The Longevity of Distinct Cultures in an Agent-Based Model of Memetic Drift

Jamie Matthews[1]

[1]Evolutionary and Adaptive Systems Group
University of Sussex, Brighton, BN1 9QH, UK
jamie.matthews@gmail.com

Abstract

When a distinct cultural region forms, its rate of absorption into the surrounding culture may be an important variable to take into account when attempting to minimise conflict. This paper describes a re-implementation of Axelrod's agent-based model of cultural dissemination, and uses it to investigate how random drift influences the longevity of distinct regions. Cultural regions are found to be surprisingly resistant to such memetic drift.

Introduction

Cultural artefacts such as beliefs, behaviours, attitudes, languages, art and music tend to spread through populations. Dawkins (1976) proposes a framework for viewing this spread as a Darwinian process. He calls the cultural replicators themselves "memes" and suggests that many aspects of human society may be explained using this paradigm.

Given that beliefs, attitudes and behaviour tend to be passed between people when they interact, how is cultural diversity maintained? Axelrod (1997) describes an abstract agent-based simulation of cultural dissemination to show that global diversity can be maintained despite local convergence.

When a distinct cultural region forms or arrives within a larger culture it may take some time before it becomes assimilated into its surroundings. An understanding of this phenomenon may be important in our desire for a peaceful society, free of tensions between cultural groups.

This paper describes a re-implementation of Axelrod's model and extends it to investigate how cultural drift (random mutation of cultural traits) affects the longevity of cultural distinctions.

Background

Memetics

In *The Selfish Gene* (1976), Dawkins introduces the concept of the *meme* to highlight the fact that there is nothing special about the gene as the fundamental unit of natural selection. This honour should be given to the more abstract *replicator*, "any unit of which copies are made, with occasional errors, and with some influence or power over their own probability of replication" (Dawkins, 2003, pp.149). Memes are another example of a replicator, and have arisen relatively recently on Earth. They are units of human culture which are passed on by imitation. Examples include ideas, melodies, beliefs, fashions, and technologies.

Like genes, memes fulfil the three criteria necessary for the Darwinian algorithm to operate. They are passed from individual to individual through imitation (heredity). Some are more successful at spreading than others (selection). Imitation may be imperfect, and new cultural artefacts may arise as novel combinations of others (variation).

Dawkins went on to suggest how this paradigm could be useful in explaining some features of human culture (such as religions: large complexes of many mutually supportive co-adapted memes). Others have taken the idea further and expanded it to other fields, including the problem of human consciousness itself, the "ultimate" meme complex (Blackmore, 1999).

Maintenance of Differences: Axelrod's Model

If this process of memetic transmission between individuals when they interact is common, how is cultural diversity maintained? Several mechanisms have been proposed to explain why cultural convergence stops before it reaches completion. Most are based on the semantics of the cultural artefacts themselves, such as "preference for extreme views" (Abelson and Bernstein, 1963) or on the specifics of the environment the population inhabits (for example, geographical isolation).

Axelrod (1997) proposes an abstract model based on the fundamental principle that "the transfer of ideas occurs most frequently between individuals . . . who are similar in certain attributes such as beliefs, education, social status, and the like."

Modelling Cultural Dissemination

The Abstract Meme

Although he never uses the terminology of memetics, Axelrod's cultural attributes are clearly analogous. His model abstracts away the *content*, or semantics, of the memes and leaves behind a list of cultural features. Each feature may take one of a range of values. The values can be thought of as metaphors for the alternative forms of the cultural artefact (if the feature is a hat, the alternative values may be a red hat, a blue hat, a green hat etc). A culture is represented as a string of digits such as "5, 2, 4, 5, 1". In this case the first feature has the fifth of its possible values, and so on.

In abstracting away the semantics of the memes, Axelrod has removed two of the three prerequisites of the Darwinian process. No trait on the cultural string is any more likely to be passed on than any other, and so there is no concept of "fitness" upon which selection may operate. Also, when a trait is passed from one culture to another it is always copied with perfect fidelity: there is no mutation and therefore no variation. His model is only one of heredity, and he asks whether cultural diversity can be maintained even in this most basic situation.

Key Assumptions

Axelrod makes two simple assumptions in his model:

- People are more likely to interact with others who are more similar to them, i.e. share more of their cultural traits.
- Interactions between people are likely to facilitate cultural transmission, increasing the number of traits shared between the two interacting parties.

The Model

Axelrod's model can be described as a randomly updated asynchronous cellular automata. The basic configuration is a square grid of cells. Each cell in the grid represents an agent, and each has its own culture string. Agents may be thought of as individual people, but due to their non-mobility, Axelrod treats them as homogeneous "villages" with a single culture string.

At each step, a site is chosen at random to be active. One of its neighbours (north, south, east or west) is also chosen at random.

> With probability equal to their cultural similarity, these two sites interact. An interaction consists of selecting at random a feature on which the active site and its neighbour differ (if there is one) and changing the active site's trait on this feature to the neighbour's trait on this feature (Axelrod, 1997).

These steps are then repeated for as many events as desired.

For example, consider an initial set of sites with randomly assigned cultures. A site is selected at random to be "active", and has the culture string "5, 2, 4, 5, 1". One of its neighbours is then selected, which has the culture string "3, 9, 4, 5, 7". By chance, these sites share two of their five cultural features (the third and fourth) and so have a 40% similarity and thus a 40% chance of interacting. If they do interact, one of the traits they do not share is copied from the active site to its neighbour. They are now 60% similar, and thus more likely to interact in the future if they are again selected at random.

Distinct Cultural Regions

Can this process of local convergence produce globally distinct cultural regions? To illustrate this, a sample run of the model is described here. The same parameters as Axelrod's initial example were used: five cultural features, each with ten possible values, on a 10×10 grid.

The similarity between two adjacent agents on the grid is shown by the opacity of the line separating them. A black line (100% opaque) indicates no shared features, while a white line (0% opaque and invisible against the white background) indicates that all features are shared. The darker the line, the lower the similarity.

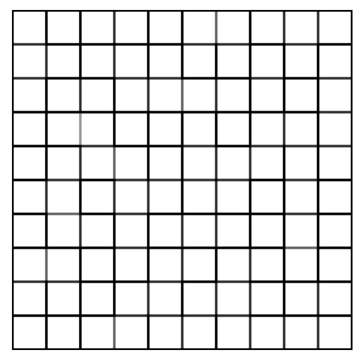

Figure 1: Initial configuration

Initially the value of each cultural feature is chosen at random for each agent (Figure 1). They are unlikely to share many features in common, and so most of the dividing lines are black.

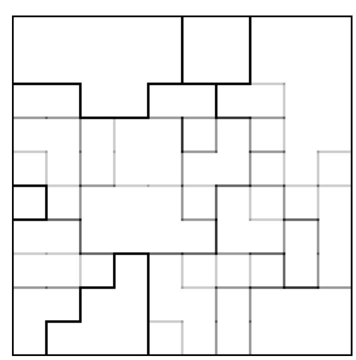

Figure 2: After 25,000 events

After 25,000 events, many cultural regions (groups of adjacent sites with identical cultures) have begun to form (Figure 2).

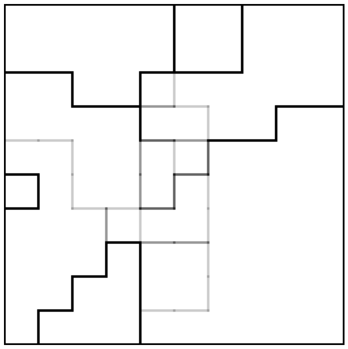

Figure 3: After 50,000 events

After 50,000 events, the cultural regions have become larger, encompassing more sites. Many of the remaining boundaries are light grey, indicating that the sites they divide differ by only one or two features. (Figure 3).

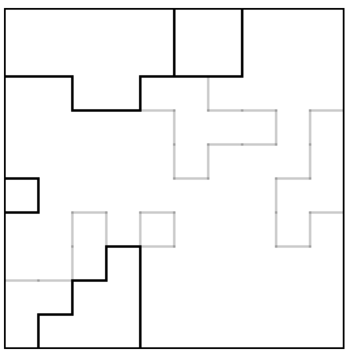

Figure 4: After 100,000 events

By 100,000 events, five clear cultural regions have emerged. The sub-regions within the largest region differ by only one feature. (Figure 4).

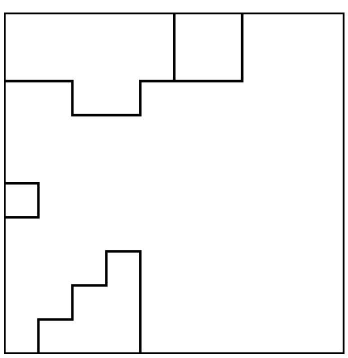

Figure 5: After 125,000 events

By 125,000 events, all sub-regions have disappeared and five main regions are clear. These are surrounded by opaque black lines, indicating that the adjacent sites at their boundaries have no features in common. The simulation is now stable as the probability of any further interactions between members of different regions is zero. (Figure 5).

The above run is a representative example of the behaviour of the model over time with different (randomly selected) initial cultural traits. It is clear that global cultural distinctions can emerge from local convergence.

The Number of Regions

To check the validity of the re-implementation, the simulation was run twenty times with different initial random culture strings. A mean of 3.35 stable regions was found, which is close to Axelrod's average of 3.2 regions for a model with the same parameters.

Axelrod's Experiments

Axelrod goes on to perform several experiments using the simulation by varying parameters (dimensions of grid, number of features in the culture string, number of possible traits of each feature, number of neighbours the active cell can interact with). He draws some interesting conclusions from these experiments, including the non-intuitive result that the average number of stable regions formed *decreases* as the size of the territory increases.

The difficulty with making such inferences from an abstract model is that it is not clear whether they are fundamental properties of the system (and thus could be applied to the more realistic, non-abstract situations upon which the model is based) or whether they are artefacts of the simplifications and assumptions built into the simulation itself.

In order to reduce this problem, it is useful to reintroduce one or more of the features that was removed for the sake of simplicity.

Cultural Drift

Axelrod suggests several possible extensions to the model, one of which he calls *cultural drift*, modelled as a spontaneous change of the value of one of the cultural features in a culture. This is analogous to a "mutation" of a meme, a random change in some aspect of culture. If the feature in question is a red hat, a mutation might involve dropping it into a bucket of green paint. [1] Intuitively, such drift may be common in real populations of interacting individuals.

Modelling Drift

It is simple to add random drift to the above model. At each step of the simulation, each feature of the current active site has probability p of undergoing a mutation. If a feature is selected for mutation, its trait is simply changed to some random value in the range of acceptable traits.

[1] The mutation may also not be random from a semantic point of view - perhaps an individual comes up with a novel new idea which can then be passed on to others. For the purposes of the model, though, such creative acts remain irrelevant. All changes are treated as random.

Analysing the Effect of Drift

As Axelrod points out, it is not obvious how to analyse the effect of drift on the basic model. Without drift, the model eventually stabilises and no further change takes place. The number of distinct regions can then be used as a measure of the heterogeneity of the grid.

When random drift is added, the model never completely stabilises, because a mutation may increase the similarity of two distinct regions (allowing future interactions) or decrease the similarity of two sites within a region (creating a slight boundary between them which may then increase due to future interactions). This raises two practical questions: how to measure heterogeneity of the grid, and when to end the simulation. Axelrod proposes several possible answers to each question, and suggests that preliminary work has shown the interaction between drift and the other parameters of the model to be quite complex.

However, this approach suffers from the same problems discussed above. While it may be possible to perform extensive experiments on the model to analyse the effect of drift, it is not clear how any results found would transfer to the real world.

For example, it may be possible to find a balance between mutation rate and the other parameters which allows the emergence of distinct regions to be preserved despite drift. However, many other factors may be present in the real world which influence this equilibrium but are ignored by the model. Gatherer (2004) uses a genetic algorithm on a similar model to Axelrod's to locate such an equilibrium, and finds that maximal memetic isolation depends on an unlikely combination of parameters. However, his model does not take into account many real-world variables which may be significant.

A more fruitful question to ask might be: given that a process of local convergence may form distinct cultural regions in the *absense* of drift, how does drift affect the stability of those regions? This approach has two main advantages:

- By analysing the effect of drift on the stability of pre-existing regions, we make no assumptions about the original source of the cultural regions themselves. Axelrod's model proposes one mechanism by which such distinctions may form, but he presents many alternative possibilities which have been suggested by others, and which may co-exist with his model. All of these can be taken into account.

- By simplifying the initial conditions of the model significantly, it is much easier to assess the stability of regions by observation, reducing the practical problems discussed above.

The Stability of Distinct Cultural Regions

To analyse how drift affects the stability of cultural regions, the initial conditions of the model were first altered to make the simulation more simple.

Instead of an initially random set of culture strings, the entire grid was made homogeneous by setting the value of every feature to zero. Then, a single mutant site was created by choosing a site at random and mutating all of its features to some random value other than zero. A sample initial configuration of the simulation is shown in Figure 6.

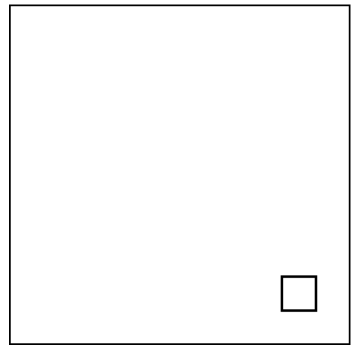

Figure 6: Initial configuration

The single mutant site is entirely distinct from the surrounding region (it shares no traits with its neighbours). In the absence of any drift, this configuration would be completely stable.

Cultural drift was then added to the model in the manner described above (see section entitled "Modelling Drift"). All other features of the model, such as the random process of convergence between neighbouring sites, remain unchanged. Also, the parameters used above (five cultural features, each with ten possible values, on a 10×10 grid) were maintained for simplicity.

The model was then run for a large number of events, and observed until the distinct mutant site disappeared (was absorbed into the surrounding region or became otherwise indistinguishable from the background activity). [2]

This process of absorption begins when a neighbouring site happens to acquire the same value in one of its features as the mutant site (either by direct mutation of that site's feature, or by the spread of that trait from elsewhere on the grid by the normal process of interaction). Now the mutant site shares a feature with one of its neighbours, it has a chance of an interaction with that neighbour which would increase the similarity yet further. In general, once a single interaction between the mutant site and its neighbour took place, the mutant site tended to disappear fairly rapidly (< 5000 events).

A sample run of this process, using a probability $p = 0.0001$ of mutation per feature at the active site per event, is shown below.

[2] To allow for unattended monitoring, this process was observed by repeatedly taking "screenshots" of the grid at 5000 event intervals. So the results are accurate to the nearest 5000 events following the disappearance of the mutant region.

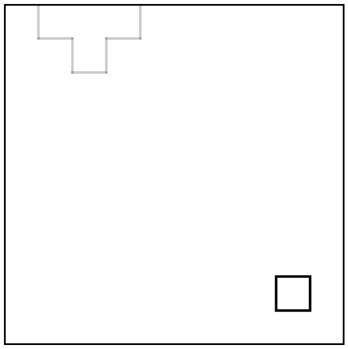

Figure 7: After 30,000 events

Single sites which are mutated become slightly differentiated from the surrounding region. Often, these are immediately reabsorbed, but occasionally they can form small clusters of differentiated sites (Figure 7).

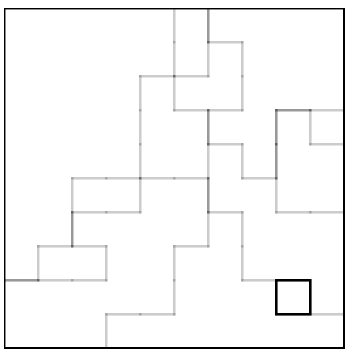

Figure 8: After 40,000 events

Usually, these small regions are short-lived, but occasionally they can "seed" larger disturbances, and chaotic patterns of differentiated regions can grow to cover much of the grid. (Figure 8).

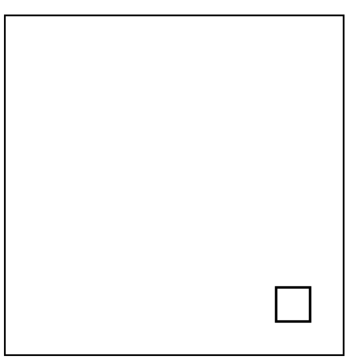

Figure 9: After 70,000 events

Often, even these large disturbances eventually resettle into stability without affecting the mutant site. (Figure 9).

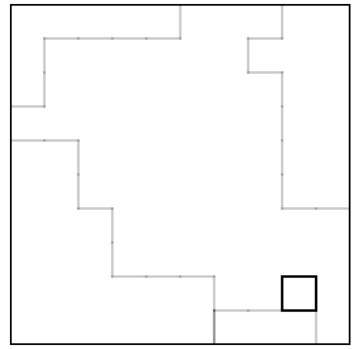

Figure 10: After 120,000 events

After 120,000 events, a new set of distinct regions has emerged and made contact with the single mutant site (Figure 10).

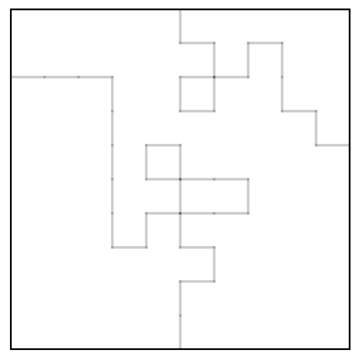

Figure 11: After 125,000 events

Just 5000 events later, the mutant site has been absorbed by the surrounding region and is no longer visible (Figure 11).

From the above run, it can be seen that with low mutation probabilities (low levels of drift), distinct cultural regions can survive significant numbers of interaction events before disappearing. Over ten such runs (with $p = 0.0001$) the mean number of events before the mutant region was absorbed was 108,000.

Rate of Cultural Drift

To further investigate these findings, the experiment was repeated with various mutation probabilities (rates of cultural drift). For each probability value, the model was run ten times, and the mean number of events to the disappearance of the mutant region was found. A graph of the results is shown in Figure 12.

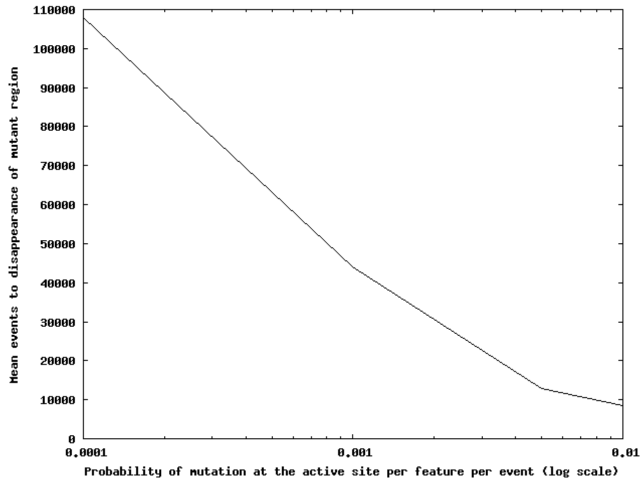

Figure 12: The effect of cultural drift on the longevity of distinct regions

What is surprising about these findings is that even as the mutation probability is increased by several orders of magnitude, the longevity of the mutant cultural region only decreases relatively slowly. One might expect that in a culture with a very high rate of drift, new cultural regions may be absorbed very rapidly as common features may appear regularly by chance, facilitating interaction across boundaries. The results of this experiment suggest that despite such high levels of drift, distinct regions may persist for significant periods of time.

Discussion

It is difficult (and probably unhelpful) to equate these findings with any concrete figures which may be found in the real world, as it is not clear what the rate of interaction (number of "events" per year, say) would be, and such values may vary widely in different regions.

In general though, it is possible to conclude that in relatively homogeneous cultures with low rates of cultural drift (as may be expected to be found in isolated, monocultural regions), any distinct cultures which do form are likely to persist for significant periods of time before being assimilated into the surrounding culture. These distinct cultures may appear through a number of possible mechanisms (including perhaps Axelrod's suggested local-interaction model), but an obvious example might be an invading or migrating group of people from a distant region with a very different culture. Finding aspects of culture in common with the invaders may be difficult, reducing the chances of further interaction and absorption.

The second result suggests that even in a culture with a high rate of drift (such as a modern, fast-changing multicultural society) it may take a considerable amount of time for a new cultural group to integrate into its surroundings. This is often intuitively true when new groups or individuals move into an established culture from afar.

Note that this finding *does not* depend on the content of the memes in either the host culture or the new, distinct culture. It is purely a stochastic interaction process between two different cultures, indifferent to the semantics of the cultural features.

It is important to bear in mind that although the reintroduction of cultural drift brought the model more closely in line with the real world, many important aspects are still missing. The main one is the third feature (in addition to heredity and variation) necessary for Darwinian evolution to take place: selection. The heart of the memetic paradigm is that individual memes or groups of memes may have a greater "reproduction" rate than others, and so may come to dominate in the population. There are several ways this could be added to the model, and this would be an interesting direction for future work. Even without this, it may prove to be useful to translate Axelrod's experiments and conclusions into the language of memes, as it may allow them to be integrated into existing memetic theory.

Summary

This paper has described a re-implementation of Axelrod's agent-based model of cultural dissemination, and discussed its parallels with memetic theory. The model was then used to investigate the longevity of distinct cultural regions in the presence of varying levels of cultural drift. It was found that cultural distinctions can be surprisingly robust, even if mutation rates are high.

Acknowledgements

Thanks to Inman Harvey for helpful advice, suggestions and encouragement.

References

Abelson, R. P. and Bernstein, A. (1963). A computer simulation model of community referendum controversies. *The Public Opinion Quarterly*, 27(1):93–122.

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

Blackmore, S. (1999). *The Meme Machine*, chapter 17. Oxford University Press, Oxford.

Dawkins, R. (1976). *The Selfish Gene*, chapter 11. Oxford University Press, Oxford.

Dawkins, R. (2003). *A Devil's Chaplain*, chapter Chinese Junk and Chinese Whispers. Phoenix, London.

Gatherer, D. (2003–2004). The inherent instability of memetic systems. *Journal of Memetics - Evolutionary Models of Information Transmission*, 8(2).