# Weak content preferences stabilise culture

Alberto Acerbi

alberto.acerbi@unitn.it

Department of Sociology and Social Research, University of Trento, Italy

Centre for Culture and Evolution, Division of Psychology, Brunel University London, UK

## Abstract

TO DO

## Keywords

Cultural attraction; cultural evolution; cultural transmission; conformity; prestige; human universals; individual-based models.

# Introduction

On the first Thursday of March, the UK and Ireland celebrate World Book Day, and in many primary schools children dress up as their favourite book characters. When children turn seven or eight, parents start to recognise in the courtyard the familiar figures of the Hogwarts School of Witchcraft and Wizardry. Apparently, children in the UK and Ireland go through generational waves, where each cohort "rediscovers" Harry Potter. Why are some things successful in spreading widely and stably, such as the fictional world of Harry Potter in the last twenty years, while others are not?

An intuitive distinction concerns the effect of social influence versus the features of the content of the traits. The "rediscovery" of Harry Potter is due to parents, elder siblings, and early adopter-peers, from which children learn, as a minimum, about its existence. At the same time, it feels that the content of Harry Potter's stories should be attractive enough to reinforce social influence, in order to be stable through years, and in many different countries. Content and social influence are likely to act together, to a different degree. For some traits, however, content seems more important: western children, on average, prefer pizza to boiled spinach, no matter the efforts of parents; the great majority of cultures use, in some occasions, masks or make-up for faces (Sperber and Hirschfeld 2004). For others, it may be the opposite: hugging or and kissing can be used as greetings in some societies, but considered inappropriate in others where, maybe, handshakes, or bowing, are used; beanie hats and skinny jeans come and go.

This intuitive distinction is reflected in evolutionary approaches to the study of culture. Epistemic vigilance distinguishes, for example, between the evaluation of the "source" and of the "content" of communicated information (Sperber et al. 2010). In the cultural evolution framework, different mechanisms have been proposed as reflecting social influence, usually under the general label of indirect-biased transmission (Boyd and Richerson 1985) or context-biased social learning strategies (Kendal et al. 2018): these mechanisms act by selecting among different cultural traits the ones that are associated to some features of the context, or of the population. On the other side, direct-biased transmission, or content-biased social learning, indicate the selection of traits based on their intrinsic features. However, preferences for content also act outside the selection process:

we can adopt cultural traits via individual learning, a process sometimes labelled guided variation, (Boyd and Richerson 1985). More generally, various individual processes can make us converge with higher probability to some traits, or particular configurations of traits: in cultural evolution terminology they are referred to as convergent transformation (Acerbi et al. 2021; Mesoudi 2021), or content-based attraction (Claidière and Sperber 2007).

In what follows, I present individual-based models of cultural evolution that consider both preferences for content and social influence, both with discrete and continuous traits. I am using the label "content preferences" for mechanisms that are *not socially influenced*, that is, they do not depend directly from features of the larger population; *directional*, that is, points towards a particular trait, or traits configuration; and, finally, *stable*: they do not change, at least at the time scale of the simulations. The content preferences modelled here are equivalent, in the discrete case, to biased mutation as in Acerbi, Mesoudi, and Smolla (2022); in the continuous case, to guided variation (Boyd and Richerson 1985), biased or convergent transformation (Acerbi et al. 2021; Mesoudi 2021), or content-based attraction (Claidière and Sperber 2007).

In opposition, in the models below, "social influence" identifies mechanisms that select traits to copy based on features of the population/source, that are not associated to particular traits, and thus change accordingly to changes in the population. The simulations consider two of the most studied mechanisms of social influence: conformity (or frequency-based indirect bias) and model-based indirect-bias. Conformity is defined as a disproportionate tendency to copy from the majority, and it is implemented, in the discrete case, following Acerbi, Mesoudi, and Smolla (2022), which, in turn, is an individual-based version of the treatment in Boyd and Richerson (1985). Conformity with continuous traits is rarely modelled: here I follow Morgan and Thompson (2020). Demonstrator-based indirect-biases instead do not depend on the frequency of traits, but on features of the demonstrator. A classic example is prestige-bias, or a tendency to copy preferentially from individuals that are considered to be "high-status" (Jiménez and Mesoudi 2019), but any tendency that make to preferentially choose some demonstrators because of features independent from the copied traits would fit the description, such as copying preferentially younger (or older) demonstrators. The models implementing demonstrator-based indirect-biases are inspired, with modification, by Acerbi, Mesoudi, and Smolla (2022) and Mesoudi (2009).

Surprisingly, there is not a clear understanding of what are exactly the consequences of dynamics driven by preferences for content or by social influence. Many studies in cultural evolution focus on social influence, possibly because heuristics like conformity or prestige-bias can produce population-level adaptive behaviours that go beyond individual cognition, a process that is considered central in cultural evolution (Henrich 2016). Other researchers have instead highlighted the importance of weak but stable preferences for content, as the main way to support cultural transmission and hence stabilise traditions (Morin 2015). Few studies have considered explicitly the difference between the two process. Henrich and Boyd (2002) showed that when both social influence and preferences for content act, social influence fully determines the outcome. However, their model assume two preferences for content, and social influence acting stably towards one of the two. In response to this work, Claidière and Sperber (2007) presented a model where social influence and the target for content preferences are separated, and they show how the final equilibrium point depends on the relative strength of social influence and content preferences (see also (Claidière et al. 2018)).

In both these models, the target of social influence is however fixed and linked to a particular trait, or trait configuration. "Pure" social influence should be instead considered as determined *only* by the context, be it the frequency of any trait in the population, or some demonstrators' features independent from the copied trait. Below, I show that, when this is the case, even weak content preferences determine cultural dynamics, as they are the only directional forces.

# Methods

The model simulates a population of $N = 1000$ individuals. Individuals possess a single cultural trait. In the "discrete traits" simulations, individuals possess a discrete cultural trait, either A or B. The trait is randomly initialised at the beginning of each run according to a parameter $p_0$, which gives the probability, for each individual, of being assigned a trait A. In the "continuous traits" simulation, the trait is a value $p$ between 0 and 1, randomly initialised at the beginning from a standard normal distribution (with $\mu = 0$ and $\sigma = 1$) rescaled between 0 and 1.

In the “discrete traits” simulations examining frequency-based social influence (e.g., conformity), at each time step, three demonstrators are randomly chosen for each individual. If all have the same trait (three As or three Bs), the individual copies it automatically. In the other cases, the majority trait (i.e., the one possessed by two demonstrators) is adopted with a probability equal to $2/3 + D/3$. The parameter $D$ goes between 0 and 1, regulating the strength of conformity. With $D = 0$ (no conformity) the probability of copying the majority trait is 2/3, equivalent to unbiased copying, and with $D = 1$ (maximum conformity) individuals always copy the majority trait. In all simulations below, $D$ is fixed to 1.

In the simulations with continuous traits, each individual adopts the mean trait in the population, with an error lower than the variance of the trait, that is, randomly extracted between $-\omega(p)$ and $+\omega(p)$.

Finally, the preference for the content is implemented: in the discrete traits scenario, with a probability of $\alpha$, individuals switch to trait A. In the simulation with continuous traits, with the same probability $\alpha$, individuals adopt instead $p = 1$. Simulations run for $T$ time steps, where this process is iterated.

For simulation examining demonstrator-based social learning (e.g., prestige), at the beginning of each run, some individuals are randomly assigned to “high” status, according to a parameter $C_s$. Each individual has a baseline probability ($x$) of being chosen as a demonstrator equal to 1. For high status individuals, this probability is increased of an integer $C_{copy}$. Probabilities are rescaled with a softmax function, $e^{x_i}/\sum_{i=1}^{N} e^{x_i}$. After copying the trait of the assigned demonstrator, individuals are subject to the same process of possibly changing traits according to preferences as described above.

Table 1 summarises the main symbols and parameters used in the models and their explanation. All codes for running the simulations and reproducing the results presented here are available at: https://github.com/albertoacerbi/attraction_social.

Table 1: List of main symbols used in the models and their explanation.

| Symbol | Meaning |
|---|---|
| $N$ | population size (fixed to 1000) |
| $T$ | maximum number of time steps |
| $p_0$ | proportion of trait $A$ in "discrete traits" simulations |
| $D$ | strength of conformity (fixed to 1) |
| $C_s$ | proportion of individuals assigned to high status |
| $C_{copy}$ | gain in probability of being copied for high status individuals |
| $\alpha$ | strength of content preference |

# Results

Let's begin with the results of the simulations mixing content preference and conformity, or frequency-based social influence. Figure 1 shows the comparison of representative runs. The top-left panel shows simulations where only conformity acts ($\alpha = 0$), starting with an equal mixture of traits A and B ($p_0 = 0.5$). Each run quickly converges towards an equilibrium where only one of the two traits is present. The trait that gains, by chance, an initial majority becomes the dominant trait. The top-right panel shows the same situation, but with content preference involved ($\alpha = 0.1$). In this case, all runs converge on the same equilibrium, with the trait A, toward which the preference points, becoming dominant.

In the "continuous traits" condition, the results are similar. When only conformity acts (bottom-left panel), the population converges, in each run, on the initial traits' mean, with the conformity process rapidly reducing variance in the population (see the inset plot in the bottom-left panel). When content preference is present, all runs converge on all individuals with $p = 1$ (bottom-right panel).

In the simulations with both content preference and conformity, the same qualitative result of convergence on trait A (or on $p = 1$ in the continuous case) are reached for all possible parameter values in the continuous case and, in particular, for any $\alpha > 0$. In the discrete case, when $p_0 = 0$, that is, when all individuals have trait B at the beginning of the simulation, trait A does not reach fixation with $\alpha = 0.1$ (see Figure 2, left panel, red lines - in the continuous case, even if all individuals are initialised with $p = 0$ results do not change, see right panel). In this situation, the proportion of individuals switching to trait A is not sufficient to overturn the majority, and it is reabsorbed at each time step. With stronger content preference ($\alpha = 0.2$) trait A is back to be fixed in the population (see Figure 2, left panel, blue lines). It is interesting, in this situation, to analyse more in detail the effect of parameters $p_0$ and $\alpha$ (see Figure 3). When there is an initial majority of trait B ($p_0 \lessapprox 0.5$), $\alpha$ needs to be higher than 0.1 for the trait A to reach fixation. In the other cases, any amount of $\alpha > 0$ is sufficient.

The results for the simulations mixing content preference and demonstrator-based social influence,

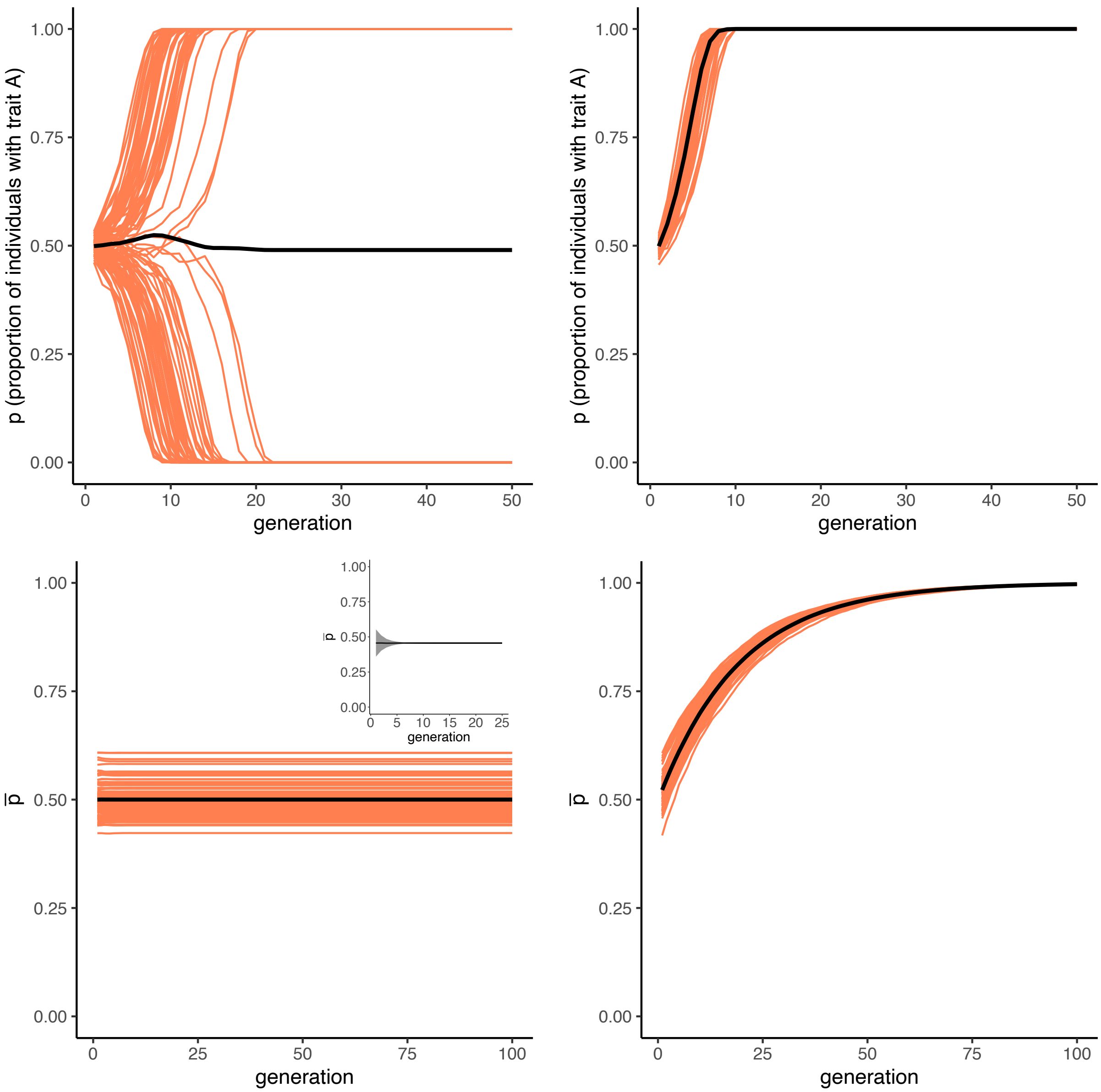


Figure 1: Simulations mixing frequency-based social influence (conformity) and content preference Each line represents a run, the thick black line is the average. Top panel: Proportion of population with trait A in the discrete traits condition. Top-left: only conformity acting ($\alpha = 0$). Top-right: content preference and content preference ($\alpha = 0.1$). For all simulations in the discrete traits condition, $T = 50, D = 1, p_0 = 0.5$ Bottom panel: Average trait value in the continuous traits condition. Bottom-left: only conformity acting ($\alpha = 0$). Inset plot: the shaded area shows the variance for a representative run. Bottom-right: conformity and content preference ($\alpha = 0.1$). For all simulations in the continuous traits condition, $T = 100$. For all simulations $N = 1000$, and 100 runs per condition.

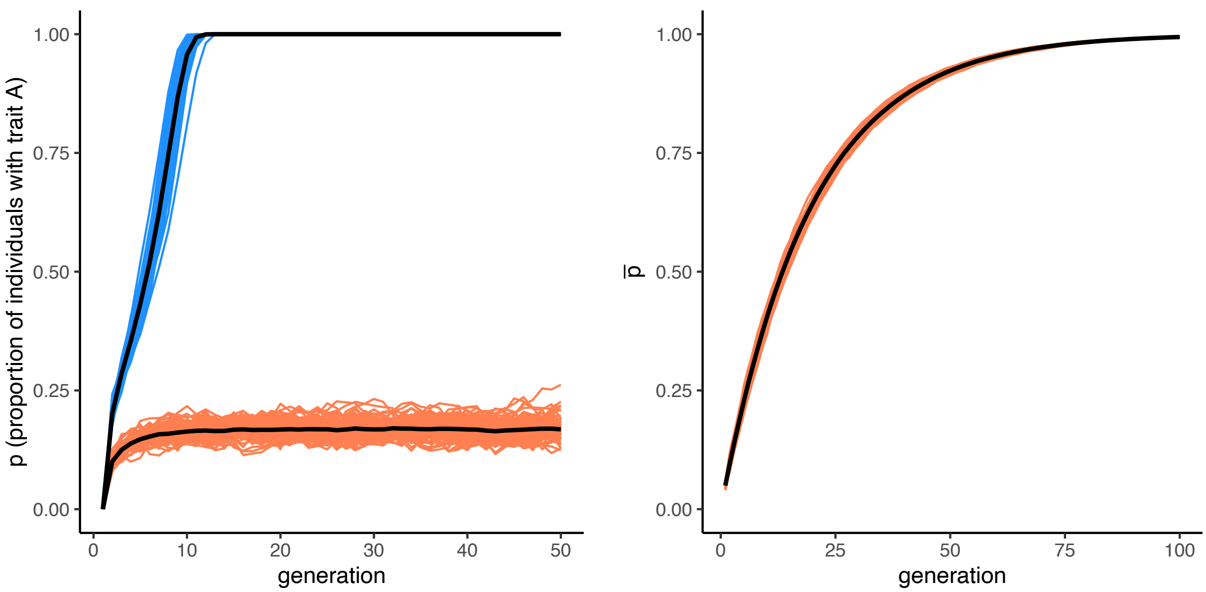


Figure 2: Simulations mixing frequency-based social influence (conformity) and content preference. Each line represents a run, the thick black line is the average. Left panel: Proportion of population with trait A in the discrete traits condition. Conformity and content preference starting from a population with only trait B ($p_0 = 0$). Red lines: runs with $\alpha = 0.1$ Blue lines: runs with $\alpha = 0.2$. For all runs: $T = 50, D = 1$. Right panel: Average trait value in the continuous traits condition. Conformity and content preference starting from a population where all individuals are initialised with $p = 0$. For all runs: $T = 100, \alpha = 0.1$ For all simulations $N = 1000$, and 100 runs per condition.

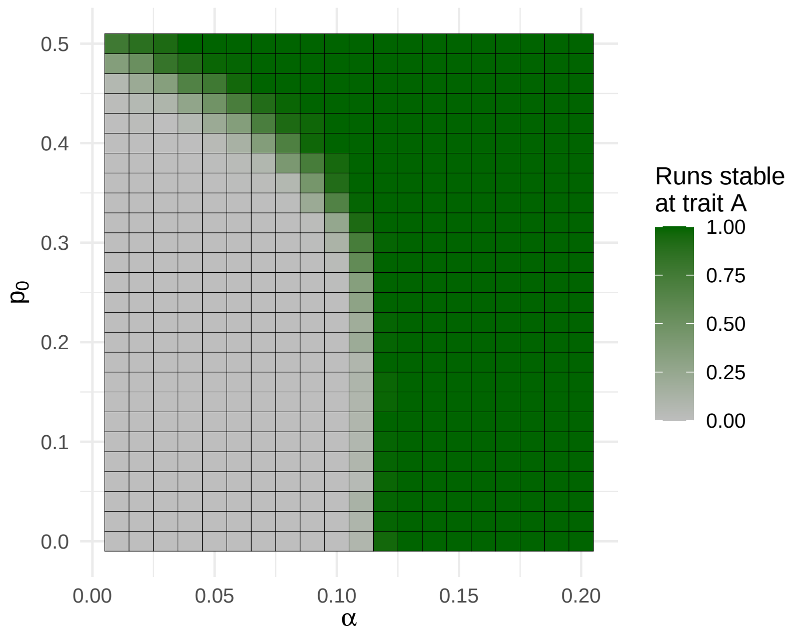


Figure 3: Simulations mixing frequency-based social influence (conformity) and content preference n in the discrete traits condition. The plot examines the results for $0.01 \leq \alpha \leq 0.2$ and $0 \leq p_0 \leq 0.5$ and shows the proportion of runs in which the trait A reach fixation (out of 100 for each parameters combination). For all simulations, $T = 50, D = 1, N = 1000$.

such as prestige, are broadly similar. Figure 4 shows representative results giving a relatively small amount of individuals with high status ($C_s = 0.01$) and $C_{copy} = 5$. As for conformity, the discrete choice without content preference produces a roughly equal distribution of runs where either A or B are fixed in the population, depending on their relative majority in the pool of "high status" demonstrators (top-left panel). As before, when content preference acts, runs converge to an equilibrium with only trait A, as long as enough high status individuals adopt it (top-right panel).

In the continuous traits condition, populations converge on one of the high status demonstrator' traits when content preference is not present (bottom-left panel), and on all individuals having $p = 1$ when it is (bottom-right panel).

With demonstrator-based social influence, all parameters combinations generate, sooner or later, an equilibrium where all individuals have trait A, or $p = 1$ in the continuous case. Even with $p_0 = 0$ in the discrete condition (all individuals starting with trait B, i.e., the analogous of the case discussed above for conformity, when $\alpha = 1$ was not sufficient to overrun the initial majority), an equilibrium

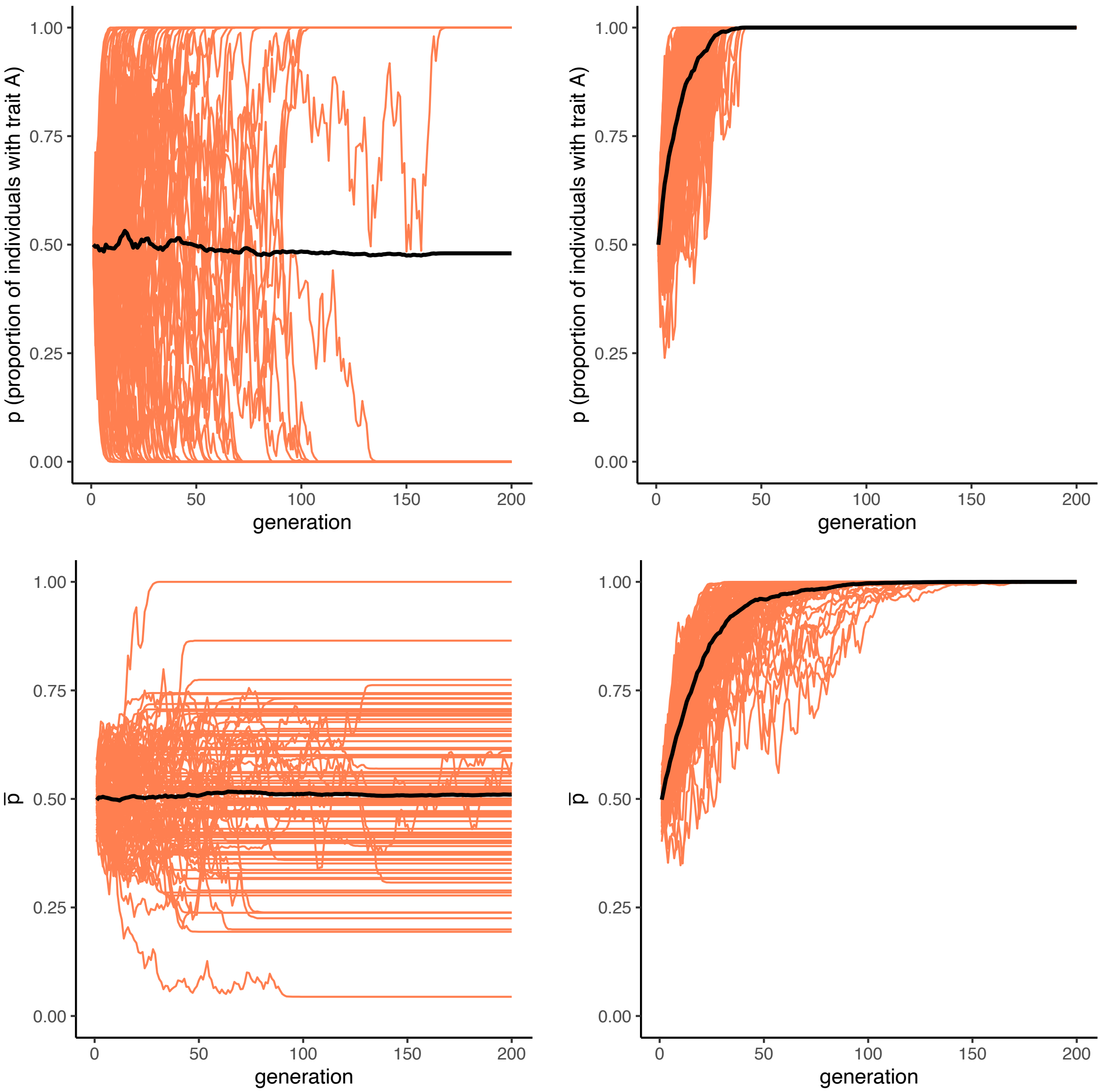


Figure 4: Simulations mixing demonstrator-based social influence and content preference. Each line represents a run, the thick black line is the average. Top panel: Proportion of population with trait A in the discrete traits condition. Top-left: only demonstrator-based social influence acting ($\alpha = 0$). Top-right: demonstrator-based social influence and content preference ($\alpha = 0.1$). For all simulations in the discrete traits condition, $D = 1$ and $p_0 = 0.5$ Bottom panel: Average trait value in the continuous traits condition. Bottom-left: only demonstrator-based social influence acting ($\alpha = 0$). Bottom-right: demonstrator-based social influence and content preference ($\alpha = 0.1$). For all simulations $N = 1000$, $T = 200$, $C_s = 0.01$, $C_{copy} = 5$, and 100 runs per condition.

with all As is reached. It is useful to examine the relation between the parameters regulating the strength of the demonstrator-based social influence ($C_s$ and $C_{copy}$) as they point out a general effect, that may appear counter-intuitive: the stronger the bias, the stronger the tendency to reach the equilibrium where the content preference points to. Figure 4 shows that, for simulations lasting $T = 200$, in the discrete traits case, trait A-equilibrium is reached more frequently with higher values of $C_{copy}$ (that is, when the probability of copying high status individuals is increased with respect to the probability of copying low status ones), and, with a small effect, with lower values of $C_s$ (i.e., when the pool of high status demonstrators is relatively smaller). (As noticed above, in all cases the equilibrium will be reached given enough time.)

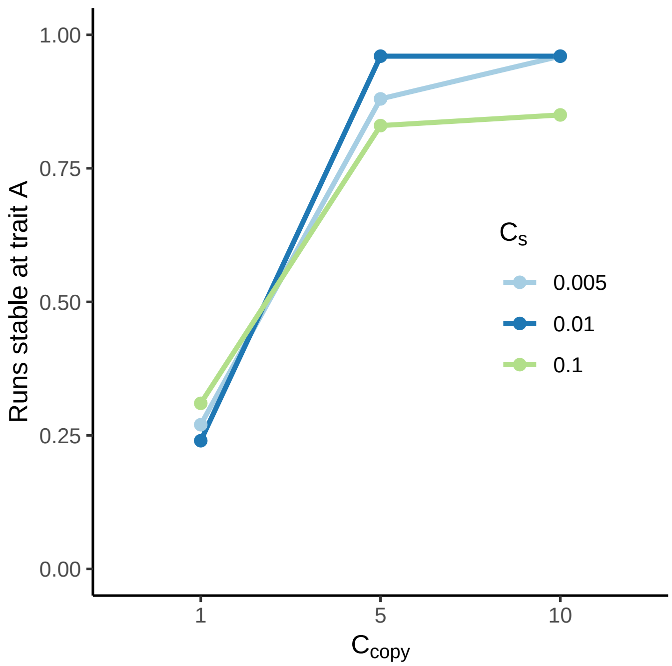


Figure 5: Simulations mixing demonstrator-based social influence and content preference in the discrete traits condition. Proportion of runs in which the trait A reach fixation (out of 100 for each parameters combination) for simulation with $T = 200$. For all simulations $\alpha = 0.1$, $C_s = (0.005, 0.01, 0.1)$, $C_{copy} = (1, 5, 10)$, and $N = 1000$.

## Discussion

The results of the simulations show that, in virtually all scenarios, content preferences stabilise culture at the point where they are directed to. When social influence mechanisms act alone, they also bring traits to fixation, but which trait/trait configuration wins is left to chance. Social influence mechanisms—both based on frequency (conformity) and on demonstrators' features (e.g., prestige bias)—, are independent from traits' features, and non-directional, so that content preferences are the only directional forces, and their existence is sufficient to determine the fate of the system. Only in one case, that is, when the majority of the population adopts the non-favoured trait in a discrete choice and conformity acts, the preference for the (minority) traits need to be sufficiently strong to overcome the initial majority (see Figure 3).

As mentioned in the Introduction, few works have explicitly addressed these questions, but the results presented here are consistent with suggestions coming from other cultural evolution models, and hopefully give them a more general background. Acerbi et al. (2021), for example, found that convergent transformation drives cultural dynamics when acting together with unbiased copying (and, similarly to here, the more faithful the copying is, the stronger the effect of convergent transformation). Morgan and Thompson (2020) found that even weak priors render conformity unable to stabilise traditions and determine the outcome, in most conditions for discrete choices, and always for continuous choices (these results are echoed in the more recent Yan, Mathew, and Boyd (2023)).

The obvious take home message of these results is that, if our question is why some things are culturally successful and others are not, weak but stable non-social forces need to be taken into account. A possibly less obvious take home message concerns the interpretation of cross-cultural regularities. The existence of strong human universals (Brown 1991) is sometime interpreted, for example by evolutionary psycohlogysts, as supporting the existence of strong cognitive evolved dispositions, or strong ecological constraints, and indicating a somehow limited role of culture. On the other end of the spectrum, socio-cultural anthropologists have tended to diminish the importance of cross-cultural regularities, to underline the importance of culture. While everyone would agree this is a false dichotomy (see e.g., (Barrett 2015)), these results suggest a way to understand why it is so: very weak directional, non-social forces, as long they are stable enough, can produce strong regularities. These can be (possibly weak) cognitive priors, physical affordances, relatively stable ecological conditions (such as the availability of certain materials), and so on.

Conceptually, it is important to think to social influence and preferences for content not as opposing forces. Non-directional social influence provide, so to say, strength, to the weak but directional preferences for content (as illustrated in Figure 5). In other words, culture magnifies individual-level tendencies, allowing them to become stable at population level. In the models presented here, the content preference is uniform in the population, i.e., only one preference was considered, but the same logic applies to more realistic situations with many different forces, and we would expect culture to homogenise population towards the stronger ones. in addition, the preference for the content is, in the models, deterministic: individuals that are subjected to it (according to $\alpha$) necessarily switch to the preferred trait. Again, this is a simplification, but a probabilistic implementation would not change qualitatively the results, only possibly producing longer times to reach fixation at the preferred trait.

Finally, the modelled social influence mechanisms are fully detached from the content of the traits. In reality, we expect that, for example, prestigious individuals would possess, on average, more adpative traits than individuals chosen at random, or that the majority targeted by conformist copying would effectively pool information from individual learning. The models above represent a "pure" social influence scenario, and are hopefully useful to strengthen our intuitions in this case.

## Acknowledgements

James Winters suggested to implement copying probabilities in the demonstrator-biased model with a softmax function.

# References


Acerbi, Alberto, Mathieu Charbonneau, Helena Miton, and Thom Scott-Phillips. 2021. "Culture Without Copying or Selection." *Evolutionary Human Sciences* 3: e50. https://doi.org/10.1017/ehs.2021.47.

Acerbi, Alberto, Alex Mesoudi, and Marco Smolla. 2022. *Individual-Based Models of Cultural Evolution: A Step-by-Step Guide Using r.* London: Routledge. https://doi.org/10.4324/9781003282068.

Barrett, H. Clark. 2015. "The Shape of Thought," January. https://doi.org/10.1093/acprof:oso/9780199348305.001.0001.

Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process.* Culture and the Evolutionary Process. Chicago, IL, US: University of Chicago Press.

Brown, Donald E. 1991. *Human Universals.* new York: McGraw Hill.

Claidière, Nicolas, Gameli Kodjo-kuma Amedon, Jean-Baptiste André, Simon Kirby, Kenny Smith, Dan Sperber, and Joël Fagot. 2018. "Convergent Transformation and Selection in Cultural Evolution." *Evolution and Human Behavior* 39 (2): 191–202.

Claidière, Nicolas, and Dan Sperber. 2007. "The Role of Attraction in Cultural Evolution." *Journal of Cognition and Culture* 7: 89–111.

Henrich, Joseph. 2016. *The Secret of Our Success.* Princeton, NJ: Princeton University Press.

Henrich, Joseph, and Robert 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.

Jiménez, Ángel V., and Alex Mesoudi. 2019. "Prestige-Biased Social Learning: Current Evidence and Outstanding Questions." *Palgrave Communications* 5 (1): 112. https://doi.org/10.1057/s41599-019-0228-7.

Kendal, Rachel L, Neeltje J Boogert, Luke Rendell, Kevin N Laland, Mike Webster, and Patricia L Jones. 2018. "Social Learning Strategies: Bridge-Building Between Fields." *Trends in Cognitive Sciences* 22 (7): 651665.

Mesoudi, Alex. 2009. "The Cultural Dynamics of Copycat Suicide." *PLOS ONE* 4 (9): e7252. https://doi.org/10.1371/journal.pone.0007252.

———. 2021. "Cultural Selection and Biased Transformation: Two Dynamics of Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1828): 20200053. https://doi.org/10.1098/rstb.2020.0053.

Morgan, Thomas J. H., and Bill Thompson. 2020. "Biased Transformation Erases Traditions Sustained by Conformist Transmission." *Biology Letters* 16 (11): 20200660. https://doi.org/10.1098/rsbl.2020.0660.

Morin, Olivier. 2015. *How Traditions Live and Die.* London & New York: Oxford University Press.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. "Epistemic Vigilance." *Mind & Language* 25 (4): 359–93. https://doi.org/10.1111/j.1468-0017.2010.01394.x.

Sperber, Dan, and Lawrence A. Hirschfeld. 2004. "The cognitive foundations of cultural stability and diversity." *Trends in Cognitive Sciences* 8 (1): 40–46. https://doi.org/10.1016/j.tics.2003.11.002.

Yan, Minhua, Sarah Mathew, and Robert Boyd. 2023. ""Doing What Others Do" Does Not Stabilize Continuous Norms." *PNAS Nexus*, February, pgad054. https://doi.org/10.1093/pnasnexus/pgad054.