Modelling cultural systems and selective filters

Fredrik Jansson[*1,2], Elliot Aguilar[†1], Alberto Acerbi[‡3], and Magnus Enquist[§1,4]

[1]*Centre for Cultural Evolution, Stockholm University, Sweden*
[2]*Division of Applied Mathematics, Mälardalen University, Sweden*
[3]*Centre for Culture and Evolution, Brunel University London, United Kingdom*
[4]*Department of Zoology, Stockholm University, Sweden*

Accepted author manuscript
Philosophical Transactions of the Royal Society B, doi:10.1098/rstb.2020.0045

A specific goal of the field of cultural evolution is to understand how processes of transmission and selection at the individual level lead to population-wide patterns of cultural diversity and change. Models of cultural evolution have typically assumed that traits are independent of one another and essentially exchangeable. But culture has a structure: traits bear relationships to one another that affect the transmission and selection process itself. Here we introduce a modelling framework to explore the effect of cultural structure on the process of learning. Through simulations, we find that introducing this simple structure changes the cultural dynamics. Based on a basic filtering mechanism for parsing these relationships, more elaborate cultural filters emerge. In a mostly incompatible cultural domain of traits, these filters organise culture into mostly (but not fully) consistent and stable systems. Incompatible domains produce small homogeneous cultures, while more compatibility increases size, diversity, and group divergence. When individuals copy based on a trait's features (here, its compatibility relationships) they produce more homogeneous cultures than when they copy based on the agent carrying the cultural trait. We discuss the implications of considering cultural systems and filters in the dynamics of cultural change.

[*]fredrik.jansson@su.se
[†]elliot.g.aguilar@gmail.com
[‡]alberto.acerbi@gmail.com
[§]magnus.enquist@su.se

1 Introduction

Most models of cultural evolution are based on cultural traits working in isolation, independently of each other, or that are strictly competing, for the one function they are expected to fill. At the most basic level, we have "neutral" models, where individuals copy each other at random and there are no inherent qualities within or between the traits that influence their transmission and longevity. A number of authors have explored the effect of random copying on various aspects of cultural change, such as the size of culture, the distribution of trait frequencies, and the rate of turnover in trait popularity (Bentley et al., 2007; Mesoudi and Lycett, 2009; Strimling et al., 2009; Eriksson et al., 2010; Lehmann et al., 2011). These models have also been expanded upon with different types of context-based transmission biases, where copying of a trait depends on its frequency in the population or the prestige of the bearer (Boyd and Richerson, 1985; Henrich and Boyd, 1998; Henrich, 2001; Henrich and Boyd, 2002; Acerbi and Bentley, 2014; Kendal et al., 2018). These biases are, however, independent from the content of the traits themselves. Boyd and Richerson (1985) also considered a content-based bias, or a 'direct bias', but this has received less attention. These cultural transmission biases are generally considered as genetically specified predispositions, even though more recent studies have underscored the importance of previous experiences and the sensitivity to different contexts (Kendal et al., 2018).

Whether traits are completely independent or influence each other through some form of biased or unbiased competition, in most of these models, they all have the same content-based influence on each other. For example, given three traits i, j and k, the effect that i has on j is the same effect that it has on k. One such effect is on the cultural transmission, meaning that i influences the transmission of j and k similarly (e.g., in a competitive setting, the existence of one trait inhibits the existence of all other traits uniformly). Another such effect is that any pair of traits (i and j, i and k, or j and k) work equally well together.

However, even a cursory examination of human culture shows the limitations of such assumptions and what they can explain. For example, religions consist of sets of moral, behavioural, and metaphysical ideas that are interdependent. Similarly, an artefact like a sword entails not only the knowledge of its manufacture and use, but also social mores about when, how, and by whom it can be used. Similar networks of relationships, often combining non-material and material elements, can be sketched for many if not all cultural phenomena, such as views about the world, identities, social institutions, political systems and society, kinship systems, food culture, ethnicity, sex and gender, and subsistence systems. It is obvious that culture has a framework formed by the relationships between different cultural elements, a feature that affects both the everyday functioning of culture and cultural change. Here, we will refer to such assemblages of traits and their relationships as *cultural systems*. Buskell et al. (2019) gave conceptual arguments for and illustrations of the potential impact of systems thinking for understanding cultural evolution, illustrating them either in a cultural state space (or automata) or as networks of traits, but remained agnostic on how to model them. We here suggest an explicit representation of cultural systems and a modelling framework for how individuals acquire and transmit culture given these structural constraints.

An important relationship between traits is that they can be mutually compatible or incompatible. Compatible traits are defined as favouring each other's appearance and maintenance

within a cultural system. For example, believing in God and donating to church can be described as compatible. Conversely, incompatible traits contrast each other's appearance and maintenance. Believing in a monotheistic God and believing in Shiva are incompatible in principle. While the rules themselves (or the beliefs in them) are the cultural traits, these relationships of consistency or compatibility form exogenous constraints between them. Compatibility and incompatibility alone are sufficient to generate complex patterns, since two traits may be incompatible with each other, while both are compatible with a third, thereby creating a conflict within the cultural system. Knowledge about climate change (a) and wanting to prevent it (b) are compatible. Going to a conference on climate change (c) increases knowledge (a) and is more likely given knowledge, so they facilitate each other. However, travelling to that same conference (c) contributes to climate change through pollution, so it is incompatible to (b). Such internal conflicts are known from balance theory (Heider, 1958).

Incompatibilities such as those above could potentially create cultures filled mainly with conflicting traits. In a world of conflicting traits, random assemblages of traits are very likely to be incompatible, at least in large sets of traits. However, if the acquisition of cultural traits by individuals were to depend on trait compatibility, then more harmonious cultural systems could evolve. Such a dependence would introduce an element of self-organisation into cultural evolution, where adopted cultural traits influence the selection of new traits.

In contrast, should all cultural elements be independent, the selection of traits must be determined by some force outside culture itself, that is, environmental factors and genetic predispositions, such as inborn transmission biases (Boyd and Richerson, 1985). At the same time, the origin of these biases are left as black boxes (Heyes, 2016), and it is unclear how common and important they are, and to what extent genetic biases can explain the occurrence of cultural systems.

A systems approach to culture could potentially provide alternative explanations, less dependent on genetic control, for many patterns and outcomes of cultural evolution. Also, transmission biases could partly have their origin in other cultural traits, for example, when selective imitation leads to "guided variation" (Boyd and Richerson, 1985).

In this paper we will study how relationships between traits may influence cultural evolution and form systems of culture. Key research questions are how and to what extent cultural evolution can organise cultural systems to become different from random assemblages of traits and promote stable systems with compatible traits.

Buskell et al. (2019) gave an overview of how we *filter* information in both acquisition and transmission, providing evidence from several fields that cultural evolution may be critically involved in their origin and how they are formed, but did not investigate this further. We will here give an explicit operationalisation of a basic selection or filtering mechanism covering the different modes of acquisition and transmission, and modelling this, we will study how they give rise to culturally evolved filters, and the extent to which these lead to the emergence of organised culture.

In Section 2 we design a mathematical model of cultural systems and selection among interdependent traits, whereby traits can be selectively rejected or acquired, and to what extent these can lead to self-organisation of cultural systems. We assume only a preference for consistent information, which has empirically well-established manifestations in the ubiquitous confirmation bias (Nickerson, 1998) and avoidance of cognitive dissonance (Festinger, 1957;

Elliot and Devine, 1994; Cooper, 2007), and is a prerequisite for building functional mental models of the world (Lewandowsky et al., 2012). We are more likely to accept and use information that is consistent with our present beliefs and values. We present results from simulations to address these questions:

1. How do relationships between traits affect the size and diversity of culture?
2. How consistent do cultural systems become through self-organisation?
3. How stable are cultural systems?

In Section 3, we explore and implement different modes of the filtering mechanism. We compare the resulting cultures in terms of size, homogeneity and consistency between the different emerging filters.

Finally, we summarise and draw some general conclusions in Section 4.

1.1 Similar models

The idea that culture should be considered as a complex system of inter-related elements is common in anthropology and it has received some attention in recent theoretical and experimental works relevant to cultural evolution, focusing on language (Enfield, 2014; Kirby et al., 2007), or on the inner recurrent structure of technology and its systemic and self-organising combinations (Arthur, 2009). Buskell et al. (2019) provide an overview of related works. So far, however, only a few attempts have been made to include interdependence among traits in modelling work on cultural evolution, and rarely are these dependencies content-based.

Axelrod (1997) presented a model where individuals copy each other based on the number of traits in common. More recently, Goldberg and Stein (2018) studied compatibility between traits that evolved culturally, through associating traits by observing other agents displaying them pairwise. Traits still operate in isolation, but by observing them in tandem, agents associate pairs of traits to each other. In contrast to our research questions, there is no external world that sets exogenous constraints on systems, and no issues of consistency, but they rather study the emergence of social agreement on preferences for clusters of arbitrarily associated traits. Similarly, Yeh et al. (2019) allowed links between traits to form and break, and for both trait variants and links to be transmitted in packages. Instead of acquiring and spreading traits, with a variable size of the culture, agents have a set number of traits, each of which can take a number of variants. A trait variant is transmitted along with the variants of the traits that the sender has linked to that trait. While this model studies package transmission of trait variants in competition with other variants, with resulting hitchhiking of less functional traits as being part of a package, our focus is on selection. Instead, we study the sequential acquisition of (exogenously constrained) culture and how that moulds our cultural cognition and subsequent filtering of new traits (c.f. Enfield's micro-scale cycle of transmission, 2014). These three previous models, however, show that links or relationships can have a variety of effects, added to those we will study here; they can lower cultural diversity or split people into groups, and traits can spread also by their association with other traits rather than their own merits.

Models where the nature of a trait influences cultural evolution include some of Acerbi et al., who showed that traits that make individuals less open to change but also more efficient as

cultural models are likely to evolve (2009), and fashion-like phenomena can emerge in cultural systems consisting of material traits and preferences for such traits (2012). These attempts have led to the emergence of phenomena that do not emerge in models where traits are independent.

Some models of cumulative cultural evolution build directly on the idea that traits have different relationships with each other (Enquist et al., 2011). Such assumptions create particular sequences of cultural evolution; for instance, a trait j might not evolve easily on its own, but the evolution of trait i facilitates the evolution of trait j.

The concept of facilitating traits was also established in the field of memetics under the label of "memeplexes" (Blackmore, 1999). These are sets of traits that are replicated together, and are thus systems of traits, or "memes", with a positive interrelationship.

Claidière et al. (2014) have a more explicit approach, defining the impact of traits on the frequency of other traits through "evolutionary causal matrices". There are also more specialised models where specific behaviours, such as selective cooperation with certain individuals, are determined by a set of traits (see, e.g., Tarnita et al., 2009; Axelrod, 1997). Our cultural systems approach shares the property with the evolutionary causal matrices that impact on transmission between traits can be represented by matrices. However, we study systems defined by essential interrelationships between traits, and treat the transmission process separately, allowing for these relationships to have various impact on cultural transmission.

2 Modelling cultural systems

Cultural systems are complex and there are many possible model formulations for investigating their emergence and change. In this paper, we focus on the effect of trait relationships on the evolution of cultural systems. We model a population of interacting agents that accumulate cultural traits through copying and innovation. In the Supplementary Section S1, we provide formal definitions of general systems and some mathematical properties. Here we make some further assumptions apart from those in the formal definition.

2.1 Description

We define a cultural system as a set of traits and a set of relationships between them. Cultural systems exist both at the level of a population (i.e., all the traits present in the population and how those traits interact) and the level of an individual (i.e., one's own traits and how they interact), roughly analogous to the ideas of a gene pool and a genome for genetic evolution.

Let $G = (V, E)$ be a trait pool (or domain) of cultural traits V and relationships E, with weights $w(E)$ assigned to them, where $w_{ij} := w(e_{ij}) \in \{-1, 1\}$ describes the relationship between traits v_i and v_j, such that when $w_{ij} = 1$, they are compatible, and when $w_{ij} = -1$, they are incompatible. We thus exclude gradual ($0 < |w_{ij}| < 1$) and neutral relationships ($w_{ij} = 0$), and we assume commutativity, that is, $w_{ij} = w_{ji}$. (Of course, in reality relationships may exist between any number of traits, and their effects need not assume discrete values. Similarly, relationships may be fundamentally asymmetric, for example language must be acquired prior to literacy, but not vice versa. While we recognise these complexities, we start with these simplifying assumptions in order to facilitate interpretation of our model.) We assign compatible relationships to pairs of traits at random with a probability $(c + 1)/2$, the

proportion of compatible trait pairs (otherwise, pairs are incompatible), such that the expected compatibility in the trait pool is $c \in [-1, 1]$. We can then vary the trait pool along this single dimension to capture the notion that for different domains of culture, the range of possibilities for cultural evolution will vary.

Consider, for example, the functional versus symbolic design features of a canoe. Rogers and Ehrlich (2008) found that symbolic designs differentiated more rapidly in Polynesian canoes than functional structures. Indeed, symbolic design demonstrates an enormous range of possibilities for combining elements. By contrast, only certain combinations of functional structures will make the canoe float. The trait pool allows us to represent these differences in design space by varying the number of compatible trait pairs.

In our simulation, agents encounter one another at random at discrete time intervals. At each time step, each agent observes one random agent, and makes two decisions: first, an agent randomly selects one of her partner's traits and has the opportunity to copy it. Next, the agent has the opportunity to invent a new trait (i.e., sample a trait at random from the trait pool), with some agent introducing a new trait on average once every ten time steps. Whether an agent copies her partner's trait is determined by its compatibility with her current traits. The agent calculates the average compatibility of the potential trait with the traits in her current repertoire, a value called the score (s). The probability of copying is then determined by the following logistic function:

$$p(s) = \frac{1}{1 + e^{-ks}} \tag{1}$$

where the parameter k determines the strength of the dependence on s, and thus how much inconsistency the individual allows. We used $k = 10$ in the following simulations. Using smaller values decreases the importance of relationships between traits, and using larger values did not alter the results qualitatively. Also using individual probability functions with k chosen uniformly randomly for each agent with mean 10 produced similar results. The logistic function maps values of the score, which in our case can be anywhere on the interval $[-1, 1]$ (but it allows for any real number, e.g. when using a summed score instead of an average), to probabilities $p \in [0, 1]$. Thus, the more compatible the new trait is on average with an agent's existing traits, the more likely it is to be copied.

We maintain a population of fixed size, N, though there is a population turnover through deaths and births. Initially, all agents possess no cultural traits, and only by sampling from the trait pool do traits accumulate in the population. Deceased agents are replaced by newborns with no culture, who acquire traits via encounters with other agents and by invention.

To summarise, the simulation model proceeds in the following steps:
Initialisation: A trait pool G of T traits $V = \{v_1, \ldots, v_T\}$ is constructed. All pairs of traits are assigned compatibility relationships $w_{ij} \in \{-1, 1\}$ at random according to a specified proportion c of compatible relationships.
Iterations: At each time step:

1. Agents select another agent at random from whom they may potentially copy a trait.

2. Agents choose one of their partner's traits at random, then calculate its average compatibility score s with their own traits. The probability of copying is then determined by Eq. (1).

3. Agents are given the opportunity to innovate (directly sample a trait from V) with probability $\frac{1}{10N}$.

4. Each agent is selected to die and be replaced with a naïve individual with probability $\frac{1}{N}$.

In this simulation, cultural systems emerge as agents in the population invent traits and transmit them through interactions. We are interested in the properties of the systems that arise. In particular, we measure the compatibilities with other traits the agent has, compatibilities and similarities between agents, and the size of the cultural systems (see Supplementary Section S2 for formal definitions of these measures).

In the following results, we varied the average compatibility in the trait pool (c) for populations of fixed size ($N = 100$) for 10^5 rounds of interaction. The average lifespan of an agent was 100 interactions and the trait pool contained $T = 10^5$ traits. We ran ten simulation runs for each constellation of parameters. We also simulated runs in which agents copied one another with a fixed probability, regardless of compatibility among traits. These unfiltered runs provide a baseline of comparison for our model.

2.2 Results

All the features of the cultural system that we measured reached stationary values well before the end of the simulations. Thus, for each measure we recorded the values for each 1000th time step during the last 20% of each simulation run, and report the averages of these. We directly address our results to the questions posed in the introduction. The results are presented in Figure 1, see the blue 'trait' lines with squares (the other curves are explained in the next section).

2.2.1 How do relationships between traits affect the size and diversity of culture?

We measured the size of culture as the number of cultural traits possessed by at least one member of the population. We also measured the average number of traits possessed by a single individual (repertoire size). The bottom panel of Figure 1 shows the culture and repertoire sizes for the model with filtering compared to the 'none' case where compatibility is not considered.

The majority of traits that enter the culture in the 'none' case are filtered out when compatibility is considered and the universal compatibility is negative, leading to small cultures, while filtering has less of an effect for positive average compatibility, both for individual agents and in the whole population. The dependence of culture size on compatibility is not linear, but follows an S-curve.

The middle panel of Figure 1 shows the average proportion of traits held in common between pairs of individuals. Filtering for compatibility between traits always produces higher sharing than unfiltered copying, except for when most traits are compatible, in which case agents are as similar with or without filters.

Overall, we see that trait relationships determine the size of culture, and results in greater similarity between individuals.

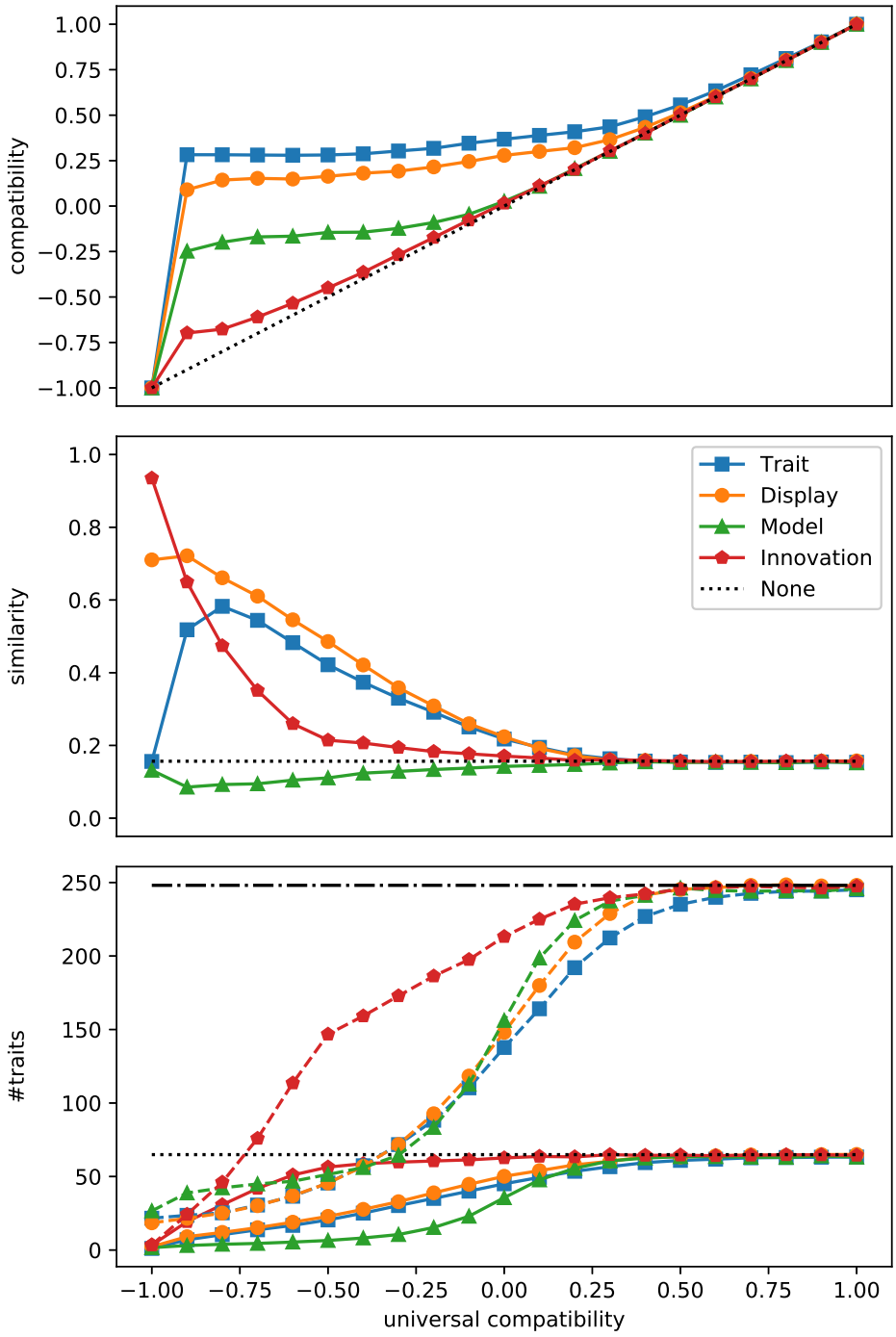

Figure 1: Internal compatibility, similarity between agents and individual (solid lines) and population (dashed lines) culture size for different modes of filtering and varying trait pool compatibilities. The dotted lines are the results without filters.

2.2.2 How consistent do cultural systems become through self-organisation?

We investigated both the average compatibility among traits within an individual's repertoire (internal consistency) and the average compatibility between individuals' repertoires.

The top panel of Figure 1 shows the internal consistency with the indiscriminate, unfiltered copying as a reference (which coincides with compatibility of the trait pool, since the cultural repertoires are then random subsets of the trait pool). Even for largely incompatible trait pools, the filtering of traits organises individual cultural systems that are more compatible than incompatible. The compatibilities between pairs of individuals align with the curve for internal consistency, giving a compatibility between individuals on average 0.03 below that within individuals (and never more than 0.07 below, except for universal compatibilities at -0.9 and below). As the trait pool compatibility increases, the cultural systems remain at largely constant compatibility, while the size of the culture increases. Interestingly, it is not possible to filter out all the inconsistencies, so systems cannot become fully consistent with the current filtering mechanism.

One way to make the systems more consistent might be to assume a more restrictive filtering mechanism. This can be achieved by reducing the score s in Eq. (1) by a constant. Reducing s by 0.5 in our model produced tiny, non-shared repertoires, and thus almost no culture. For there to be any culture, we thus need to accept some inconsistencies.

With relatively few compatible traits in the pool, the filters thus produce a small consistent culture that is shared by most agents, and as compatibility increases in the pool, more traits can be added to the individual cultural systems, which start to diverge, making agents more dissimilar, while the consistencies of the individual cultural systems and the compatibilities between them stay constant.

2.2.3 How stable are cultural systems?

Cultural practices often remain largely stable over time, even if other groups of people exercise other practices and cultural transmission is imperfect. This has been explained through cognitive factors related to mental reconstruction of transmitted content (Sperber and Hirschfeld, 2004) and conformist biases (Henrich and Boyd, 2002). Cultural systems with mutually reinforcing traits may provide a mechanism for both stability within clusters of traits and diversity between them.

We ran a simulation for 200,000 iterations, after which the population was duplicated and the two initial clones evolved independently for another 200,000 iterations. Figure 2 shows the average compatibility and similarity over time between the cultural systems of the individuals, where each individual is compared to each individual in the other population. Comparing the populations to the population at the time of the split produces roughly the same figure.

For largely incompatible trait pools (average compatibility -0.6), most pairs of populations remain almost constantly similar over the 200,000 iterations, except for a few pairs that rapidly decrease somewhat in similarity in the middle of the time period (due to one of the two populations diverging from the cultural system at the time of the split), and then remain at the new level.

For moderately incompatible trait pools (average compatibility -0.2), for which there are

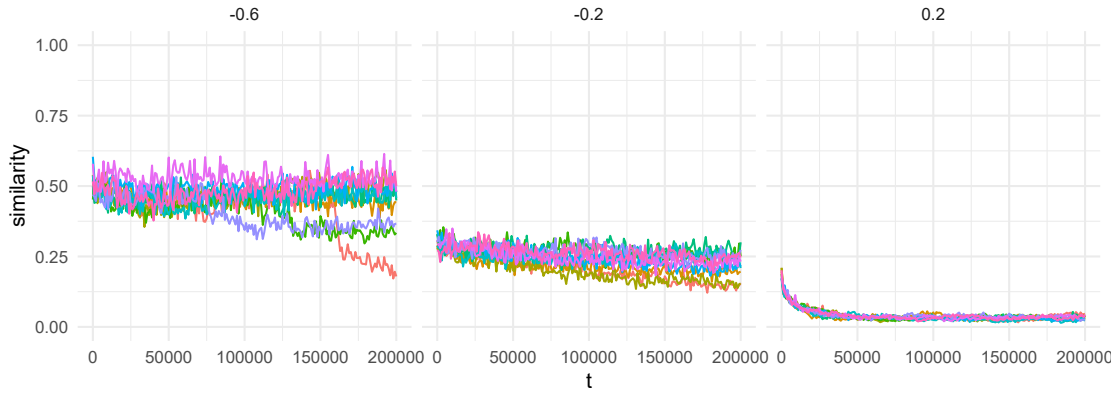

Figure 2: Average similarity over time, for ten simulation runs, between agents compared pairwise between two populations that were initially identical, for different average compatibilities in the trait pool.

more compatible clusters of traits and there should be more opportunities for gradual movement between these, the similarity between the populations decrease slowly and gradually.

Finally, with moderately compatible trait pools (average compatibility 0.2), the populations quickly diverge. This is also expected, since when traits are more compatible than incompatible, new traits easily enter into the existing cultural systems, and by random sampling, these traits will be different in the two populations.

The larger stability in mostly incompatible trait universes is consistent with the empirical observation of the Polynesian canoes (Rogers and Ehrlich, 2008) that symbolic features, which should have few compatibility constraints, change more quickly than functional aspects, with physical constraints on compatibility. As in this empirical example, the systems in the incompatible universe simulations are not stable in the sense of being attractors, but by changing slowly and gradually. At the same time, the sudden drop in similarity followed by little change for some of the cultures in the incompatible universe suggests that mutually supporting clusters can be replaced by other mutually supporting clusters (cf. punctuated equilibria).

Further simulations suggest that the stability does not seem to be a result of there being few compatible clusters. When populations were separated from start, compatibility levels between the populations were on par with the universal compatibility of the trait pool, and similarity levels were close to zero. Not only can cultural systems be mostly stable, but there are also many mutually supporting clusters, and the different outcomes suggest there is strong path dependence.

3 Additional instances of filtering

3.1 Description

There are different ways in which the compatibility relationships delineated in the trait pool can affect the introduction and transmission of traits in the population. In our first model, we

assumed that transmission was determined by the compatibility of the observed trait to the current traits of the observing agent, that is, agents will filter new social information based on how well it fits with their current knowledge. However, rather than placing the emphasis on the trait, the receiver could evaluate the sender. It is well-known that transmission is not only based on content, but is also influenced by the carrier of the trait (Jiménez and Mesoudi, 2019). There are also two more parts of the aquisition and transmission chain of individual and social learning where filtering can take place: what will agents invent, and what will they pass on? We will here introduce filtering also for these parts, based on the same consistency mechanism. There are some similarities to content-based and model-based biases suggested by Boyd and Richerson (1985), but rather than being equipped with such biases innately, filters emerge from a simple rule of cognitive consistency along with culturally aquired traits, and are thus learnt. For example, a content bias is inborn and linked to certain content at the outset, while a filter evolves culturally and acquires links over the lifetime of the agents. We will use the term *filter* for the selection based on cultural traits that emerges in agents and *filtering* for the basic mechanism.

What and whom to copy: As in our first model, agents can choose to copy a displayed trait based on its compatibility to their existing repertoire (*trait filtering*). Instead of evaluating the single trait, the receiver could evaluate the sender (cultural model). We operationalise this by letting agents base copying on the overall compatibility of the repertoire of their interaction partner with their own (*model filtering*).

What to transmit: Agents sample a trait from their repertoire and can decide whether to display that trait and make it available for copying by others according to the compatibility to other traits within their repertoires (*display filtering*). Otherwise the agent displays nothing.

What to innovate: An agent can decide whether to introduce a new trait from the trait pool based on its overall compatibility with the agent's existing repertoire (*innovation filtering*).

The mechanism is the same as in the first model, so the relationships between traits are used to calculate a score, s, which then gives the probability to copy, display or innovate, based on the logistic function $p(s)$ in Eq. 1. The score is calculated in the following ways, depending on where filtering is active (see Supplementary Section S3 for equations):

Trait filtering: s is the average compatibility of the observed trait with the existing traits in the observing agent's repertoire.

Display filtering: s is the average compatibility of the randomly selected trait with the existing traits in the model agent's repertoire.

Model filtering: s is the average compatibility between both agents' trait repertoires.

Innovation filtering: s is the average compatibility of the new trait with the existing traits in the agent's repertoire.

Note that $-1 \leq s \leq 1$ in our simulations, since the underlying compatibility relationships assume values of 1 or -1.

To isolate the effects of when filtering takes place, each filtering mechanism is implemented independently. When trait or model filtering is in effect, agents display traits at random (uniformly) and are allowed to innovate with a fixed probability that leads to, on average, one innovation per ten agents in their lifetime. When innovation or display filtering is in effect, the observer always copies in an interaction where a trait is displayed.

We ran simulations for each mode of filtering in the same way as in our first model, varying c, with $N = 100$, 10^5 rounds of interaction, average lifespan of 100 interactions, and a trait pool of $T = 10^5$ traits. As before, we ran ten simulations for each constellation of parameters, and will compare the results to runs without any filters in effect.

3.2 Results

As for trait filtering in the previous section, for each measure we recorded the values for each 1000th time step during the last 20% of each simulation run, and report the averages of these. The results after 10^5 iterations are presented in Figure 1. The between-individual compatibilities were similar to the internal compatibilities, with a few alterations commented below, and are therefore not included in the figure.

In general, basing the transmission of a trait on its compatibility to other traits of the sender or receiver provides the highest compatibility of cultural systems and similarity between them. Looking in more detail at each mode of filtering, we find the following.

Display filtering Display filtering makes agents slightly more similar than trait filtering does, and their repertoires slightly less internally consistent. It is the sender's repertoire that determines the probability for a trait to be transmitted, so the receiver cannot tailor its repertoire to be compatible, and avoid unfit traits from senders. Even if agents are more similar, since their internal consistency decreases compared to trait filtering, so does the compatibility between agents (not in the figure), to levels similar to the internal consistency.

Model filtering Model filtering produces dissimilar, incompatible (close to the trait pool compatibility) agents with a negative internal consistency (when the trait pool compatibility is negative), even if it filters out some incompatibility. The lack of similarity may seem counter-intuitive. Should agents not become more similar to their cultural models, given that they copy based on mutual compatibility, as opposed to cherry-picking for compatible traits? The small number of traits in agents' individual repertoires compared to the relatively large number of traits in the population may provide an explanation. Consider a randomly sampled innovation. In a mostly incompatible universe, such a trait is most likely to reduce both the internal compatibility and the compatibility between two agents. This will decrease the probability of transmission, but only marginally if the agent with the new trait already has several other traits. In an interaction, the new trait has an equal chance to the other traits to be spread, contrasting to trait filtering, which impedes mostly incompatible innovations. This means that innovations are likely to survive, increasing the number of traits in the population, and to decrease the compatibility, decreasing the number of traits that are transmitted and thus the

size of the individual repertoires. With different innovations spreading to different agents, the agents become dissimilar.

Innovation filtering Innovation filtering remedies the influx of incompatible traits into the population and individual repertoires. Contrary to the model filtering, this limits the introduction of new traits into the population, while those that are introduced can spread freely, since transmission is unfiltered. For mostly incompatible trait pools, the agents in a population share most of the traits, leading to high similarity, and culture and repertoire sizes to be almost equal.

Innovation filtering is not as effective as trait filtering at maintaining internally consistent repertoires. A possible explanation comes from the fact that agents are born without culture, and before they have acquired most of the culture of the other agents, they have the opportunity of inventing traits that are incompatible with those other traits, and due to indiscriminate transmission, the new trait will spread. There is no filtering in the cultural transmission, only in individual learning. Since most learning is social, most traits are acquired without filtering. Adding to this effect, since old agents have larger, incompatible, repertoires than young agents, when given the opportunity to invent, young agents are more likely not to filter out the potential innovation, and are thus more likely to introduce new traits. When sampling from a mostly compatible trait pool, few traits are filtered out and the culture becomes so large that agents will acquire only a fraction of all traits during their lifetime, making agents more dissimilar.

Multiple filtering In reality, filtering would likely be at work in both acquisition and transmission simultaneously, and the model allows for combining copy filtering (trait and model) with display and innovation filtering. As expected, for the parameter values used here, combining the trait, display and innovation filtering produces slightly smaller and more homogeneous cultures, with higher compatibilities both within and between agents, than any one type of filtering does alone, when the universal compatibility is negative. For positive compatibilities, the results are on par with trait filtering, but with slightly fewer traits in the population.

4 Conclusions

We propose taking a systems view of culture as a next step in the development of theory in cultural evolution (building further on ideas by Buskell et al., 2019). We believe that in order to understand the differences between cultural assemblies, as well as how different assemblies emerge from previous ones, the relationship between traits must be considered. In order to explore the consequences of this view, we have here suggested a modelling framework that implements the idea of structural dependencies between cultural traits (in the form of pairwise relationships of compatibility and incompatibility) and emergent ways for these dependencies to influence acquisition and cultural transmission (filters).

4.1 Summary

Contrasting to recent models on interrelated cultural traits (Claidière et al., 2014; Goldberg and Stein, 2018; Yeh et al., 2019), we here focus on the developmental process of individuals, and

how cultural information filters emerge from basic mechanisms of striving for cognitive consistency along with previously acquired information that has a nonneutral relationship to new information the individual encounters. Since agents in our model are continuously exposed to each other's ideas and learn socially, they come to share culture to a great extent and develop similar filters. Dependencies between traits may affect the probability of the introduction of a new trait (innovation filtering), of copying it from someone else, contingent on the trait (trait filtering) or whether we deem others as suitable role models (model filtering), and of using it in a way such that others can observe and copy it (display filtering). In isolation, the trait filtering provided the most compatible individual repertoires and the most compatible agents, even if other types of filters would sometimes make agents more similar.

Filters organise consistent systems at a level that is largely independent of trait compatibility in the trait pool. Filters are thus highly efficient in incompatible traits pools, but there seems to be a limit to how much inconsistency can be eradicated. We cannot eradicate inconsistency without eradicating culture. The more compatible the trait pool, however, the larger the systems that emerge.

The resulting cultural systems are highly stable in domains of low compatibility, for which there is little gradual evolution, so the relatively rare changes towards new equilibria seem to occur in leaps. Meanwhile, populations without common history almost exclusively end up with different cultural systems, so there is no convergence to any particular cluster of compatible traits for agents that do not interact with each other. Thus, as in the dissemination of culture model (Axelrod, 1997), we do find local and stable convergence among interacting agents, but polarisation between separated populations, without assuming a direct mechanism for becoming more similar to other agents, nor a specific spatial structure.

4.2 Discussion

We have seen that considering systems has consequences for cultural evolution. With simple filtering tools, even universes of mostly incompatible cultural traits, cultural systems become organised and consistent, without assuming specific learning biases. In fact, it does not seem to matter how hostile the traits are to each other (except at the extremes), but the systems reach similar levels of consistency. What varies is how large the systems can be, and the degree of diversity between individuals. While filters can organise mostly compatible systems from any universal compatibility, they cannot create completely consistent systems. As long as incompatible traits are not completely blocked (as they might be if there are physical constraints, but not e.g. if the traits are beliefs), some incoherence will adhere.

Comparing to innate biases in the cultural evolution literature (Boyd and Richerson, 1985), what is being filtered is here an emergent property that sifts and sorts and organises social information, rather than skewing it in any particular direction. Nevertheless, there are similarities between these properties and the effects of certain biases. Filters whose function most closely resemble content biases in cultural transmission are the most efficient ones, especially if the filtering is made on the receiver's side (trait filtering) rather than the sender's (display filtering). Such filters are less efficient for individual learning (innovation filtering) alone. Model filters are most similar to context biases, determining whom rather than what to copy, and are less efficient at organising consistent systems.

We have assumed that the basic machinery that enables the evolution of these filters is inborn. In our models, agents prefer information that is consistent with their present information to information that causes conflict. Depending on the domain of cultural traits, the anticipation from agents' cognition varies. At one extreme, only some combinations are functional together, and agents can learn to associate functional combinations, or there may even be physical constraints that limit certain combinations, in which case laws of nature determine consistency. At the other end, as would be the case for example for belief systems, the agent is expected to use experience-guided learning. As mentioned earlier, there is ample empirical evidence that people can even experience psychological discomfort from keeping contradictory ideas, and thus avoid information that inflicts cognitive dissonance (Festinger, 1957; Elliot and Devine, 1994; Nickerson, 1998; Cooper, 2007). It would be hard for the individual to create functional, non-chaotic, mental models of the world without such a preference.

The realised filters are emergent properties from social interactions and part of individual development. How these filters will actually operate, and which specific traits they will filter out, is subject to which other traits agents acquire first. The manifestation of filtering mechanisms is thus a result of path-dependent cultural transmission.

There are several paths that can be taken in filtered acquisition of traits, with many possible cultural systems (different simulations produce different systems), but once a system has emerged, it tends to be surprisingly stable, at least if the universal compatibility of the trait pool is low. With highly compatible trait pools, systems are less stable, since new traits can easily enter, and filters are in less operation.

4.3 Model assumptions and future directions

Our model assumed that the dependencies between cultural traits were delineated and fixed at the start of the process, and that there are no higher-order relationships between clusters of traits. In reality, new compatibility relationships between individual traits arise over time and as a result of historical contingency, due to higher-order relationships. Thus, in reality these dependencies will be both a force and a product of cultural evolution.

For example, a swastika and a peace symbol are unlikely to be considered compatible symbols, not because of some a priori nature of their meanings or appearances, but because of a particular history of cultural associations attached to both symbols. While we recognise this fact, constraining the dependencies at the outset of the simulations and limiting the model to relationships between individual traits allowed us to examine the effect of structural dependencies more directly and avoid unmanageable complexity; by varying c we could explore a range of scenarios for cultural copying. Another simplification, for ease of interpretation, that could easily be removed, is the assumption of symmetric relationships between traits. Building further, asymmetric relationships, which might indicate sequential learning of traits, could be explored, and, as a greater challenge, allowing relationships between triads or higher number tuples of traits would lead to more complex structures.

One consequence is that also the basic filtering mechanism could potentially evolve culturally. We have here assumed that all agents have a preference for consistency, while their manifestations are results of learning. However, also the filtering rules, how to take compatibilities into account, might themselves be results of cultural evolution. This could be modelled by

taking higher-order relationships into account, where traits regulate the pairwise relationships between other traits. For example, a trait that dictates a decreased importance of consistency would lower all compatibility weights between traits.

Our modelling framework is an attempt to formalise the concept of cultural systems (Buskell et al., 2019). Researchers in cultural evolution have revealed a number of important phenomena using simple copying models. These approaches, often inspired by population genetics, are quite different from the traditional views of cultural anthropologists and other students of cultural change. We hope that incorporating a system-wide view will help bridge the gap between cultural evolutionists and cultural anthropologists, and hopefully lead to new insights into cultural change.

Acknowledgements

This work was supported by the Knut and Alice Wallenberg Foundation (grant №2015.0005).

References

Acerbi, A. and R. A. Bentley (2014). Biases in cultural transmission shape the turnover of popular traits. *Evolution and Human Behavior 35*(3), 228–236.

Acerbi, A., M. Enquist, and S. Ghirlanda (2009). Cultural Evolution and Individual Development of Openness and Conservatism. *PNAS 106*(45), 18931–18935.

Acerbi, A., S. Ghirlanda, and M. Enquist (2012). The logic of fashion cycles. *PLoS One 7*(3), e32541.

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution 41*(2), 203–226.

Bentley, R. A., C. P. Lipo, H. A. Herzog, and M. W. Hahn (2007). Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior 28*(3), 151–158.

Blackmore, S. (1999). *The Meme Machine*. Oxford, United Kingdom: Oxford University Press.

Boyd, R. and P. J. Richerson (1985). *Culture and the Evolutionary Process*. University of Chicago Press.

Buskell, A., M. Enquist, and F. Jansson (2019). A systems approach to cultural evolution. *Palgrave Communications 5*(131), 1–15.

Claidière, N., T. C. Scott-Phillips, and D. Sperber (2014). How Darwinian is cultural evolution? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 369*(1642), 20130368.

Cooper, J. (2007). *Cognitive Dissonance: 50 Years of a Classic Theory*. London, United Kingdom: SAGE.

Elliot, A. J. and P. G. Devine (1994). On the Motivational Nature of Cognitive Dissonance: Dissonance as Psychological Discomfort. *Journal of Personality and Social Psychology 67*(3), 382–394.

Enfield, N. J. (2014). *Natural causes of language: Frames, biases, and cultural transmission*. Language Science Press.

Enquist, M., S. Ghirlanda, and K. Eriksson (2011). Modelling the evolution and diversity of cumulative culture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 366*(1563), 412–423.

Eriksson, K., F. Jansson, and J. Sjöstrand (2010). Bentley's conjecture on popularity toplist turnover under random copying. *The Ramanujan Journal 23*(1), 371–396.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Redwood City, California: Stanford University Press.

Goldberg, A. and S. K. Stein (2018). Beyond social contagion: Associative diffusion and the emergence of cultural variation. *American Sociological Review 83*(5), 897–932.

Heider, F. (1958). *The psychology of interpersonal relations*. New York, New York: John Wiley & Sons Inc.

Henrich, J. (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*(4), 992–1013.

Henrich, J. and R. Boyd (1998). The Evolution of Conformist Transmission and the Emergence of Between-Group Differences. *Evolution and Human Behavior 19*(4), 215–241.

Henrich, J. and R. Boyd (2002). On Modeling Cognition and Culture: Why cultural evolution does not require replication of representations. *Journal of Cognition and Culture 2*(2), 87–112.

Heyes, C. (2016). Blackboxing: social learning strategies and cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences 371*(1693), 20150369.

Jiménez, Á. V. and A. Mesoudi (2019). Prestige-biased social learning: current evidence and outstanding questions. *Palgrave Communications 5*(1), 1–12.

Kendal, R. L., N. J. Boogert, L. Rendell, K. N. Laland, M. Webster, and P. L. Jones (2018). Social learning strategies: bridge-building between fields. *Trends in cognitive sciences 22*(7), 651–665.

Kirby, S., M. Dowman, and T. L. Griffiths (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences 104*(12), 5241–5245.

Lehmann, L., K. Aoki, and M. W. Feldman (2011). On the number of independent cultural traits carried by individuals and populations. *Philosophical Transactions of the Royal Society B: Biological Sciences 366*(1563), 424–435.

Lewandowsky, S., U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, Supplement 13*(3), 106–131.

Mesoudi, A. and S. J. Lycett (2009). Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior 30*(1), 41–48.

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology 2*(2), 175–220.

Rogers, D. S. and P. R. Ehrlich (2008). Natural selection and cultural rates of change. *PNAS 105*(9), 3416–3420.

Sperber, D. and L. A. Hirschfeld (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences 8*(1), 40–46.

Strimling, P., J. Sjöstrand, M. Enquist, and K. Eriksson (2009). Accumulation of independent cultural traits. *Theoretical Population Biology 76*, 77–83.

Tarnita, C. E., T. Antal, H. Ohtsuki, and M. A. Nowak (2009). Evolutionary dynamics in set structured populations. *PNAS 106*(21), 8601–8604.

Yeh, D. J., L. Fogarty, and A. Kandler (2019). Cultural linkage: the influence of package transmission on cultural dynamics. *Proceedings of the Royal Society B 286*(1916), 20191951.

Supplementary information

S1 Formal definitions and mathematical properties

First, we designate a structured *trait pool* or *cultural domain*, which is a set of potential cultural traits, along with a network of relationships, or exogenous constraints, of compatibility or incompatibility that exists between each pair of traits.

Let $G = (V, E)$ be a directed graph, representing a pool of interdependent cultural traits, with a set of vertices, or nodes, V, where each vertex v_i represents a potential cultural trait, and a set of edges, or connections or ties, E, consisting of ordered pairs of elements from V, where an edge $e_{ij} = (v_i, v_j)$ represents a relationship between v_i and v_j where v_i influences v_j. The graph is weighted, such that there is a function $w : E \to [-1, 1]$ that assigns a weight $-1 \leq w(e_{ij}) \leq 1$ to each edge e_{ij} in E, representing the influence that v_i has on v_j, that is, where $|w(e_{ij})|$ represents the extent to which v_i inhibits v_j if $w(e_{ij}) < 0$, and the extent to which v_i facilitates v_j if $w(e_{ij}) > 0$. Weights do not necessarily need to be on a scale, but the weight functions could be generalized to arbitrary values, such that $w : E \to \mathbb{R}$. The graph G is the *trait pool*, or *cultural domain*, forming constraints on how culture evolves.

A *cultural system* is a subgraph $G' = (V', E') \subseteq G$, where $V' \subseteq V$ and $E' \subseteq E$, such that if $v_i, v_j \in V$ and $e_{ij} \in E$, then $v_i, v_j \in V' \Leftrightarrow e_{ij} \in E'$, that is, G' is a sample of cultural traits from G together with the relationships in G between all of these traits.

We will here focus on how relationships between traits affect copying and innovation. There are many ways in which these relationships could be parsed or incorporated in cultural transmission, and we refer to these as cultural filters. These filters can intervene in cultural transmission in many ways. For example, they may affect how agents make decisions about what and when to copy; or they may affect what agents make available for copying, through teaching or demonstration; finally, they may affect which traits are invented and by whom.

S2 Measures to characterise systems

We are interested in the size, similarity, and compatibility of cultural systems, both internally, within agents, and between them, and whether different systems based on the same trait pool converge or diverge. Similarity is a measure of the variety between cultural systems, and is thus a comparison of systems between agents. Compatibility can be defined both between individual traits and between whole systems.

Let $G_k = (V_k, E_k)$ denote the cultural system possessed by agent $k \in \{1, 2, \ldots, N\}$ with weights $w_{ij} = w(e_{ij})$ for traits v_i and v_j in V_k. Further, let $V = \bigcup_k V_k$, that is, the population-level set of all traits possessed by any agent.

The size of the culture in the population is simply $|V|$, while the agent's repertoire has the average size $\frac{\sum_k |V_k|}{N}$.

The similarity $S_{k\ell}$, for a pair of agents k and ℓ, is a measure of how many traits in their total pools of traits are shared, that is,

$$S_{k\ell} := \frac{|V_k \cap V_\ell|}{|V_k \cup V_\ell|}$$

Averaging over each pair of agents k and ℓ for a population measure gives

$$\frac{\sum_{k=1}^{N-1} \sum_{\ell=k+1}^{N} S_{k\ell}}{\binom{N}{2}}$$

The internal compatibility for an agent k measures how compatible a trait $v_i \in V_k$ is on average with all of ks other traits $v_j \in V_k \setminus \{v_i\}$, that is,

$$C_k := \frac{1}{|V_k|(|V_k| - 1)} \sum_{i:v_i \in V_k} \sum_{j:v_j \in V_k \setminus \{v_i\}} w_{ij}$$

Averaging over all agents gives

$$\frac{\sum_{k=1}^{N} C_k}{N}$$

The compatibility between a pair of agents k and ℓ is computed similarly, that is,

$$C_{k\ell} := \frac{1}{|V_k||V_\ell|} \sum_{i:v_i \in V_k} \sum_{j:v_j \in V_\ell} w_{ij}$$

Averaging of each pair of agents gives

$$\frac{\sum_{k=1}^{N-1} \sum_{\ell=k+1}^{N} C_{k\ell}}{\binom{N}{2}}$$

S3 Filters

In calculating the probabilities of copying or innovation with filters, we made use of a quantity, called the score, s, determining the probability, p, of taking the action related to the filter, that is, to innovate, display or copy the cultural trait in question. The probability is determined by the following logistic function:

$$p = \frac{1}{1 + e^{-k(s-s_0)}} \tag{1}$$

where the parameter k determines the strength of the filter effect. For the simulation results reported in the paper, $k = 10$ and $s_0 = 0$. The score is computed in the following ways for the various filters. We denote the set of traits in the trait pool as $V = \{v_1, v_2, \ldots\}$ and the compatibility, or weight, of the relationships between two traits v_i and v_j as w_{ij}. In our simulations, $w_{ij} \in \{-1, 1\}$. We assume below that the set of traits $V_i \neq \emptyset$ for a focal agent i, and otherwise $s := 0$.

Trait filtering

When a learner ℓ encounters a model agent k, it selects one of her traits, $v_i \in V_k$, at random for potential copying. Let $V_\ell \subseteq V$ be the learner's set of traits. We define the score as

$$s = \frac{\sum_j w_{ij}}{|V_\ell|} \tag{2}$$

where j are the indices of the traits in V_ℓ.

Display filtering

When a learner ℓ encounters a model agent k, the model agent selects one of her traits, $v_i \in V_k$, at random for potential copying. Let $V_k \subseteq V$ be the model agent's set of traits. We define the score as

$$s = \frac{\sum_{j \neq i} w_{ij}}{|V_k|} \tag{3}$$

where j are the indices of the traits in V_k.

Model filtering

When a learner ℓ encounters a model agent k, it selects one of her traits, $v \in V_k$, at random for potential copying. We define the score as

$$s = \frac{\sum_i \sum_j w_{ij}}{|V_k||V_\ell|} \tag{4}$$

where i and j are the indices of the traits in V_k and V_ℓ, respectively.

Innovation filtering

When an agent is given the opportunity to invent, it selects a trait, $v_i \in V$, at random from the trait pool, for potential innovation. The score is then calculated according to Eq. 2.