated above) the relevant differences between populations will be differences in inferential biases. That is, under such conditions, the force of attraction can explain why it is that one population occupies a Clinton-supporting area of culture-space, while the other occupies a Clinton-detracting area of culture-space.

But this example can also show how the two kinds of explanations can intersect with one another. Facts about initial distributions of variants in a population can change populational evolvability. Even if individuals have an innate psychology that disposes one towards supporting Clinton (whatever that would look like), if such individuals are raised in an environment of Clinton detractors, then they are likely to end up as Clinton detractors as well.

On the back of these reflections, one can see why distribution and cultural evolvability explanations overlap: both appeal to the structure of heritable variation in a population and how this is likely to change over time. But this appeal is made in different ways, and this generates distinct kinds of explanation. In cultural evolvability explanations, unlike distribution explanations, there is an implicit acknowledgement of a reference group, a group which differs in the set of relevant difference-makers. With distribution explanations, by contrast, we appeal to a zero-force state of a population and show how various forces can impinge upon this population drive it towards particular distributions.

6 Some Explanatory and Methodological Worries

I hope to have shown in the previous sections that CAT can be fruitfully reconstructed in such a way that it contributes to two explanatory projects. CAT can generate both distribution explanations and cultural evolvability explanations. Further, I hope to have shown that both explanatory projects should appeal to the force of attraction. With all this ground clearing out of the way, however, I want to highlight some outstanding empirical, methodological, and explanatory worries that have been, or might be, levelled at CAT and the account on offer. I show that some of these worries are pressing, while others are less so.

6.1 *Harmonisation Problems*

The first worry concerns the complexity and heterogeneity of the force of attraction. As I showed above, the force of attraction is a composite, where a mix of affective, cognitive, and doxastic factors together combine to influence reconstructive learning. One worry that might be leveraged at the account I have developed here is what Christopher Stephens (2004) calls a *harmonisation* problem: when one is attempting to describe or characterise the force of attraction, how is one to weigh the various affective, cognitive and doxastic factors? How do they add up? Let us call the problem of ‘adding up’ the various affects on reconstructive learning the problem of *intra-force harmonisation*.

What does this intra-force harmonisation problem actually problematise? The first thing to note is that this is not a problem for general formalisations or models like those given by Sperber and Claidière (2007). As I have suggested above, forces can be explanatory even if one lacks knowledge about the causal underpinnings of that force. One does not need to know what struck the billiard ball to explain its trajectory towards the corner pocket. Instead, what this intra-force harmonisation problem renders difficult is the *application* of general formalisations or models to real-world scenarios. Without a clear sense of how various factors impact reconstructive learning and how these factors are to be composed, we lack a way of making predictions, or explaining past scenarios in anything but an *ad hoc* manner.

But is this really a problem? I am not sure that it is, or at least, that it is a problem unique to the account on offer. Evolution by natural selection also relies on heterogeneous sets of causal processes and complex biological and ecological trade-offs to explain adaptations and evolution (Sober 1993; Sterelny and Griffiths 1999). Evolution by natural selection, too, faces a problem of *intra-force harmonisation*. And if evolution by natural selection can be successful theory despite providing *ad hoc* explanations, there does not seem to be a reason why CAT should not be treated in an analogous way.

So there does not seem to be much to worry about with the first kind of harmonisation problem. But there is a second harmonisation problem: How is it that the force of cultural attraction should be combined or integrated with other cultural evolutionary forces? That is,

how does the force of attraction work with or against forces that determine who or what one should pay attention to, or the fitness ramifications of adopting particular cultural variants? Let us call this the problem of *inter-force harmonisation*. It points to the difficulty in combining the various cultural evolutionary forces—for example, the force of attraction, the force of cultural selection, and perhaps some cultural evolutionary analog of drift (Bentley et al. 2004). Is this a problem for this reconstruction of CAT?

Here too, I would suggest that it is not. Or, at least, is not a problem that uniquely bedevils CAT. As Sober writes, each theory of forces “must discover how to combine the forces it describes [...], there being no antecedent recipe that is guaranteed to work for all cases” (Sober 1993, 31–32). This second harmonisation problem is one for cultural evolutionary theory as a whole, not specifically for CAT.

6.2 *CAT and Human Psychology*

Is CAT wedded to a particular view of human psychology? In reading work by cultural epidemiologists, one might reasonably think so. CAT seem to be wedded to a speculative picture of massively modular minds (e.g. Sperber 1996; Sperber and Hirschfeld 2004). Why might this be the case? Here is one plausible reason: shared aggregate transformation rates allow for modelling the “convergence of [one’s] affective and cognitive processes with those of many people toward some psychologically attractive type” (Sperber 1996, 106). But why should we expect such convergence?

Cultural epidemiologists have long defended populational ‘convergence’ as being the result of “evolved domain- or task-specific autonomous devices or ‘modules’” (Sperber and Hirschfeld 2004, 40). Indeed, Sperber (1996, 2006; Sperber and Hirschfeld 2004, 2007) has long maintained an idiosyncratic version of the massive modularity thesis (cf. Fodor 1983; Sterelny 2001), where ‘modules’ are special-purpose evolved psychological mechanisms that have been selected for robust processing stimuli. According to Sperber, assuming massive modularity can ground the assumption that human beings will have broad similarities in their cognitive and affective biases.

We do not need to go further into the details of Sperber’s account of modularity here. All I seek to argue for here is that one can be agnostic about an appeal to “a modular organization of human cognitive abilities” (Sperber and Hirschfeld 2004, 45). One does not need to ‘go modular’ to ground the assumption that individuals in population can share inferential biases.

Consider here Cecilia Heyes and Chris Frith’s (2014) recent argument that capacities for mind reading (capacities for predicting, explaining and manipulating the behaviour of others) are culturally inherited in a manner analogous to print reading. Marshalling a wide range of cross-cultural evidence, Heyes and Frith show that there exists widespread variation in the ontogeny of such capacities. Children in western countries like Australia and the United States seem first to grasp the fact that individuals have beliefs and may vary in their beliefs before they attribute ignorance or knowledge to individuals. By contrast, in countries such as China and Iran, children seem to acquire these capacities in precisely the opposite order: understanding first that individuals can be ignorant or knowledgeable before grasping that individuals may have diverse beliefs (Heyes and Frith 2014, 1243091–2). Heyes and Frith’s suggestion here is that such cross-cultural variation can be explained by positing an innate capacity for *implicit* mind reading that is developed and bootstrapped in distinct ways in different cultures into *explicit* mind reading capabilities.

Implicit mind reading is a mechanism or set of mechanisms that enables automatic, fast and fluid tracking of other’s behaviour. Using an ingenious experiment, Santiesteban and colleagues

(2014) showed that adults quickly and automatically take perspectival information of others into account. This automatic incorporation of other's perspective occurs even when the 'other' is a non-human stimulus with orienting features (e.g. an arrow-shaped avatar). This suggests a quite general mechanism that quickly and implicitly takes orienting information into account (cf. Apperly 2011; Heyes 2014). According to Heyes and Frith (2015; cf. Heyes 2012), this implicit mechanism is then *exapted* or bootstrapped by various local cultural traditions and practices into more familiar explicit mind reading, exemplified in folk psychology.

This account relies on an innate capacity for *implicit* mind reading: a mechanism that orients one towards and incorporates the viewpoints of others. To the extent that Heyes and Frith appeal to an innate capacity, then, one might think there is little light between their account and Sperber's. But there are deep disagreements here. For Heyes and Frith, what explains explicit mind reading are cultural practices: variation in mind-reading development results from variation in early childhood environments. This is in comparison to Sperber's (1996; Sperber and Hirschfeld 2004), where such explicit mind reading capacities are universal and innate.

So while both accounts appeal to specific adaptations or 'start-up kits' that predispose human beings to attend to and orient to specific stimuli—they differ in the extent to which these adaptations invariably end up generating specific capacities for mind reading. For Heyes and Frith, implicit mind reading "is radically insufficient for the development of explicit mind reading" and "no amount of individual learning—implicit mind reading, introspection, and watching the behavior of others—would be enough for the development of explicit mind reading" (1243091–5). If Heyes and Frith are right, then while some cognitive structures are needed for explicit mind reading, further scaffolding is needed. And this scaffolding can vary across cultures and shape explicit mind reading in different ways.

Whatever you make of Heyes and Frith's account, it seems a plausible alternative to Sperber's nativist view of the mind. Such accounts are further buttressed by recent work in the philosophy of biology. Kim Sterelny (2003), for instance, has defended an account where social learning accounts for folk psychological capacities. Matteo Mameli (2004), too, has defended the idea that there can be selection for capacities to learn, and that such learning can do the work of reliably acquiring adaptive traits generation after generation. Taken together, this suggests that one can purchase the robust, shared capacities that CAT theorists appeal to without buying into modular minds.

If this is right, then the reconstructed version of CAT on offer here is not strictly beholden to any particular view of the mind. The inferential biases appealed to in CAT models need not say anything about their causal origins. This is a big difference between my reconstruction and the CAT literature—but I see this as a feature, and not a bug.

Still, the abstract character of the force of attraction does raise tricky issues for the causal interpretation of such models. Cultural epidemiologists have long assumed that 'going modular' solves what I have called the intra-force harmonisation problem. The default interpretation of CAT models has been that deep cognitive architecture is the most important causal factor in determining the strength and direction of the force of attraction. This default assumption is visible in both CAT theoretical writings (Sperber 1996; Heintz and Claidière 2015) and empirical work (Miton et al. 2015). But if something like Heyes and Frith's account is right, then this assumption needs to be jettisoned. Local cultural variation may be more important than yet recognised.

The issues at hand reflect ongoing empirical and philosophical issues, and here is not the place to settle the debate. Though there seems to be momentum in scientific and philosophical opinion towards accounts like Heyes's or Sterelny's, the pendulum of opinion may swing back to a nativist one in the future. Whatever the long-term outcome, it is important to note that

on the reconstruction on offer, CAT is insulated from these debates; one can appeal to and use inferential biases in one's models regardless of their origins in evolutionary history or local cultural context.

6.3 *Circularity and Pan-Attractivism*

In a recent paper, I argued that CAT sometimes verges on circularity (Buskell 2016). Occasionally, cultural epidemiologists write as though certain variants are attractive because they are prevalent, and are prevalent because they are attractive. Is there anything to this claim?

To see what is at stake, we are well-advised to return to the discussion of cultural attractors. I argued above that we can isolate four characterisations of the cultural attractor concept (TRANSFORMATION PATTERN, CLUSTER POINT, INCREASING FREQUENCY, and EQUILIBRIUM FREQUENCY) and that cultural epidemiologists should not restrict themselves to identifying cultural attractors with a sub-set of possible population outcomes. This is because there is a wide variety of possible distribution outcomes that are not well captured by one or all of these distributional characterisations. It is this identification that underpins my criticism. CAT verges on circularity when it identifies attractors with particular distribution outcomes.

To be sure, it can be the case that aggregate transformation rates generate distributions with increasing frequencies of variants, clustering, and equilibria points. But the problem is that forces other than attraction can generate such distributions. A force of cultural selection could also generate clustering, increasing frequency, and stable equilibria. But such distributions result from discriminate sampling and not from inferential biases.

The solution to all of these worries is to recognise the idealisations at work in cultural epidemiological work. In their models and discussions, researchers typically hold an implicit assumption that cultural attraction is the only force at work. Additionally, they hold that the force of attraction will generate distributions where attractive cultural variants both increase in relative frequencies and cluster together (e.g. Claidière and Sperber 2007). But such clustering and increases in relative frequencies can be counteracted by other forces. Indeed, just as a ball need not roll downhill if there is a contravening force pushing it back up; contravening forces in cultural evolution may drive populations to adopt variants that are not attractive—if what 'attractive' picks out is the variant that one *would have* ended up with *had no other forces been acting*.

This seems to be the most charitable way of interpreting many of the claims of cultural epidemiologists. Rather than seeing their claims as circular, one should see them as making—to coin an ugly term—*attractionist* or *pan-attractionist* claims. These are claims that hold that in the absence of other forces, the force of attraction will drive populations to particular distribution and equilibrium outcomes.⁹

Adopting an attractionist viewpoint is not a bad thing. After all, adopting an adaptationist perspective allows for useful modelling of a wide variety of traits. But one should be cautious in adopting a *pan-attractionist* viewpoint: a view that the force of attraction is the most significant of the forces acting on cultures. One should be cautious of such a position not only because analogous pan-adaptationist stances (*sensu* Lewens 2009) have been widely criticized in evolutionary theory (notably, by Gould and Lewontin (1979), and most recently by Elisabeth Lloyd (2015)), but also because accounts that appeal to discriminate sampling seem to get real traction on cultural evolution and the evolution of human cognition.

⁹Of course, as I note above, the force of attraction—the aggregate of transformation rates—need not generate such clustered or increasing frequency distribution outcomes or even equilibria. This can be the case *even if* the force of attraction is the only force impinging on a population.

Humans are responsive to information gradients in the local environment (Sterelny 2003). Information gradients exist where useful information is unequally distributed among individuals in a population. Given such gradients, natural or cultural selection can produce strategies for navigating these gradients; strategies like learning from the most successful or most prestigious in one's society. Indeed, the most well-developed and robust models within cultural evolutionary theory assume information gradients to explain human capacities for culture (e.g. Boyd and Richerson 1985, 2005). These capacities in turn have been used to explain a variety of cultural phenomena, notably, the differential success of populations in similar environments and why specific individuals are appealed to for help or learning experiences in particular domains of knowledge (Henrich 2001, 2016; Henrich and Henrich 2010).

Nonetheless, much like the issue of massive modularity, we need not settle the issue of pan-attractionism here. Whatever the views of CAT theorists, one can recognise pan-attractionism as a viable modelling strategy. Adopting such a strategy produces models demonstrating what culture would look like if the force of attraction were the only force impinging on a population. But one can adopt this strategy without making the questionable inference that whatever cultural distributions and form one observes in a population is the result of the force of attraction.

6.4 *Explanatory Scope*

The reconstruction I have offered in this paper employs a range of conceptual tools—notably, 'force talk'—to show how CAT can contribute to explanatory projects at work in cultural evolutionary theory. As I hope to have shown, this is a reconstruction with a number of positive features. But there are also likely to be a number of negative features, at least from the perspective of practicing cultural epidemiologists.

First, the reconstructed account on offer here is modest about CAT's theoretical potential. It holds that CAT is an enterprise concerned with offering distribution and evolvability explanations. This is in contrast to several cultural epidemiologists who hold the scope of CAT to be much broader than this. Sperber (1996), for instance, argues that CAT is at least partly an account of human cognitive evolution, and Claidière et al. (2014) hold that CAT is general framework for research in cultural evolution. Though both of these claims are contentious (Sterelny 2001; Acerbi and Mesoudi 2015; Buskell 2017), they are legitimate epistemic aims not well represented in this reconstruction.

Second, the account I have reconstructed here does not employ all the explanatory resources that CAT has to offer. In particular, it focuses on inferential biases to the exclusion of motivational and ecological factors of attraction. While I believe that a focus on inferential biases is warranted, many cultural epidemiologists might disagree. Indeed, they might argue that a focus on inferential biases represents an unjustified jettisoning of useful explanatory resources.

Third, the account I offer here might make CAT hard to distinguish from other cultural evolutionary accounts. In particular, understanding inferential biases in terms of a 'force of attraction' might undermine its distinctiveness from other kinds of biases that can similarly be construed in terms of a force impinging on cultural population—notably content-based biases, context-based biases, or guided variation (Richerson and Boyd 2005).

Personally, I see this as a positive feature of the current account. One of the main benefits of using 'force talk' is to allow the integration of multiple distinct kinds of forces in a single explanatory framework. And this integration can take place without losing the distinctiveness of such forces. I have elsewhere defended reconstructive learning from arguments that identify reconstructive learning with content-based biases and guided variation (Buskell 2017). Nonetheless, because casting CAT in terms of a theory of forces makes it more amenable to modelling work

present in other cultural evolutionary work, it does lose some distinctiveness as an independent research programme.

7 Conclusion

CAT is not merely concerned with giving a realistic account of learning. As I interpret CAT, reconstructive learning is the centrepiece of rich explanatory framework, with the tools to contribute to two distinct explanatory projects. The first accounts for how cultural variants are distributed in populations, while the second accounts for the robust dispositional property of populations to explore culture-space. I have argued that both of these explanatory projects are best achieved if one takes CAT to be a theory of forces, one that posits a force of attraction. Understanding CAT's explanatory aims in this way has a number of payoffs: it clarifies the explanatory role of cultural attractors; it shows how the force of attraction can generate distribution and evolvability explanations; it is grounded in real causes; and it is agnostic about lower-level details to do with structure of human psychology. While there are several potential drawbacks, as I see it, none of these are sufficient to detract from the attractive version on offer here.

Acknowledgments

I am grateful to Jonathan Birch, Adrian Currie, Tim Lewens, Olivier Morin, and Thom Scott-Phillips for their comments and criticisms. This paper also benefited from the feedback provided at the 'New Directions in the Evolutionary Social Sciences' conference in Cambridge, and from two anonymous reviewers. The research leading to this paper was supported by a grant from the John Templeton Foundation, and by funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 284123.

Literature cited

- Acerbi, Alberto, and Alex Mesoudi. 2015. "If We Are All Cultural Darwinians What's the Fuss About? Clarifying Recent Disagreements in the Field of Cultural Evolution." *Biology & Philosophy* 30: 481–503.
- Apperly, Ian. 2011. *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Hove: Psychology Press.
- Bentley, R. Alexander, Matthew W. Hahn, and Stephen J. Shennan. 2004. "Random Drift and Culture Change." *Proceedings of the Royal Society B: Biological Sciences* 271: 1443–1450.
- Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Boyd, Robert, and Peter J. Richerson. 2005. *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.
- Brown, Rachael L. 2014. "What Evolvability Really Is." *The British Journal for the Philosophy of Science* 65: 549–572.
- Buskell, Andrew. 2016. "Cultural Longevity: Morin on Cultural Lineages." *Biology & Philosophy* 31: 435–446.
- Buskell, Andrew. 2017. "What are Cultural Attractors?" *Biology & Philosophy* 32: 377–394.
- Cavalli-Sforza, L. L., and Marcus W. Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton: Princeton University Press.

- Claidiere, Nicolas, Thom C. Scott-Phillips, and Dan Sperber. 2014. "How Darwinian is Cultural Evolution?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130368.
- Claidiere, Nicolas, Kenny Smith, Simon Kirby, and Joël Fagot. 2014. "Cultural Evolution of Systematically Structured Behaviour in a Non-Human Primate." *Proceedings of the Royal Society B: Biological Sciences* 281: 20141541.
- Claidière, Nicolas, and Dan Sperber. 2007. "The Role of Attraction in Cultural Evolution." *Journal of Cognition and Culture* 7: 89–111.
- Currie, Thomas E., and Ruth Mace. 2014. "Evolution of Cultural Traits Occurs at Similar Relative Rates in Different World Regions." *Proceedings of the Royal Society B: Biological Sciences* 281: 20141622.
- Fodor, Jerry A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Fortunato, Laura, and Aaron Clauset. Preprint. "Revisiting the Effect of Red on Competition in Humans." doi: 10.1101/086710
- Godfrey-Smith, Peter. 1996. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Godfrey-Smith, Peter. 2009. *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.
- Gould, Stephen J., and Richard C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: a Critique of the Adaptationist Programme." *Proceedings of the Royal Society of Edinburgh* 205: 581–598.
- Hacking, Ian. 1995. "The Looping Effects of Human Kinds." In *Causal Cognition: A Multidisciplinary Debate*, edited by Dan Sperber, Davie Premack, and Anne James Premack, 351–383. New York, NY: Oxford University Press.
- Heintz, Christophe, and Nicolas Claidière. 2015. "Current Darwinism in Social Science." In *Handbook of Evolutionary Thinking in the Science*, edited by Thomas Heams, Philippe Huneman, Guillaume Lecointre, and Marc Silberstein, 781–807. Dordrecht: Springer.
- Henrich, Joseph. 2001. "Cultural Transmission and the Diffusion of Innovations: Adoption Dynamics Indicate That Biased Cultural Transmission Is the Predominate Force in Behavioral Change." *American Anthropologist* 103: 992–1013.
- Henrich, Joseph. 2004. "Demography and Cultural Evolution: How Adaptive Cultural Processes Can Produce Maladaptive Losses: The Tasmanian Case." *American Antiquity* 69: 197–214.
- Henrich, Joseph. 2016. *The Secret of our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making Us Smarter*. Princeton: Princeton University Press.
- Henrich, Joseph, and Richard Boyd. 2002. "On Modeling Cognition and Culture." *Journal of Cognition and Culture* 2: 87–112.
- Heyes, Cecilia. 2012. "Grist and Mills: On the Cultural Origins of Cultural Learning." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 2181–2191.
- Heyes, Cecilia. 2014. "Submentalizing: I Am Not Really Reading Your Mind." *Perspectives on Psychological Science* 9: 131–143.
- Heyes, Cecilia, and Chris D. Frith. 2014. "The Cultural Evolution of Mind Reading." *Science* 344: 1243091.
- Hill, Russell A., and Robert A. Barton. 2005. "Red Enhances Human Performance in Contests." *Nature* 435: 293.

- Lewens, Tim. 2004. *Organisms and Artifacts*. Cambridge, MA: MIT Press.
- Lewens, Tim. 2009. "Seven Types of Adaptationism." *Biology & Philosophy* 24: 161–182.
- Lewens, Tim. 2015. *Cultural Evolution: Conceptual Challenges*. Oxford: Oxford University Press.
- Lloyd, Elisabeth. A. 2015. "Adaptationism and the Logic of Research Questions: How to Think Clearly About Evolutionary Causes." *Biological Theory* 10: 343–362.
- Mace, Ruth, and Clare J. Holden. 2005. "A Phylogenetic Approach to Cultural Evolution." *Trends in Ecology & Evolution* 20: 116–121.
- Matthen, Mohan, and André Ariew. 2002. "Two Ways of Thinking about Fitness and Natural Selection." *The Journal of Philosophy* 99: 55–83.
- Matthen, Mohan, and André Ariew. 2009. "Selection and Causation." *Philosophy of Science* 79: 201–224.
- Mameli, Matteo. 2004. "Nongenetic Selection and Nongenetic Inheritance." *The British Journal for the Philosophy of Science* 44: 35–71.
- Mesoudi, Alex, and Andrew Whiten. 2008. "The Multiple Roles of Cultural Transmission Experiments in Understanding Human Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 3489–3501.
- Millstein, Roberta L. 2002. "Are Random Drift and Natural Selection Conceptually Distinct?" *Biology & Philosophy* 17: 33–53.
- Morin, Olivier. 2013. "How Portraits Turned Their Eyes Upon Us: Visual Preferences and Demographic Change in Cultural Evolution." *Evolution and Human Behavior* 34: 222–229.
- Morin, Olivier. 2015. *How Traditions Live and Die*. Oxford: Oxford University Press.
- Miton, Helena, Nicolas Claidière, and Hugo Mercier. 2015. "Universal Cognitive Mechanisms Explain the Cultural Success of Bloodletting." *Evolution and Human Behavior* 36: 303–312.
- Odling-Smee, F. John, Kevin N. Laland, and Marcus W., Feldman. 2003. *Niche Construction: The Neglected Process in Evolution*. Princeton: Princeton University Press.
- Powell, Adam, Stephen J. Shennan, and Mark G. Thomas. 2009. "Late Pleistocene Demography and the Appearance of Modern Human Behavior." *Science* 324: 1298–1301.
- Ramsey, Grant, and Andreas De Block. 2017. "Is Cultural Fitness Hopelessly Confused?" *The British Journal for the Philosophy of Science* 68: 305–328.
- Reisman, Kenneth. 2005. "Conceptual Foundations of Cultural Evolution." PhD Diss., Stanford University.
- Richerson, Peter J., and Robert Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
- Santiesteban, Idalmís, Caroline Catmur, Senan Coughlan Hopkins, Geoffrey Bird, and Cecilia Heyes. 2014. "Avatars and Arrows: Implicit Mentalizing or Domain-General Processing?" *Journal of Experimental Psychology: Human Perception and Performance* 40: 929–937.
- Scott-Phillips, Thom C., and Dan Sperber. 2015. "The Mutual Relevance of Teaching and Cultural Attraction." *Behavioural and Brain Sciences* 38: e63.
- Sober, Elliott. 1993. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, Chicago: University of Chicago Press.
- Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.
- Sperber, Dan. 2000. "An Objection to the Memetic Approach to Culture." In *Darwinizing Culture: The Status of Memetics as a Science*, edited by Robert Aunger, 163–174. Oxford: Oxford University

Press.

- Sperber, Dan. 2001. "Conceptual Tools for a Natural Science of Society and Culture." *Proceedings of the British Academy* 111: 297–317.
- Sperber, Dan. 2006. "Why a Deep Understanding of Cultural Evolution is Incompatible With Shallow Psychology." In *Roots of Human Sociality: Culture, Cognition and Interaction*, edited by Stephen C. Levinson and Nicholas J. Enfield, 431–449. Oxford: Berg.
- Sperber, Dan, and Nicolas Claidière. 2008. "Defining and Explaining Culture (Comments on Richerson and Boyd, Not by Genes Alone)." *Biology & Philosophy* 23: 283–292.
- Sperber, Dan, and Lawrence A. Hirschfeld. 2004. "The Cognitive Foundations of Cultural Stability and Diversity." *Trends in Cognitive Science* 8: 40–46.
- Sperber, Dan, and Lawrence A. Hirschfeld. 2007. "Culture and Modularity." In *The Innate Mind: Volume 2: Culture and Cognition*, edited by Peter Carruthers, Stephen Laurence, and Stephen Stich, 149–164. Oxford: Oxford University Press.
- Stephens, Christopher. 2004. "Selection, Drift, and the 'Forces' of Evolution." *Philosophy of Science* 71: 550–570.
- Stephens, Christopher. 2010. "Forces and Causes in Evolutionary Theory." *Philosophy of Science* 77: 716–727.
- Sterelny, Kim. 2001. "Explaining Culture: A Naturalistic Approach, by Dan Sperber." *Mind* 110: 845–854.
- Sterelny, Kim. 2003. *Thought in a Hostile World: The Evolution of Human Cognition*. Oxford: Blackwell.
- Walsh, Denis M. 2007. "The Pomp of Superfluous Causes: The Interpretation of Evolutionary Theory." *Philosophy of Science* 74: 281–303.
- Walsh, Denis M., Tim Lewens, and André Ariew. 2002. "The Trials of Life: Natural Selection and Random Drift." *Philosophy of Science* 69: 429–446.

© 2017 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license, which permits anyone to download, copy, distribute, or display the full text without asking for permission, provided that the creator(s) are given full credit, no derivative works are created, and the work is not used for commercial purposes.

ISSN 2475-3025