BRILL

Journal of Cognition and Culture 7 (2007) 89-111

www.brill.nl/jocc

Commentary
The role of attraction in cultural evolution*

Nicolas Claidière
Dan Sperber
Institut Jean Nicod, Paris
nicolas.claidiere@normalesup.org; dan@sperber.com

Abstract
Henrich and Boyd (2002) were the first to propose a formal model of the role of attraction in cultural evolution. They came to the surprising conclusion that, when both attraction and selection are at work, final outcomes are determined by selection alone. This result is based on a determistic view of cultural attraction, different from the probabilistic view introduced in Sperber (1996). We defend this probabilistic view, show how to model it, and argue that, when both attraction and selection are at work, both affect final outcomes.

Keywords
Attraction, cultural evolution, cultural transmission, selection

Two naturalistic research programmes relevant to the explanation of cultural phenomena that started in the 70s – the evolutionary approach of Boyd and Richerson (1985, Richerson and Boyd 2005), and their collaborators, and the cognitive approach of Atran (1990, 2002), Boyer (1994, 2001), Hirschfeld (1996), Sperber (1996), and their collaborators – have to a certain extent converged over the years, the first, more evolutionary programme going into greater detail into the cognitive bases of cultural evolution, and the second, more cognitive programme paying an ever increasing attention to the evolution of mind and culture. Part of the reason why this relative convergence went almost unnoticed is the fact that these programmes were generally pursued in mutual ignorance

* Reply to J. Henrich and R. Boyd, "On modeling cognition and culture", *Journal of Cognition and Culture*, 2 (2), 2002). We thank Rob Boyd for very useful comments on an earlier version of this article.

 DOI: 10.1163/156853707X171829

with no discussion of the work in the other tradition, or, worse with misrepresentation, as when Sperber extended his criticisms addressed to Dawkins and memeticists to the work of Boyd and Richerson without attending to the relevant differences between these two approaches.

In their article "On modeling cognition and culture," Henrich and Boyd (2002) open a serious discussion of the cognitive approach. They overestimate, however, the points of divergence: we happen to agree with much of what they present as objections. Let us, to illustrate this point, add comment in square brackets and in italics to their concluding paragraph:[1]

> The crux of Sperber, Atran and Boyer's position is that the transmission of culture requires domain specific cognitive mechanisms [*yes, with qualifications*], and that therefore population dynamic models of culture proceed from untenable assumptions [*some population dynamic models, memetic ones in particular, proceed from untenable assumptions, but they need not; what we want is to contribute to improving these models, not reject them*]. We accept that social learning, like all other forms of learning, requires innate expectations about objects in the environment and the nature of relationships among them. How these innate structures shape the human mind is obviously of great importance for understanding human culture. The mistake is to see these ideas as incompatible with making population dynamic models of cultural change [*this is a mistake we have never been tempted to make*]. It will never be enough to focus on the mind and ignore the interactions between different minds [*of course*]. To keep track of such interactions some kind of population dynamic models will be necessary. What is needed is both more effort by coevolutionary theorists to incorporate rich cognition into formal models of social learning, and more effort by cognitive scientists to consider how innate cognitive structure interacts with social processes and the cognition of social learning to influence the epidemiology of representations and its associated behavioral products [*total agreement*].

Henrich and Boyd article presents and discuss three models. The second and the third models illustrate the claims that population-scale conformity-biased and prestige-biased transmission can play a role in compensating for high error rates in inter-individual transmission and in securing adaptive cultural evolution, and that discrete units of transmission are not necessary for this to happen. Contrary to what Henrich and Boyd seem to expect, we[2] are in general agreement with these claims.

[1] We discuss the views of Boyd and Richerson in greater detail in Sperber and Claidière (in press).

[2] We cannot speak for Atran and Boyer whom Henrich and Boyd also cite, but we don't believe that their views are importantly different from ours on the issues at hand.

Still, there is an important point of disagreement between Henrich and Boyd and us regarding the respective roles of attraction and selection in cultural evolution. They argue, with the use of the first model presented in their article, that, to put it succinctly, in cultural evolution, selection trumps attraction. We reply that what looks like a demonstration is in fact based on quite inadequate modeling of attraction. Our response is in two parts, a first part where the arguments are presented informally, and a second, more formal part presenting and discussing models and simulations.

1 – The arguments

The idea of cultural attraction was introduced in Sperber 1996, ch. 5. It is intended to help reconcile two observations:

1) at the micro-level, transmission of information among humans is generally not a copying process and typically results in modifications of the information transmitted;
2) at the macro-level, cultural information is relatively stable within whole populations and often across generations.

The micro-processes of transmission are not faithful enough to come near explaining this macro-stability (unlike the faithfulness of gene replication that does provide the core of the explanation of the relative inertia of gene pools).

As we just mentioned, the approach defended by Henrich and Boyd identifies mechanisms – conformity-biased and prestige-biased transmission – that can contribute to the explanation of this macro-stability. These mechanisms tend to favor some cultural contents not because of properties of these contents, but because of their distribution in the population either as contents adopted by the majority, or as contents adopted by the most prestigious individuals. The idea of attraction, on the other hand, aims at explaining the relative prevalence and stability of cultural contents as a function of properties of the contents themselves. We believe that both kinds of phenomena – distribution-based transmission biases and content-based attraction – play a role in explaining cultural stability and evolution, and we leave for another occasion the discussion of what their respective roles might be.

Here is an account of the idea of cultural attraction simplified as much as possible for the purpose of this discussion. 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 construct a variant of her own. This variant is likely to depart from the variants on which it is based both because some information may be lost in the process, and because the goal of acquisition is generally to acquire not a replica of other people's variants, but, rather, a piece of knowledge or a skill that suits the individual own dispositions and preferences. It would be misleading therefore to talk of these departures from model or models in cultural transmission as "failures to replicate", "mutations", or "noise". Even if these departures from the model often do involve poor cognitive or behavioral performance, they occur not as accidents or malfunctions but as normal outcomes of the constructive processes involved in cultural transmission.

If each individual variant of a cultural item departed at random from the variants that had inspired it (and in the absence or insufficiency of compensating factors such as the biases described by Henrich and Boyd), it is hard to see how cultural items would ever reach the minimal level of stability within a population over time without which the very notion of culture does not make sense at all. If, on the other hand, individual variants do not depart at random from their model, but tend to gravitate around the same positions in the space of possibilities, then, even without any strict replication ever, one would end up with clusters of cultural items around these attractors and therefore at least the modicum of stability that culture presupposes.

Attractors as points or areas in the space of possibilities are abstract objects similar in this respect to proportions or centers of gravity. They exist because there are concrete factors of attraction that affect the probability that individual variants of a cultural item will depart from their models in one direction rather than in another and that cause all the variants of a given item to gravitate around the same point. Factors of attractions can be of different kinds. At the most general level, they may have to do with psychological dispositions or with environmental constraints and affordances (contrary to what Henrich and Boyd suggest, it has never been part of the theory that factors of attraction should be exclusively cognitive). Attractors themselves can and do change over time as an effect of the factors that explain them, but they change in historical time, that is, slowly enough to uphold the relative stability of culture.

To illustrate in the simplest possible way (and in a manner that will help us discuss Henrich and Boyd's model) the idea of attraction and its relationship to replication and selection, consider a schematic version of the evolution of cigarette consumption in a population (see figure 1a – this is not meant to be realistic, but just to make the idea more concrete, and the presentation in the text of the article will be informal, with formal details presented in Appendix 1). Members of some population smoke each between zero and 30 cigarettes per day, so there are 31 variants of their smoking pattern. Every year, a new age cohort of youngsters joins this population and select, from among the members of the

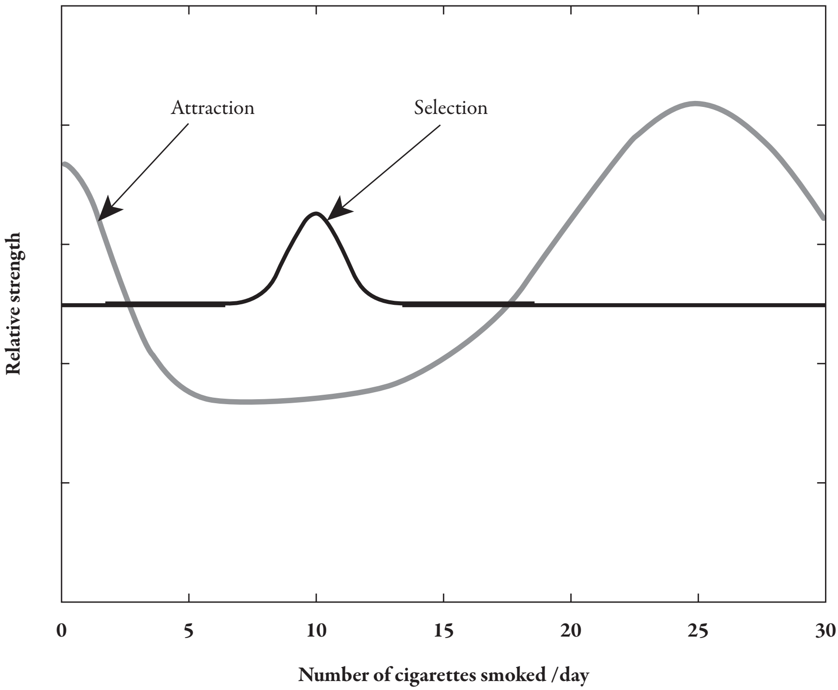

Figure 1a. The cigarettes model, with two peaks of attraction and one peak of selection (details in Appendix 1)

cohort just above them, a person whose smoking pattern they would like to adopt. Depending on their smoking pattern, some people have a greater probability than others of being selected as models to imitate. More specifically, let us assume that relatively light smokers who smoke 10 cigarettes a day are the people most likely to be selected as models. This probability of an individual being selected as model given his or her smoking pattern is represented in figure 1a as a black curve.[3] New smokers, however, end up, in less than a year, with a variant that may differ from that of the model they selected. This is so for a variety of reasons, in particular because of the lack of correct estimation of the smoking pattern of the people they chose to imitate, because of carelessness in imitative behavior, and, above all, because of the fact that smoking is an addictive acquired

[3] Incidentally, when we speak of "selection" here we refer, as do Henrich and Boyd, to the probability of being selected as a model, and to nothing else. Selection in this sense is independent of fidelity in copying the model and differs therefore from Darwinian selection, which presupposes a rate of mutation much lower than the selection bias.

taste so that people tend either not to smoke at all or to smoke more cigarettes than they intended to.

People's smoking pattern is likely to depart from the variant they selected not at random, but, we assume, in the direction of one of two attractors. One attractor is abstinence, or zero cigarette, and the other, based on the addictive properties of tobacco, is at 25 cigarettes per day. The 0-cigarette attractor has a strong effect on people who choose to imitate non-smokers and who tend to remain non-smokers themselves, and also on people who select as models smokers of one to five cigarettes per day, and who are likely to end up as non-smokers. So, the 0-cigarette variant is a very strong but very local attractor. Even so, some people decide to imitate a non-smoker but end-up, through weakness of will, becoming smokers themselves. Attraction is probabilistic. The 25-cigarettes attractor is also quite strong and has much wider effect. The people who select as models smokers smoking from 7 to 30 cigarettes per day tend to end up smoking a number of cigarettes between the variant they selected and 25. Even so, some people who decide to imitate a light or even a heavy smoker end up non-smokers. Again, this is an improbable but not an impossible outcome. The attractive force of different smoking patterns is represented in figure 1a as a grey curve.

This toy model illustrates several interesting properties and cases:

1) The curve of attraction indicates probabilities of transformation in one direction rather than another.
2) A curve flat on both side of a given variant (as around the 7-cigarettes variant) indicates that transformations in either direction are equally probable.
3) A curve slanted in the same direction on both sides of the variant indicates that the variant is more attractive than variants on the descending side and less attractive than variants on the ascending side (as for, say, 15).
4) An attractor is a peak in the curve of attraction, such that the neighboring variants on both sides (or just on one side, if it is at one end of the range of possibilities) are less attractive than it is (as for 0 and 25).
5) An attractor with very steep curve on both sides (or just on one side, if it is at one end of the range of possibilities) indicates that when this variant is selected as a model, it is very likely to be replicated. In other terms *a very steep attractor is equivalent to a replicator* (as for 0).

Imagine that each age cohort has 310 members and that, in the initial cohort at time t_0, each of the 31 variants is followed by exactly 10 people. We can ask how the relative success of each variant will evolve with successive cohorts. If there was only attraction and no selection, we would expect after some time the distribution of smoking patterns to correspond to the attraction curve. A simulation

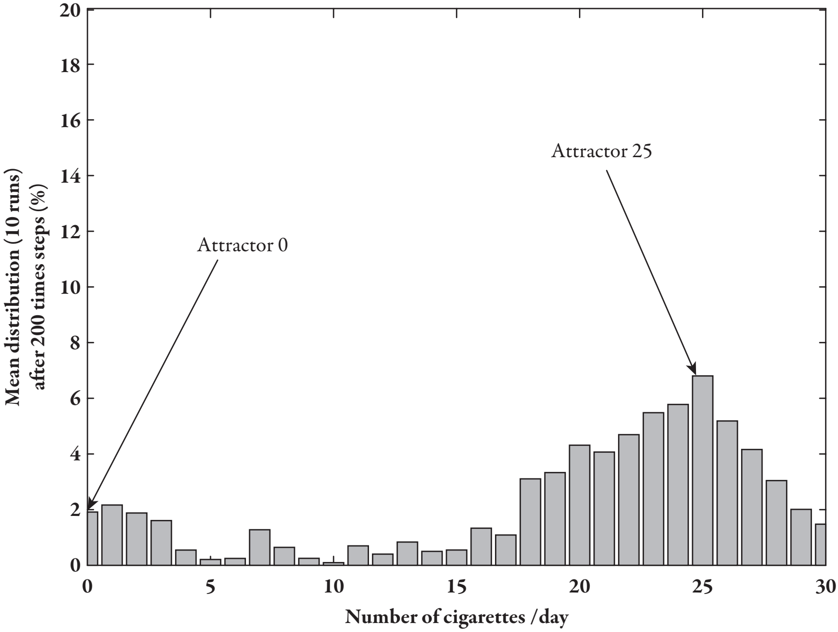

Figure 1b. The cigarette model with attraction and without selection: distribution of the population after 200 steps (details in Appendix 1)

with 200 time steps and 10 runs confirms this prediction (see figure 1b). If there was only selection, no attraction, and accurate copying of the model, we would expect to find that, after a few time steps, the population is concentrated at the selection peak of 10 cigarettes/day, and this is indeed what we found (this result being trivial, the data is not shown). On the other hand, if there was selection but inaccurate copying of the model, we would expect to find most of the population concentrated around the selection peak and this is what we found (see figure 1c).

The more interesting situation is that where both attraction and selection are at work. Imagine that, in such a situation, we track the "descendants" – descent being through selection as a model – of an individual A smoking 8 cigarettes a day. We might observe that, because selection at this point is quite strong, 2 individuals in the second age cohort, B and C, select A as a model. Because, at that point, attraction is nearly symmetrical B might end up smoking 5 cigarettes, and C 10 cigarettes. Now, a third age cohort arrives and, because selection is lower for 5 cigarettes than for 10, only one individual, D, might select B (who smokes 5 cigarettes) as model, and 3 other individuals, E, F, and G, might select C (who smokes the 10 cigarettes). D, imitating the 5 cigarettes pattern, might

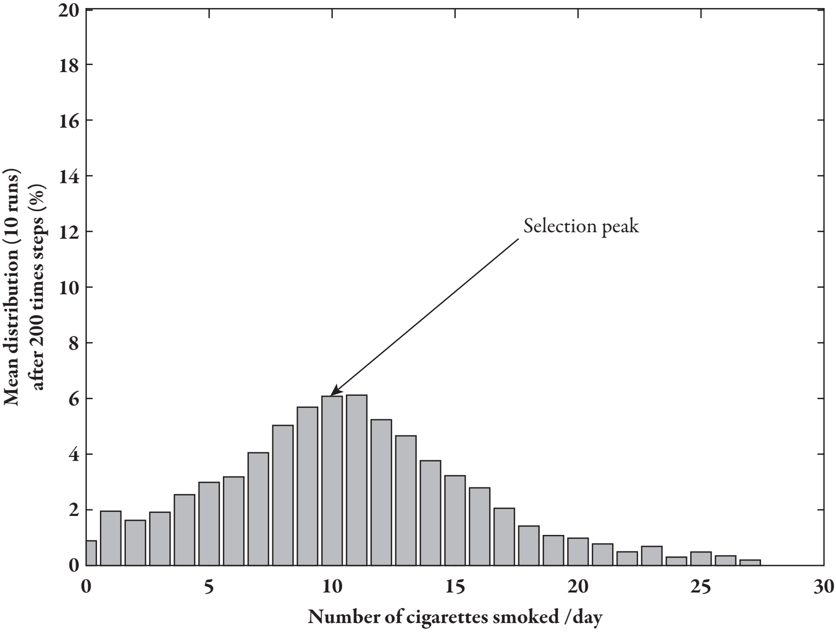

Figure 1c. The cigarette model with selection and inaccurate copying, and without attraction: distribution of the population after 200 steps (details in Appendix 1)

end up smoking 0 cigarette since attraction toward 0 is high at that point. E, F, and G, imitating the 10 cigarettes pattern, might end up smoking 13.8, and 12 cigarettes respectively because attraction is relatively flat at that point. With such lines of descent, we should not be surprised if both selection and attraction had an effect of the distribution of the population among the various smoking patterns, with the 10-cigarettes pattern being better represented than if there were no selection, and the 0 and 25 patterns, and those in their neighborhood being better represented than if there was no attraction. This is indeed what we found (see figure 1d). Of course, with different parameters, we might render the effect of selection or those of attraction negligible, but the point we have illustrated so far is that, in principle, when both attraction and selection are at work, they may both have noticeable effects on the distribution of variants in the population.

Even without this example, it seems intuitively implausible that, when both attraction and selection are involved in a cultural evolution process, only attraction or only selection should systematically determine the final outcome. Henrich and Boyd claim however to have demonstrated that, in particular when attaction is strong, the final outcome is determined by selection alone.

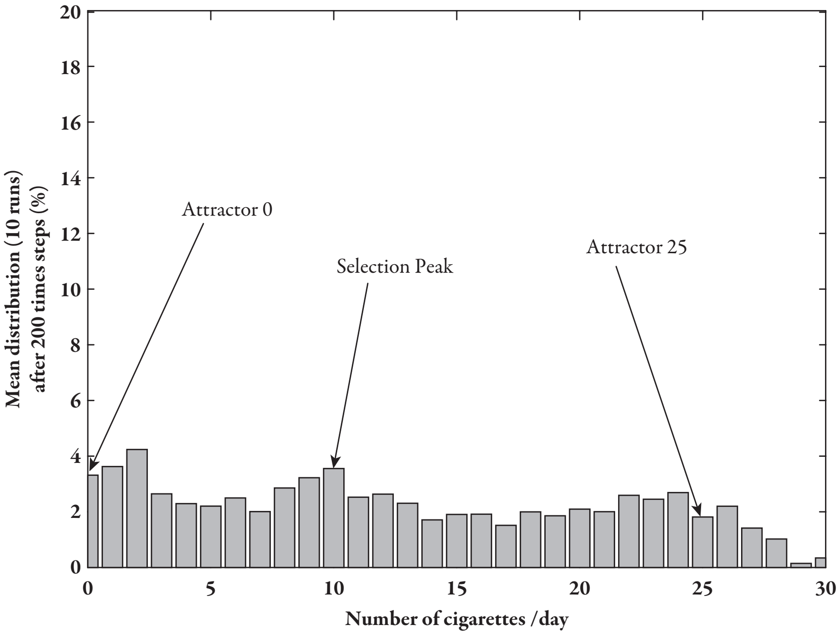

Figure 1d. The cigarette model with both attraction and selection: distribution of the population after 200 steps (details in Appendix 1)

Henrich and Boyd, while granting the reality of attraction, suggest that the dynamics of cultural evolution reduce to that of replication and selection where selective forces determine the ultimate outcome. If this were correct, the notion of attraction might still be relevant to a detailed description of the processes involved – and in particular, as we will see, of its initial stages –, but not to modeling the dynamics of cultural evolution. The argument is based on the use of a formal model that scholars interested in culture and cognition but with no competence in modeling may not have fully understood, let alone felt confident enough to evaluate. They may have been left with the idea that a demonstration had been given of a surprising and even paradoxical conclusion that would severely limit the claim of relevance to cultural evolution of the cognitive approach. This is not so. It is not so, to begin with, because such models cannot yield such decisive conclusions. They are great tools for asking novel questions about cultural evolution, imagining possible answers, and sharpening our conceptual tools. They allow demonstrations of what happens *in the model*. On the other hand, in the absence of a clear methodology for judging the fit between the model and the reality it purports to represent and to test non-trivial predictions

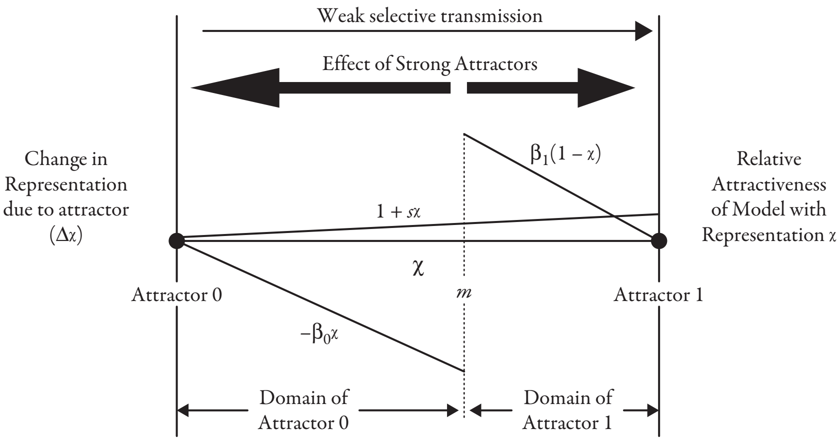

Figure 2. Henrich and Boyd's model. Detailed description in section 2

of the model on the basis of (preferably quantifiable) empirical evidence, these models don't demonstrate or even provide compelling argument about what is actually the case in the real world. This should not be understood as a criticism, but as a reminder. So, even if the model used by Henrich and Boyd were adequate, what it would show – and this would be interesting enough – is that attraction *might* work in a manner such that, quite generally, its effects on cultural dynamics would collapse into those of replication plus selection. As it happens, their model is, we believe, based on misunderstandings and is not a good tool to explore the issue.

Henrich and Boyd's model assumes a population whose members hold mental representations the content of which is a value x represented by real numbers between 0 and 1. During each time period, people in the population choose each an individual as their model and try to acquire his or her representation. However people's construal of this representation is biased towards one of two attractors, which are situated at the two ends of the continuum, i.e. at 0 and at 1. There is an arbitrary cut-off point m between 0 and 1 such that, when the variant selected has a value between 0 and m, people invariably end up with a representation that is closer to 0 than the variant selected, and when the variant selected has a value between m and 1, people invariably end up with a representation that is closer to 1 than the variant selected (see figure 2 reproduced from Henrich and Boyd's figure 1)

To make all this a bit more concrete, let us translate this into a version of our cigarette model (we take it that the fact that one model involves a continuous variable between 0 and 1 and the other 31 discrete variants between 0 and 30 is irrelevant to the issue at hand). We have the same general situation regarding the transmission of smoking patterns as in our initial model, but there are only two attractors at 0 and at 30 cigarettes, and there is a cut-off point at, say, 17 cigarettes. People who decide to imitate someone who smokes less than 17 cigarettes end up smoking even less than their model, whereas people who choose to imitate someone who smokes 17 or more cigarettes ends up smoking more than their model. There is no probabilistic element left regarding the direction of attraction. Attraction is wholly in one direction or wholly in the other. The population is therefore partitioned into two groups, those under the 17-cigarettes threshold who are attracted towards 0, and those at or above this threshold who are attracted towards 30.

Whereas in our initial model, anyone at any variant could be attracted in either direction and just the probability of transformation in one direction rather than the other changed from one variant to another, here the direction of transformation is a sure thing. This is not strong probabilistic attraction, but deterministic attraction. Departing from Sperber's notion of "attraction" defined in terms of greater probabilities of transformations towards, rather than away from a given point or "attractor" (Sperber 1996:112), Henrich and Boyd's understanding of "attraction" is not probabilistic but determistic (an understanding possibly "attracted" towards the standard deterministic notion of "attraction" in systems dynamics). They do talk of stronger or weaker force of attraction, but actually, what they mean by "force" of attraction is not the relative probability of departing from the model in one direction rather than another, but the variable size of the departure from the model always in one and the same direction, that of the attractor. With a "stronger attractor" so understood descendents of a given variant will reach the attractor in fewer steps than with a "weaker attractor", but, in any case, after a shorter or longer time interval, all items will be at an attractor, and there will be no role left for attraction.

Deterministic cultural attraction is to regular, probabilistic cultural attraction what black holes are to regular physical attraction. Nothing ever gets out of a black hole. No line of cultural descent ever moves in any direction other than that of its attractor. The descendants of variants below 17 cigarettes will, after a few time periods, end up non-smokers and stay so forever. The descendents of variants at or above 17 cigarettes will, after a few time periods, end up at 30 cigarettes per day and stay there forever. As we noted, very steep attraction – i.e. a much higher probability of change in one direction rather than the other – culminates in attractors that are equivalent to replicators. In Henrich and Boyd's model not only are the two end points, 0 and 1 (or, in our cigarette version of their model, 0 and 30) perfect replicators, but so are also two other, less obvious traits, that of being attracted towards 0 and that of being attracted towards 1

(0 and 30 in our version). No wonder that replicator dynamics seems uniquely relevant to the evolution of the model!

What about selection in Henrich and Boyd's model? They assume that, in selecting whom to emulate, individuals are likely to prefer someone whose representation has a higher value. The selective force increases continuously from 0 to 1. As a result, people whose representation has a value above *m* are all more likely to be selected as models to be imitated than any people whose representation has a value below *m*, and people altogether most likely to be selected as models are those with the representation 1, which also happens to be an attractor. Translating into the cigarette model, this would mean that the greater the number of cigarette an individual smokes, the greater his or her likelihood to be imitated, with selective force, i.e. the probability of being imitated, peaking at the maximum number of 30 cigarettes per day. All variants at or above 17-cigarettes would be more likely to be selected than any variant under that threshold.

Henrich and Boyd's model has three relevant peculiarities:

1) The variants in the model fall into two groups, above and below a threshold, and the trait of belonging to one or the other of these two groups strictly replicates.
2) Attraction is determistically towards 0 in the group below the threshold, and towards 1 in the group above the threshold, which the effect that 0 and 1 are strict replicators.
3) Selective force is wholly in favor of the upper group and peaks at his attractor.

Given these three peculiarities, it should be intuitively clear that:

1) With each time period, there will tend to be more people with variants in the upper group selected as models, until all the people have variants in this upper group.
2) The variants in the upper group will evolve toward the upper attractor until this perfect replicator is the only variant represented in the population: deterministic attraction self-eliminates.
3) Moreover, if attraction is strong enough, it self-eliminates in a few steps and, from early on, the process is simply one of selection between two replicators.

So, in Henrich and Boyd model the only variant remaining in the end is 1, and in the cigarette version, it is 30 cigarettes a day. The fact that, in both versions, 0 was also an attractor does not make any difference to this ultimate outcome, since selection favors the higher group and attractor .

Henrich and Boyd used formal considerations and equations, but, in fact, their conclusions regarding what happens in their model follow quite commonsensically from plain properties of this model that can be informally understood. However, nothing of interest follows regarding the relationship between attraction and selection in cultural evolution, because what obtains in this model is an artifact linked to the peculiarities of the model. To give just one intuitive illustration of this, there is no *a priori* reason why selective force should peak at an attractor (it does not in our initial cigarette model). Imagine, then, the following variation of Henrich and Boyd's model: everything is as they describe it except that maximum selective force is at the threshold *m*, the selective force of the variants above and below the threshold have on average the same probability of being selected, and, in particular, the selective force of 1 and of 0 are equal. It should be intuitively obvious that, in this case, however strong the selective forces, it would not matter at all to the ultimate outcome, which would be exclusively determined by initial conditions, attraction, and drift (with all the descendents of variants below the threshold ending up at attractor 0, and all descendents of variants above the threshold ending up at attractor 1). If Henrich and Boyd had used this modified model (which is of course quite arbitrary, but so is their own model), and had generalized from it, they would have come to the surprising and equally unwarranted conclusion that, when you have both selective force and attraction at work, in the end, only attraction matters.

Even informally, it seems clear that the model used by Henrich and Boyd has such peculiar properties (in particular the non-probabilistic character of attraction and the coincidence of the selective peak with an attractor) that it does not help, unlike many of other models developed by Boyd, Richerson, and their collaborators (including the two other models in the article under discussion), get a better grasp of questions and possible answers in the study of cultural evolution. Henrich and Boyd's model is even less capable of giving any support to the implausible theoretical claim that, even in the presence of strong attraction, only selection determines the final outcome.

In the next section, we present a formal treatment of our arguments and show that by manipulating the parameters of Henrich and Boyd own model, one may reach very different conclusions. We first show that the results of Henrich and Boyd do not depend on what they call the force of attraction or of selection but just on the peculiarities of their model. We then extend their model and show that, when the representation most selected does not coincide with an attractor, the outcome is not anymore that predicted by selection alone. And finally, by making attraction probabilistic, we show that, in general, the outcome depends on the relative strength of both attraction and selection.

2 – Models and simulations

2.1 – *Confirming Henrich and Boyd's own results*

First we replicated Henrich and Boyd's own simulation, using the same parameters (see figure 3a). This served both to confirm their results and to establish that we were following the same procedure.

What is represented here (and in figure 2 above borrowed from Henrich and Boyd), is the evolution of a pool of mental representations in a population. The content of these representations is a real number x between 0 and 1. During each time period, people in the population observe the behavior of another individual, infer from this behavior the mental representation of the model, and adopt the mental representation they have inferred their model must have. Not all individuals have the same probability of being selected as model. Rather, the probability that an individual be selected as a model increases with the value of his or her representation and equals 1+sx. People's inferences are moreover biased towards attractors, which happen to be x = 0 and x = 1. As a result, instead of inferring the actual value of a representation x, people interpret it as having the value $x + \Delta x$. Which of the two attractors biases the interpretation of a given representation x is determined by a point m between 0 and 1 that marks the limit between the two basins of attraction of the two attractors. If x is greater than m, it is attracted toward attractor 1. If x is smaller than m, it is attracted toward attractor 0. The "force" of attraction – we have questioned this use of the notion of force in the first section and won't raise the issue again here – is expressed by a number, β_0 for attractor 0 and β_1 for attractor 1. If $x < m$, then $\Delta x = -\beta_0 x$, and if $x > m$, then $\Delta x = \beta_1(1\text{-}x)$.

Using the same parameters as Henrich and Boyd (i.e., m = 0.6, s = 0.05, $\beta_0 = \beta_1 = 0.5$, n = 200), we indeed replicate their results. The evolution of the pool of representations fits the prediction of replicator dynamics, and attraction plays a negligible role. Before reading too much into this result, one should pay attention to the two curves indicating the average value of x in group 0 (containing all and only variants below m) and in group 1 (containing all and only variants above m). They indicate that after about 10 time periods (see the shaded area), all the representations have either the value 1 or the value 0 and are not anymore subject to attraction. From the 10-steps point in the time scale, the process involves only replicators and there is no way attraction could play any role at all. Given this, the fact that the dynamics at work is plain replicator dynamics is quite trivial. As selection favor representations with value 1 over representations with value 0, in the end, all representations have a value of 1 (as can be seen from the distribution at time t = 250).

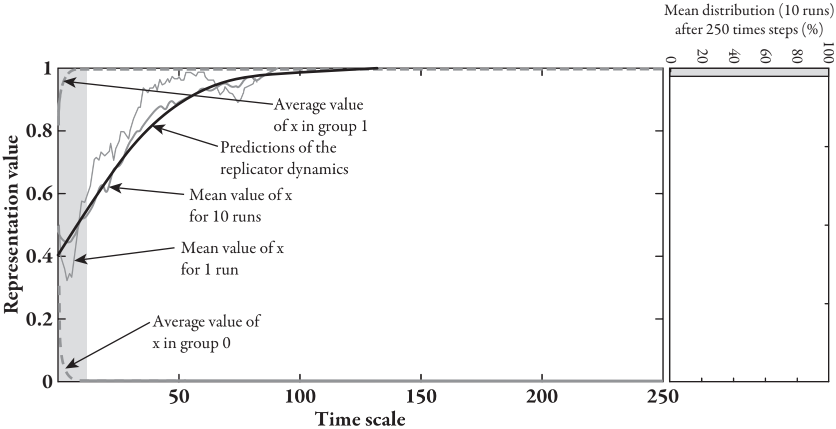

Figure 3a. Replication of the simulation used by Henrich and Boyd in support of the claim that weak selection overrides even strong attraction. The left frame represents the evolution through time of values of x as observed and as predicted by replicator dynamics with the following parameters: m = 0.6, s = 0.05, $\beta_0 = \beta_1 = 0.5$, n = 200. With these parameters, attraction self-eliminates in about 10 time steps (shaded area). Thereafter (unshaded area), only selection is at work. The right frame represents the distribution of representations after 250 time steps for the 10 simulations.

2.2 – *When attraction is weaker or when selection is stronger: Same outcome*

What would happen if attraction was much "weaker" in Henrich and Boyd sense, while still being non-probabilistic? Intuitively, it would take many more steps to eliminate the impact of attraction, but, selection would still be the sole determinant of the final outcome. We performed a simulation with the same value as before except for β_0 and β_1 which were divided by 20. As the shaded area in figure 3b shows, it does take more steps to get rid of the values between 0 and 1, and during all these steps, the dynamic of the population does not follow replicator dynamic. However, once practically all representations have values 0 or 1 and are therefore not subject to attraction anymore, the dynamics converges with replicator dynamics and the end result is solely determined by selection (see the distribution graph).

Raising the selection by increasing s does, on the other hand, make the population dynamics even closer to that of replicators, and the equilibrium is reached much faster (since this result is quite trivial, the data is not shown).

So far, our simulations show that the end result of the model of Henrich and Boyd does not depend on the force of either attraction or selection. The claim that the final outcome is determined only by selection is in fact related to two

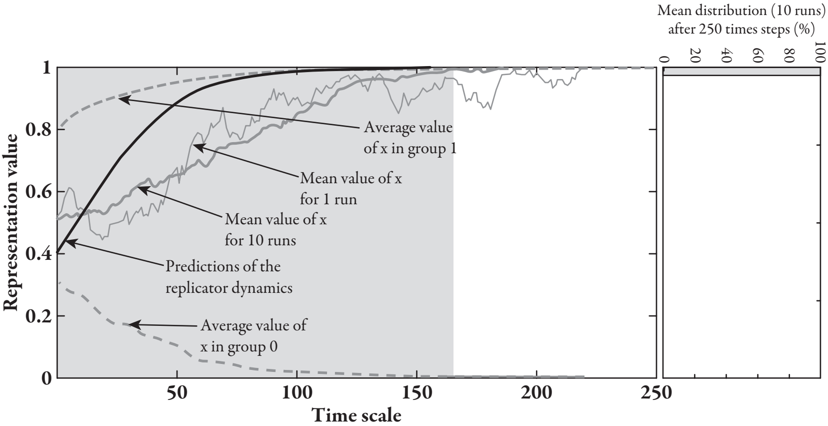

Figure 3b. If attraction is weak, it takes more steps (shaded area) for it to self-eliminate. Still, once all representations have converged to 0 or 1, selection determines the same outcome as previously. The left frame represents the evolution through time of values of x as observed and as predicted by replicator dynamics with parameters as in Fig. 3a except $\beta_0 = \beta_1 = 0.025$. The right frame represents the distribution of representations after 250 times step for the 10 simulations.

artifacts of the model: first attraction is non probabilistic and second selection happens to favor an attractor. What would happen if we altered these two special features of Henrich and Boyd's model?

2.3 – *When selection does not peak at an attractor: Different outcome*

We believe that Henrich and Boyd's would-be demonstration that selection determines the final outcome irrespective of attraction is an artifact of their choice of selective function and, even more importantly, of the non-probabilistic character of attraction in their model. We first present simulations where we leave their attraction parameters untouched but where we modify their selection function and in particular their selection peak.

There is no principled reason to assume that attractors, that is, points towards which transformations tend to be biased, should coincide with variants most likely to be selected as models. After all, in real life, people typically choose as models the most skilled performers (craftsmen, warriors, artists, and so on) even though their own performance tends to be biased towards easier and less admirable outcomes. Henrich and Boyd used a linear function of x as the selective function (viz. $w(x) = 1 + sx$) which makes the value 1, which happens to be an attractor in their model, the one most likely to be selected. To keep attraction

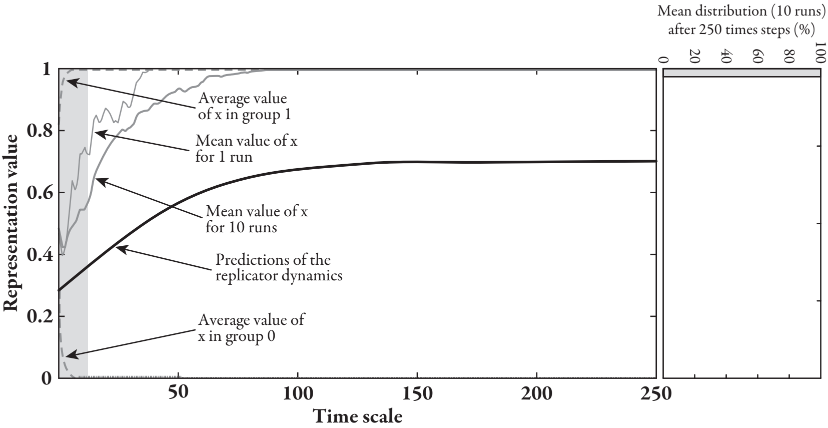

Figure 4. Selection peaks at x = 0.7, while the attractors are at 0 and 1. Because selection favors values closer to 1 over values closer to 0 the mean representation value in the population converges toward 1. The left frame represents the evolution through time of values of x as observed and as predicted by replicator dynamics with the following parameters: $\mu = 0.70$, $\sigma = 2$, $\beta_0 = \beta_1 = 0.5$, m = 0.6, n = 200. The shaded area corresponds to the time span where attraction has some effect. The right frame represents the distribution of representations after 250 times step for the 10 simulations.

and selection properly apart, we used a Gaussian function of x as the selective function: $w(x) = \exp(-(x-\mu)^2/(2\sigma^2))$. In Fig 4, it is holders of the representation x = 0.7 who are the most likely to be chosen as models. However, far from converging towards 0.7, *in fine* all representations have a value of 1, that is, the value of one of the two attractors. Why it should be so is not mysterious. The selection peak (0.7) is above m (0.6), and therefore variant 1 is favored by selection over variant 0. In group 1 however, the force of selection is dominated by that of deterministic attraction, and variants favored by selection are eliminated in favor of variants favored by attraction, i.e. variants with the value of 1. In this case therefore the final outcome is the combined effect of attraction, which eliminated all variants other than 0 and 1 (including 0.7, the variant most favored by selection), and of selection, which favored 1 over 0 (see figure 4).

2.4 – *When attraction is probabilistic: Different outcome*

The very idea of attraction is intended to capture the observation that, in cultural transmission, departures from the model are not purely random and tend to be biased in certain direction. To reintroduce stochasticity in the idea of attraction while staying as close as possible to Henrich and Boyd model, we allow

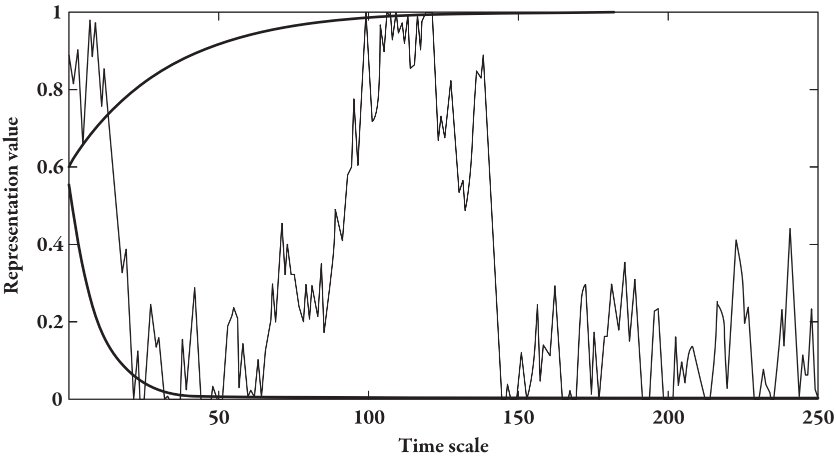

Figure 5a. Attraction with and without a degree of randomness. The two thick lines represent the lines of descent, in the absence of randomness, of two representations, one with an initial value above m (here 0.6) converging toward 1, and the other with an initial value below m converging toward 0. The thin line represents the line of descent, with a degree of randomness added to attraction, of a representation with an arbitrary initial value. All values between 0 and 1 can be reached by this line of descent. Parameters are as follows: r = (0 for thick lines and 0.2 for thin one), $\beta_0 = 0.1$, $\beta_1 = 0.03$, m = 0.6.

for the representation value acquired by an individual to vary between an interval of [x − r + Δx ; x + r + Δx].[4] To help visualize the effect of this probabilistic reinterpretation of attraction, we show in figure 5a the lines of descent of three individual representations: two obeying a non-probabilistic force of attraction à la Henrich and Boyd and beginning, one, just above the cut-off point m, and the other just below it, and a third representation with a random initial value and subject to probabilistic attraction. Without some positive degree of randomness, attraction is a deterministic mechanism that drives representation values toward 0 or 1 at a speed depending on the 'force' of attraction (in figure 5a attraction toward 0 is 3 times 'stronger' than attraction toward 1). With randomness, attraction is the probability for a representation to have a certain value given the value of the model from which it is inferred. As the figure well illustrates, with probabilistic attraction all values have a certain probability of being reached. But since, in this model, the attraction bias towards 0 is three times greater than the one towards 1, overall, values closer to 0 are more often reached.

[4] We take care of border effects by resampling new values until they fall between 0 and 1.

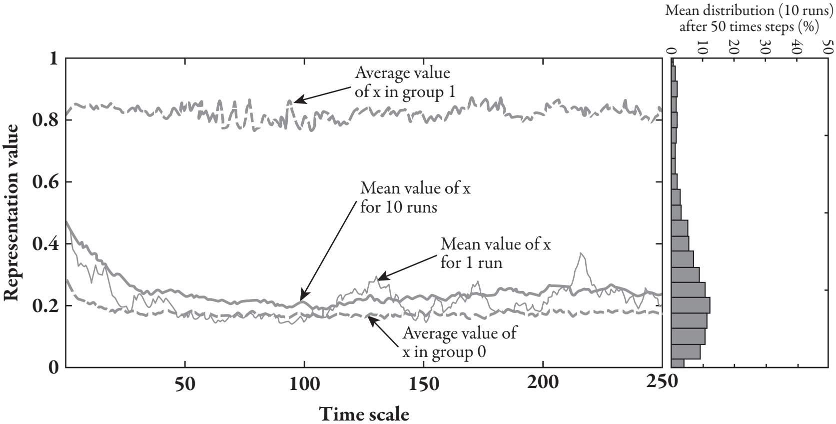

Figure 5b. Evolution of the population with no selection and probabilistic attraction three times stronger toward 0 than toward 1. The left frame represents the evolution through time of values of x as observed with the same parameters as in 5a except r=0.2. The right frame represents the distribution of representations after 250 times step for the 10 simulations.

If we represent now the whole population (n=200) with probabilistic attraction and otherwise the same parameters as in figure 5a, we observe that all values are reached but that they are more or less represented depending on the force of attraction.

What if we add to the parameters of figure 5b a weak selection force peaking at 0.7? Both selection and attraction are important factors, with selection favoring values close to 0.7 and attraction favoring values close to 0 or to 1. Because attraction remains dominant, the most often selected variants (close to 0.7) are immediately attracted toward 1 or 0 (see figure 5c). If we increase selection, we expect values around 0.7 (and therefore also around 1) to be better represented. Strong selection may indeed force the dynamics to look like replicator dynamics for mean values, but attraction remains crucial to account for the distribution we observe at equilibrium (fig 5d). Only with selection quite strong and probabilistic attraction quite weak could attraction be ignored. In general however, when you have both attraction and selection at work, both contribute to the evolution of the population. If Henrich and Boyd had shown otherwise, it would indeed have been surprising, but they have not.

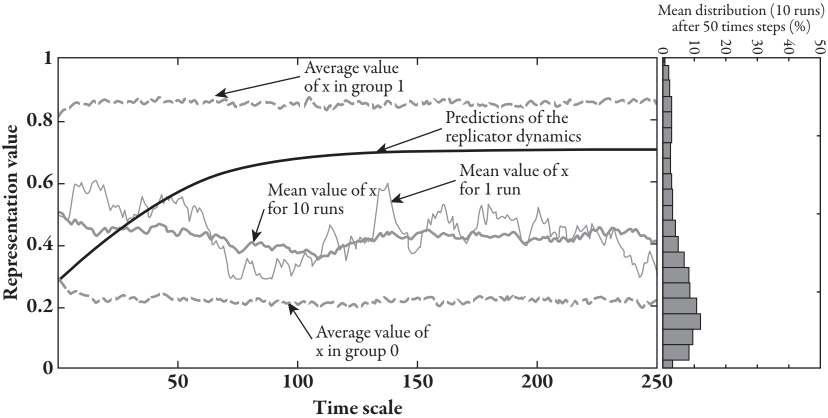

Figure 5c. Adding weak selection to attraction changes the distribution of representations in the population (see fig 5b for comparison) but it does not bring the population dynamic close to replicator dynamics. Both selection and attraction are important to explain the equilibrium distribution we observe (see the right frame). Selection favors values close to 0.7 and attraction values close to 0 or 1. Parameters are as follows: $\mu = 0.7$, $\sigma = 1.5$, $r = 0.2$, $\beta_0 = 0.1$, $\beta_1 = 0.03$, $m = 0.6$, $n = 200$.

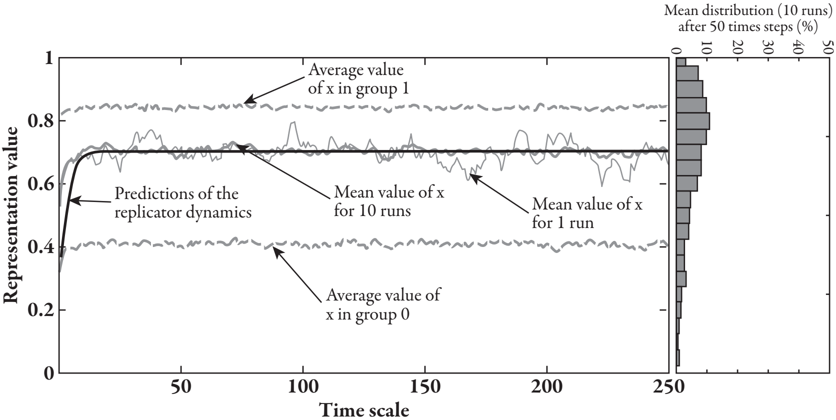

Figure 5d. Stronger selection may drive the dynamic closer to the replicator dynamic (see fig 5c and 5b for comparison) but it still does not account for the distribution we observe in the right frame. Parameters as in Fig 5c, except $\sigma = 0.4$

References

Atran, S. 1990. *Cognitive foundations of natural history: Towards an anthropology of science*. Cambridge: Cambridge University Press.

—— 2002. *In gods we trust: The evolutionary landscape of religion*. Oxford: Oxford University Press

Boyd, R., and P. J. Richerson. 1985. *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyer, P. 1994. *The naturalness of religious ideas : a cognitive theory of religion*. Berkeley, CA: University of California Press.

—— 2001. *Religion explained: the evolutionary origins of religious thought*. New York: Basic Books.

Henrich, J. and R. Boyd 2002. "On modeling cognition and culture", *Journal of Cognition and Culture*, 2 (2) 87-112.

Hirschfeld, L. A. 1996. *Race in the making: cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.

Richerson, P. J., and R. Boyd. 2005. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Sperber, D. 1996. *Explaining Culture: a Naturalistic Approach*. Oxford: Blackwell.

Sperber, D. and N. Claidière (in press). Defining and explaining culture (comments on Richerson and Boyd, Not by genes alone). *Biology and Philosophy*.

Appendix 1: The cigarette model

The 'cigarette model' informally presented in the text was meant to illustrate as simply as possible ordinary relationships between attraction and selection. Here we explain the model in more technical details.

Principles

Members of a population may each smoke between 0 and 30 cigarettes a day, so there are 31 different cigarettes patterns. Initially each smoking pattern is equally represented by 10 individuals, thus the size of the population is 310. Every year, a new age cohort of 310 youngsters joins this population and each select, from among the preceding age cohort, the individual whose smoking pattern he or she want to imitate. Imitation is imperfect and individuals typically end up, in less than a year, with a smoking pattern different from that of the individual they chose to imitate. Departure from the model are not purely random and tend to be in the direction of attractors. Thus, the first uniform distribution progressively changes with time due to both selection and attraction.

Selection

Depending on their smoking pattern, some people have a greater probability than others of being selected as models to imitate. More precisely, we suppose that the likelihood of an individual smoking x cigarette a day to be selected as a model is given by the following function:

$$W(x) = 0.15 \exp\left(\frac{-(x-10)^2}{2}\right) + 0.5$$

W(x) is greatest for x = 10 and decreases before and after that peak value (see the selection curve in figure 1a). This simply means that people smoking 10 cigarettes a day have a higher chance of being selected as models than others. In particular, if selection alone were at work and imitation were accurate, other smoking patterns, because of their lower probability of being selected, would progressively disappear, and all individuals would end up smoking 10 cigarettes per day.

Randomness

Imitation, however is not perfect. Consider first the case where the probability of a departure from the model is equal in both directions (towards smoking a greater or a lesser number of cigarettes than the model) and decreases with the distance from the model. For instance, an individual trying to imitate a person who smokes 10 cigarettes a day, has the same probability to end up smoking 8 or 12 cigarettes and a lesser probability of ending up smoking 6 or 14 cigarettes than 8 or 12. To model this case, we define a probability function r(y,x):

$$r(y,x) = \frac{\int_{y-0.5}^{y+0.5} \exp\left(\frac{-(y-x)^2}{8}\right) dy}{\int_{0}^{30} \exp\left(\frac{-(y-x)^2}{8}\right) dy}$$

Here x is the value selected and r(y,x) is the probability that an individual having selected a model smoking x cigarettes a day ends up smoking y cigarettes a day (y varying between 0 and 30). Notice, that whatever the smoking pattern of the individual imitated, the imitator may end up with any of the 31 patterns, but the probabilities are quite different for each pattern. For instance, if individual A selects as model an individual smoking 5 cigarettes a day, the probability that A will smoke 6 cigarettes by the end of the year is r(6,5) = 0.17, while the probability that A will end up smoking 10 cigarettes a day is r(10,5) = 0.09. Given that the first age cohort is uniformly distributed and the probability of going either to the left or to the right is the same, we would of course expect, in the absence of selection, to find a uniform distribution of patterns. With randomness combined with selection (and W(x) as characterized above), we find the pattern illustrated in figure 1c: most of the population is concentrated around the selection peak, as one would expect.

Probabilistic attraction

We are interested in the case where people's smoking pattern is likely to depart from the variant they selected not at random, but, we assume, in the direction of two attractors (0 and 25). We stipulate that people smoking less than 5 cigarettes are strongly attracted toward 0 and people smoking more than 5 cigarettes are progressively attracted toward 25. To represent this case, we redefine the probability function r(y,x) as follows:

$$r(y,x) = \frac{\int_{y-0.5}^{y+0.5} (0.6\exp\left(\frac{-y^2}{8}\right) + 0.75\exp\left(\frac{(-y-25)^2}{50}\right) + 0.5)/1.5\exp\left(\frac{-(y-x)^2}{8}\right) dy}{\int_{0}^{30} (0.6\exp\left(\frac{-y^2}{8}\right) + 0.75\exp\left(\frac{(-y-25)^2}{50}\right) + 0.5)/1.5\exp\left(\frac{-(y-x)^2}{8}\right) dy}$$

In this equation, the first term $0.6\exp\left(\frac{-y^2}{8}\right)$ represent the attractor 0. Thus, r(y,x) is high when y is close to 0 and decreases rapidly when y increases. The second term $0.75\exp\left(\frac{-(y-25)^2}{50}\right)$ represents the attractor 25. Thus, r(y,x) is high when y is close to 25 and decreases progressively as y depart from 25 (see figure 1a, the first term is mainly responsible of the part below 5 of the attraction function, the second of the part above 5). Finally, the third term, $\exp\left(\frac{-(y-x)^2}{8}\right)$, correspond to the previous randomness function. Now for instance, the probability that an individual selecting a person smoking 5 cigarettes a day as model should end up smoking 6 cigarettes r(6,5) = 0.15 is lower than the probability of that individual ending up smoking 4 cigarettes r(4,5) = 0.17 because attraction is lower towards 1 than towards 0 at this point. As before, r(y,x) is never 0 which means that there is always a certain probability to end up smoking any given pattern. What we expect, if attraction is acting alone (that is, without selection) is that the most frequent patterns will be 25 and 0 cigarettes and those close to them (see figure 1b).

Considering both attraction and selection

If we have both selection and probabilistic attraction (each with the parameters specified above) in play, we would expect both to affect the distribution of variants in the long run and indeed this is what we observe (see figure 1d).