Universitat Autònoma de Barcelona

Social and cognitive constraints on the evolution of culturally transmitted variants:
Models, experiments and theory

José Segovia Martín

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: http://es.creativecommons.org/blog/licencias/

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

Social and cognitive constraints on the evolution of culturally transmitted variants:

Models, experiments and theory

PhD thesis
by

José Segovia Martín

Supervisors:
Sergio Balari
Mónica Tamariz

Universitat Autònoma
de Barcelona

PhD programme in Cognitive Science and Language
Facultat de Filosofia y Lletres
Universitat Autònoma de Barcelona
Spain

2020

Declaration

I declare that this thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own and that has been supervised by Sergio Balari and Mónica Tamariz. The work presented in Chapter 2 has received feedback from Nicolas Fay and Bradley Walker, who contributed with comments and appear as co-authors in a version of this paper which has been submitted for publication. All the work presented here has been conceived by me and I have carried out the writing and the literature review.

I certify, to the best of my knowledge, that my thesis does not infringe upon anyone's copyright and that any ideas, techniques, quotations, or any other material from the work of other people are fully acknowledged.

All the code that was used to run the simulations presented in this thesis is freely available at: https://github.com/jsegoviamartin

José Segovia Martín
24 March 2020

Abstract

The emergence of shared cultural conventions in a population is shaped by the interaction between individuals' cognition and the structure of the society. Humans, more than any other species in the animal kingdom, are able to learn and transmit vast amounts of information, through language and other cultural products. Individual cognitive constraints include cognitive biases, value systems and memory among others. Additionally, humans have an extraordinary capacity to build developmental environments and construct social niches that can be modelled as complex systems. We are born into particular societies with specific social structures, which constitute our ecological niches. Societies are shaped by the structure of the social network and other high-level hierarchical entities that constitute integrated systems of rules that structure social interactions (e.g. institutions). In this thesis I formalise some of the relationships between these factors using a variety of approaches. In particular, I explore the following three main research questions:

(1) How do the interactions between individual cognitive traits and the temporal dynamics of social network connectivity, i.e. the order in which individuals in a population interact with each other, affect the spread of cultural variants?

(2) How do the interactions between individual cognitive traits and institutions affect the evolution of cultural diversity and the emergence of cultural conventions?

(3) How might current iterated learning models, niche construction and evolutionary developmental biology be synthesised into a compatible framework for language evolution?

Chapter 1 contains a review of the literature and an introduction to the assumptions underlying the models presented in this thesis. In Chapter 2, I present an agent-based model manipulating specific network connectivity dynamics, cognitive biases and memory. I show that connectivity dynamics affect the time-course of variant spread, with lower connectivity slowing down convergence of the population onto a single cultural variant. I also show that, compared to a neutral evolutionary model, content bias (i.e. a preference for variants with high value) is the main driver of convergence and amplifies the effects of connectivity dynamics, whilst larger memory size and coordination bias, especially egocentric bias, slow down convergence. In Chapter 3, I report on an experiment in the lab in which participants engage in a Pictionary-like graphical communication task as members of a 4-participant micro-society. The experiment has two main goals: First, to evaluate the effect of two network connectivity dynamics (early and late) on the evolution of the convergence of micro-societies on shared communicative conventions under controlled conditions. Second, to compare the predictions of the agent-based model described in Chapter 2 against experimental data, and calibrate the model to find the best-fitting parameter setting. Our experimental data shows that, as predicted by the model, an early connectivity dynamic increases convergence and a late connectivity dynamic slows down convergence. Expanding on the model developed in Chapter 2, Chapter 4 explores how the interactions between content bias, value systems and institutional performance affect cultural diversity and the emergence of cultural conventions at the population level. Simulation results show that high hegemony (i.e. one or few variants are strongly preferred over the others) and homogeneity of value systems among agents accelerate the extinction of cultural traits and thus erode cultural diversity. In certain regions of the parameter space, institutions that do not reinforce original value systems tend to be effective at preserving cultural diversity. However, an important limitation of this model is that institutional performance remains fixed over time. In Chapter 5, I attempt to overcome previous shortcomings of the model by developing a co-evolutionary model of value systems, institutions and choice. To the best of my knowledge, this model constitutes the first attempt to quantify the propagation of cultural variants by incorporating a comprehensive parameter combination of compliance, confirmation, content and frequency biases into the learning and production algorithm. Results show that, in general, institutional power facilitates the emergence of cultural conventions when compliance biases increase. In general, a compliance bias pushes diversity up when institutions are diverse, and pushes diversity down when institutions convey value systems with strong dominance of one or few cultural variants. In some regions of the parameter space, global conventions can also emerge in the absence of institutional power and therefore of institutions that are in place to guide convergence. On a more conceptual level and applied to the case of language, Chapter 6 reviews the literature on iterated learning and ecological evolutionary developmental biology to explore their compatibility. I use the concept of niche construction to build bridges between eco-evo-devo accounts for cognitive capacities and cultural evolution guided by iterated learning processes. Additionally, based on recent insights from both domains, I propose an integrated conceptual model that might be useful to connect biological and cultural approaches, as well as act as a hypothesis-generating framework around which cognitive scientists can structure new triple-inheritance formal models. In Chapter 7, I summarise the most relevant findings of this thesis and I discuss some potential implications.

Acknowledgements

I want to thank my supervisors Sergio and Mónica for sharing their knowledge with me and for being so generous and supportive throughout this PhD. From them I have socially inherited not only a large amount of technical skills and academic strategies, but useful tools to do better science. However, what I am most grateful to them for is their genuine way of being, their spontaneity and closeness. Without the former my PhD would have been worse, without the latter I would not have completed it.

I feel lucky to have been part of the doctoral program in Cognitive Science and Language (CCiL-UAB) and the Centre de Lingüística Teòrica (CLT). I want to thank director Anna for always maintaining fast and efficient communication with the students and for her great ability to lead multidisciplinary research projects. I also feel very fortunate to have spent a large part of my PhD research at the Centre for Language Evolution (University of Edinburgh) and the Centre for Biological Diversity (University of St.Andrews). Thank you Kenny, Simon and Kevin for your extremely useful feedback.

Thank you all the students and postdocs I came across and shared stimulating conversations with: Alejandra, Andres, Carmen, Dària, Edith, Fausto, Fiona, Helen, Jasmeen, Jingtao, Jon, Jonas, Marieke, Svenja, Thomas, Tom. I want also to thank fellow researchers with whom I have crossed paths (physically or virtually) during my PhD: Nicolas Fay, Bradley Walker, Sean Roberts and Paul Smaldino. Their projects, work and feedback was very insightful.

Thank you Ash for your contribution with ideas, corrections and revisions. Without them, this research would be much worse. But above all, thank you for all the shared things that cannot be explained with science and words. Thank you Spooki (my cat) too for being so cuddly, playful and for always listening and sharing your advice in the good and the bad moments.

Gracias mamá, papá, Edu, abuelas Elvira y Eugenia, Petra y demás frimos y gansos por el apoyo que siempre me dais para que siga adelante

con mis locas aventuras. En especial, gracias Edu por la implicación y ayuda técnica para montar un laboratorio online.

Last but not least, I want to thank Lucy, my beloved van. Many of the best ideas that appear in this thesis emerged within her magical yellow walls.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cultural diversity and cultural conventions are the foundation for human life. Whether linguistic, social, political, ethical or religious, human cultural conventions have the unique power to capture and share thoughts, meanings and traditions, which in turn has played a major role in the development of civilisations and the spread of human culture all across the globe and even beyond. Although we are not the only species with the ability to copy and transmit culture, everything indicates that there is no case in the animal kingdom comparable to the vast human capacity to conventionalise, that is, to create languages, technology, norms or political systems.

Human life is thus shaped by our culture, that is, by socially transmitted information that determines our behaviour, attitudes and values (Boyd & Richerson, 1985). Language, technology or beliefs propagate in populations following evolutionary dynamics (Cavalli-Sforza & Feldman, 1981). Since conventions imply selection of one variant from among two or more alternatives, by definition, the more conventional a cultural variant, the less diversity for that variant. For example, the processes of language shift are a paradigmatic example of cultural loss: Although language birth and death are natural processes, it is estimated that 90% of the about 6000 languages currently spoken may become extinct by the end of this century (Krauss, 1992). However, conventions can also facilitate the formation of different in-groups, which in turn would rise global cultural diversity. The idea that social influence combined with certain social structures might be a prominent mechanism for the preservation of cultural diversity has been studied in the past empirically. Indeed, if there is enough heterogeneity in the society, the formation of conventions in similar groups can lead to the emergence of dissimilar cultural groups, and in turn, the maintenance of cultural diversity (Axelrod, 1997; Centola, Gonzalez-Avella, Eguiluz, & San Miguel, 2007).

Yet the formation of human conventions is not a simplistic process where each level of analysis is disentangled from each other. Instead, they are subject to complex networks of interactions between individual-level and society-level features (Labov, 2011). Individuals' learning and production cognitive mechanisms (Acerbi, Van Leeuwen, Haun, & Tennie, 2016; Boyd & Richerson, 1985; Fay, Garrod, Roberts, & Swoboda, 2010; Ferdinand, Thompson, Kirby, & Smith, 2013; Kirby, 2017; Tamariz, Ellison, Barr, & Fay, 2014) affect humans' cultural choices. From such choices, complex niches emerge, giving rise to social structures that are shaped by population network structures (Becker, Brackbill, & Centola, 2017; Centola & Baronchelli, 2015) and other high-level hierarchical entities that constitute ecological niches, such us human institutions (Bowles, 2000; Bowles, Choi, & Hopfensitz, 2003; J. Kendal, Tehrani, & Odling-Smee, 2011). Both, individuals and cultural structures co-evolve in dynamic developmental process where the boundaries of each process can not be easily unraveled from the others (Balari & Lorenzo, 2013; Laland, Odling-Smee, & Feldman, 2000; R. C. Lewontin, 1983). Individual underlying cognitive mechanisms bring about patterns of collective action at the population level, promoting the construction of complex niches. These human niches in turn yield additional pressures that alter the way in which cultural evolution proceeds (Laland, 2017a).

The complex dynamics that shape the formation of human conventions are therefore related with the preservation of cultural diversity, and due to their crucial role in the evolution of human behavior and culture, these patterns of collective behavior are one of the most interesting and useful human processes to be understood. Not only because of their intrinsic research value, but because a better knowledge of these patterns is crucial to develop more integrated and coherent theoretical models of cultural evolution, and in turn help to better structure the bases upon which public policies for cultural preservation stand.

The purpose of this thesis is to explore some specific relationships in the formation and spread of cultural conventions. Although I will use a broad definition of culture where cultural variants can be thought of as a proxy for words, opinions, beliefs, norms or even artistic tastes (music, painting), an important part of the models I will construct throughout this thesis derive from a tradition of mathematical and theoretical frameworks grounded in the fields of linguistics and cultural evolution. For this reason, the reader of this thesis will find that certain discussions are more focused on the emergence and formation of communicative conventions and language. However, all models I will present here are built with an eminently practical and general intention.

1.1 Complex dynamical systems: society and individuals

The idea that these patterns of collective behavior are shaped by varying selection pressures at both the individual and the population level has been largely put forward by several theorists in the social, behavioral and cognitive sciences (Boyd & Richerson, 1985; Lass, 1997; Lewis, 2008; Young, 1993). Social evolutionary theories maintain that cultural evolution is constrained by individual traits and social structures (Labov, 2011;

R. C. Lewontin, 1983). However, although there is a widespread interest in approaching the study of the evolution of culture form an integrated perspective, the complexity of the social process has prevented the development of models of dissemination of culture that include both a comprehensive set of cognitive traits and specific features of the social structures at the population level.

Here I ground my exploration in two main sources of variation: Society and individuals. Firstly, the transmission of cultural features in a population (e.g. language variants) depends on a number of characteristics of the society. One of these features is the arrangement of connections among individuals in the social network (Centola, 2018). For example, who speaks to whom, how often people exchange information or how many neighbourhoods there are in a population are all important characteristics of a social network. Social networks shape the spread of cultural information by connecting or disconnecting people over time and in turn they can affect the dissemination of cultural variants and the emergence of cultural conventions. Previous models have successfully shown how the co-evolutionary dynamics of social influence and network structures can affect cultural change (Becker et al., 2017; Centola, 2010; Centola & Baronchelli, 2015; Centola et al., 2007; Muthukrishna & Schaller, 2019). Inspired by those models and expanding on previous work in cultural evolution (Fay et al., 2010; Tamariz et al., 2014), here I focus on a specific manipulation of the social network: The order in which connections between individuals unfold over time. I present a micro-scale model of dissemination of culture and I apply my simulations to a number of specific social learning scenarios.

Secondly, societies are formed by individuals. And individuals may be disproportionately in favour or against an idea, a norm or a word (Boyd & Richerson, 1985). For example, individuals can be biased towards adopting one particular word just because it is more iconic or easy to remember, or against one particular institutional norm just because it was promoted by an unpopular government. People also have personal beliefs or value systems, and they can be more prone or reluctant to use new information, such as scientific information, institutional information or religious information to update their prior beliefs or value systems. In this thesis, I am particularly interested in the construction of models that include a comprehensive set of cognitive biases that have been shown to be crucial mechanisms underlying the adoption of cultural variants at the individual level, and therefore critical for explaining cultural diversity at the population level. In particular, throughout this thesis I explore the effects of a *general content bias*, or a preference for variants with high value (Boyd & Richerson, 2008; J. Henrich & McElreath, 2007; Tamariz et al., 2014), an *egocentric bias*, or a preference for self-produced variants (Tamariz et al., 2014), an *allocentric bias*, or a preference for others-produced variants (Garrod & Pickering, 2007; Pickering & Garrod, 2004; Tamariz & Kirby, 2015), a *compliance bias*, or a tendency to conform to institutions (Cialdini & Goldstein, 2004), a *confirmation bias*, or a tendency to process new information in a way that confirms one's prior beliefs (Del Vicario, Scala, Caldarelli, Stanley, & Quattrociocchi, 2017; Nickerson, 1998; Quattrociocchi, Scala, & Sunstein, 2016) and a *memory-frequency bias*, or a preference for variants that are more frequent in the agents' history (Ferdinand et al., 2013; Hudson Kam & Chang, 2009; Hudson Kam & Newport, 2005; Tamariz & Kirby, 2015). My models, thus, require also the implementation of memory, which is operationalised as the maximum amount of history that can affect agents' variant choice at a given moment. All these cognitive traits are, therefore, important selection pressures that can affect the way in which cultural variants spread out in a population. Throughout my exploration, I examine the impact of these parameters using different versions of my agent-based models.

Human societies, however, are not just a number of individuals or groups learning and sharing information within a social network. Humans have an extraordinary capacity to build developmental environments and complex niches. Niche construction is particularly relevant to human evolution because by modifying the environment, humans create artifacts that act as additional sources of biological and cultural selection (Laland, 2017a; Laland et al., 2000). Human institutions are a paradigmatic example of these human artifacts (Bowles, 2000). Institutions have been defined by researchers on institutionalism as 'integrated systems of rules that structure social interactions' (Hodgson, 2015, p. 501). Similarly, Boyd and Richerson (2008) define social institutions as norms and conventions that give durable structure to social interactions within a population. These institutions are now recognised to have played important roles in the co-evolution of individual cognition and human culture by creating norms, spreading beliefs, establishing language rules or promoting ideas. In the later versions of the computational models I present in this thesis, I pay special attention to the emergence of representative institutions and how they interact with human cognition, value systems and the selection and production of cultural variants in a population. On a more conceptual level, the notion of niche construction will be also used in the last part of this thesis to propose a conceptual model applied to the case of the emergence of language regularities. I gather evidence from iterated learning models (Chater, Reali, & Christiansen, 2009; Kirby, 2017; Kirby, Dowman, & Griffiths, 2007; Smith et al., 2017; Zuidema, 2002) and ecological evolutionary developmental biology (Gilbert & Epel, 2009, 2015; R. C. Lewontin, 1983; Müller, 2020; Sultan, 2015, 2017) to show that these two frameworks are compatible on the basis of recent interdisciplinary studies that are stressing the necessity to construct more integrated and less simplistic models of cultural evolution (e.g. Balari & Lorenzo, 2013; Charbonneau, 2016; Deacon, 2010; Fisher & Vernes, 2015; Laland, Odling-Smee, Hoppitt, & Uller, 2013; Laland, Odling-Smee, & Turner, 2014; Mesoudi, Whiten, & Laland, 2006, among others).

The processes of cultural transmission are therefore shaped by social structures (e.g. network connectivity dynamics and emergent institutions) and by individual cognitive traits (e.g. cognitive biases, value systems and memory). All these factors take part in complex dynamic systems that require an integrated treatment of the different levels of the evolutionary process.

1.2 Plan of the thesis

This thesis aims to formalise some of the relationships between the social and cognitive features outlined above. In particular, I will explore three main research questions:

(1) How do the interactions between individual cognitive traits and the temporal dynamics of social network connectivity, i.e. the order in which individuals in a population interact with each other, affect the spread of cultural variants?

(2) How do the interactions between individual cognitive traits and institutions affect the evolution of cultural diversity and the emergence of cultural conventions?

(3) How might current iterated learning models, niche construction and evolutionary developmental biology be synthesised into a compatible framework for language evolution?

Broadly speaking, question 1 will be addressed in Chapters 2 and 3, question 2 in Chapters 4 and 5, and question 3 in Chapter 6. In my research

I will adopt a variety of approaches, such as modelling, experiments and theory. Although there are obvious connections between chapters, each chapter of this thesis can be read independently, because each of them corresponds to an independent study that has been written as an independent academic article. Both Chapter 2 (a model) and Chapter 3 (an experiment) are devoted to the same topic: Network connectivity dynamics, memory and cognitive biases. Chapters 4 and 5 are two models of dissemination of culture that aim to explore the interactions between institutions and value systems. Finally, Chapter 6 proposes a conceptual model of language evolution that attempts to integrate iterated learning, evo-devo and niche construction.

In Chapter 2, I construct an agent-based model to explore how the interactions between population network connectivity dynamics (i.e. the order in which agents interact over time), content bias, coordination bias and memory affect the evolution of social conventions and cultural diversity. This study starts with an explanation of drift and selection pressures in cultural and language evolution. It then reviews the relevant literature on social networks and proposes a specific manipulation of the connectivity dynamic according to how many agents could potentially share the same cultural variant at a given round. Then, I describe an agent-based model that includes the parameters mentioned above. Simulations show that content bias is the main driver of cultural diversity and that it amplifies the effect of network connectivity dynamics on the time-course of variant spread. Chapter 2 then continues with a discussion and stresses the necessity of testing the model against experimental data obtained from multiple connectivity dynamics.

Chapter 3 reports an experiment in the lab in which participants engage in a Pictionary-like game as members of 4-agent micro-societies. This study has two goals. First, to assess the effect of two specific network connectivity dynamics on the evolution of communicative conventions; and second, to compare the predictions of the model developed in Chapter 2 against experimental data. Experimental results support the hypothesis that the convergence of a small-scale society of human participants on shared communicative conventions can be affected by the order in which connections between individuals unfold over time.

In Chapter 4 I develop an agent-based model to explore the effects of the interactions between content biases, value systems and institutional performance on cultural diversity. Value systems are operationalised as the set of values that each agent assigns to each cultural variant at each time step, and they can evolve according to the agents' choices and the type of institutional pressure. However, one limitation of the model developed in Chapter 4 is that institutions are fixed and hence they are not part of a co-evolutionary model of cumulative culture. In Chapter 5, I construct an agent-based model that attempts to overcome those limitations. The co-evolutionary model includes value systems, institutions and a comprehensive set of cognitive biases. Simulations show that institutions modulate the emergence of cultural conventions and the evolution of cultural diversity. They are, however, dependent on the particular combination of value systems and cognitive biases in the population.

Chapter 6 proposes a conceptual integrated framework for language evolution that includes insights from iterated learning, evo-devo and niche construction. I start by challenging the idea of domain specificity with a review of some relevant iterated learning models. These models show that the emergence of near universal properties of language do not require the preexistence of strong specific constraints. Then I provide a brief review of the history that led to the development of eco-evo-devo models. Finally, I revisit a variety of studies that might be adding evidence to support the main hypothesis of the study: The conceptual apparatus of eco-evo-devo is compatible with the findings of iterated learning models. Chapter 7 sums up the most relevant findings of this thesis and discusses several implications for the cognitive and social sciences, such us the link to studies of social influence, dissemination of culture and language history among others. This discussion should prompt some serious thinking by those researchers working on collective behavior, social networks and institutions, as it might help to develop more coherent theoretical models of cultural evolution and better structure the bases upon which public policies for cultural preservation and information management stand.

Due to the particular nature of this thesis, which can be defined as a compilation of related but also independent studies, some of which are versions of submitted articles to academic journals, the reader of this thesis might find that language and terminology in some chapters are used for such purpose. That is, in order to acknowledge the contribution of my supervisors and collaborators I generally use plural forms. In the end, this thesis consists of a number of studies that I have developed throughout my PhD in Cognitive Science and Language, and therefore maintaining the original attributes of each study helps present each piece of work, albeit in chapter form, in a more fluent and less forceful way.

In particular, Chapters 2 and 4 are revised versions of two papers that were developed as independent studies, on which the work reported on Chapters 3 and 5 builds upon. The work presented in Chapter 6 also materialised as an independent study and as a corollary of the material that precedes it in this thesis. All these works have been submitted for publication and are currently either under revision or accepted. For the sake of completeness, the full credits for these articles are, respectively:

- José Segovia-Martín, Bradley Walker, Nicolas Fay & Mónica Tamariz, (in press). Network connectivity dynamics, cognitive biases and the evolution of cultural diversity in round-robin interactive micro-societies. *Cognitive Science*. An early version of this paper is available online at arXiv.

  Affiliations: Universitat Autònoma de Barcelona, University of Western Australia & Heriot-Watt University.

  JSM conceived, designed the study, coded, performed and analysed simulations, and wrote the manuscript. MT contributed in designing the study and writing the manuscript. BW and NF contributed useful comments.

- José Segovia-Martín & Mónica Tamariz, (2020). Value systems, content bias and institutional performance: a micro-scale model of cultural diversity. Manuscript submitted for publication.

  Affiliations: Universitat Autònoma de Barcelona & Heriot-Watt University.

  JSM conceived, designed the study, coded, performed and analysed simulations, and wrote the manuscript. MT contributed in designing the study and writing the manuscript.

- José Segovia-Martín & Sergio Balari, (2020). Eco-evo-devo and iterated learning: Towards an integrated approach in the light of niche construction. Manuscript submitted for publication

  Affiliations: Universitat Autònoma de Barcelona.

  JSM conceived, designed the study, developed the mathematical and conceptual explanations and wrote the manuscript. SB contributed in designing the study and writing the manuscript.

Chapter 2

Network connectivity dynamics and cognitive biases

Abstract

The distribution of cultural variants in a population is shaped by both neutral evolutionary dynamics and by selection pressures. The temporal dynamics of social network connectivity, i.e. the order in which individuals in a population interact with each other, has been largely unexplored. In this chapter we investigate how, in a fully connected social network, connectivity dynamics, alone and in interaction with different cognitive biases, affect the evolution of cultural variants. Using agent-based computer simulations, we manipulate population connectivity dynamics (early, mid and late full-population connectivity); content bias, or a preference for high-quality variants; coordination bias, or whether agents tend to use self-produced variants (egocentric bias), or to switch to variants observed in others (allocentric bias); and memory size, or the number of items that agents can store in their memory. We show that connectivity dynamics affect the time-course of variant spread, with lower connectivity slowing down convergence of the population onto a single cultural variant. We also show that, compared to a neutral evolutionary model, content bias accelerates convergence and amplifies the effects of connectivity dynamics, whilst larger memory size and coordination bias, especially egocentric bias, slow down convergence. Furthermore, connectivity dynamics affect the frequency of high quality variants (adaptiveness), with late connectivity populations showing bursts of rapid change in adaptiveness followed by periods of relatively slower change, and early connectivity populations following a single-peak evolutionary dynamic. We evaluate our simulations against existing data collected from previous experiments and show how our model reproduces the empirical patterns of convergence.

Keywords: cultural evolution, convergence, adaptiveness, connectivity, network topology, content bias, coordination bias, memory, punctuational evolution

2.1 Introduction

Human life is shaped by our culture, that is, by socially transmitted information that determines our behaviour, beliefs, attitudes and values (Richerson & Boyd, 2008). Cultural variants such as technology, language and beliefs propagate in populations following evolutionary dynamics (Boyd & Richerson, 1988; Cavalli-Sforza & Feldman, 1981; Neiman, 1995)—individuals inherit cultural traits from ancestors or peers and occasionally generate new trait variants. Over generations, cultures evolve: some variants are lost, while others spread in a population, sometimes to the point of fixation, when we can say the population has converged on a variant.

In the absence of innovation, drift may cause cultural variants to disappear and, in turn, reduce cultural variation. Drift models can thus lead, over generations, to convergence on a single variant. This is particularly true for smaller populations because random sampling can remove variants faster, resulting in a change of variant distributions over generations. A variant has been fixed when its frequency is 1, and has been extinguished when its frequency is 0. Therefore, since the influence of stochastic variation is higher for small populations, the probability of fixation is also higher in smaller populations (Frankham, Briscoe, & Ballou, 2002). Interestingly, it is well known that drift may explain the propagation of cultural variants including baby names, pottery decorations and patents (Bentley, Hahn, & Shennan, 2004), dog breeds (Herzog, Bentley, & Hahn, 2004) and some diachronic changes in language (DeGraff, 2001; Komarova & Nowak, 2003; Kroch, 1989). These applications and properties make drift models usable as null models against which other models can be tested (Hahn & Bentley, 2003; Lipo, Madsen, Dunnell, & Hunt, 1997; Neiman, 1995; Reali & Griffiths, 2009; Shennan & Wilkinson, 2001).

2.1.1 Cognitive biases and memory

Some variants spread more rapidly than others. In these cases, evolution is not neutral, but subject to biases, or selection pressures. *Content bias*, also termed *direct bias* by Boyd and Richerson (1985), relates to individuals' sensitivity to intrinsic properties of traits, and results in the more learnable, efficient or effective variants having a higher probability of being adopted by others (Hagen & Hammerstein, 2006; J. Henrich & McElreath, 2007; Vale et al., 2017), and therefore spreading faster through a population than a neutral, drift model would predict (Gong, Shuai, Tamariz, & Jäger, 2012; Tamariz et al., 2014). *Coordination biases* may involve a preference to use variants we have used before (egocentric bias) or variants produced by others (allocentric bias). In communicative tasks, for instance, an allocentric bias can be observed under certain mechanistic accounts of dialogue, since speakers tend to adopt labels used by their interlocutors (Garrod & Pickering, 2007; Pickering & Garrod, 2004), which in turn favours cooperation and coordination (Fusaroli et al., 2012; Fusaroli & Tylén, 2016). Content and coordination biases also interact with each other: Egocentric bias maintains variation, which improves the chances that content bias will select for the most adaptive variant in a population (Segovia-Martín, Walker, Fay, & Tamariz, 2019).

Cultural transmission is also affected by the memory record of cultural variants. Some authors claim that the type of variation that learners produce can be explained by memory limitations: for example, memory can affect language regularisation (Hudson Kam & Chang, 2009; Hudson Kam & Newport, 2005), compressibility (Chater & Vitányi, 2003) or conventionalisation (Tamariz & Kirby, 2015). In general, memory limitations reduce variation (Tamariz & Kirby, 2015). Because frequency learning is a prominent aspect of social learning in linguistic and non-linguistic tasks, regularisation behaviour is consistent with a domain-general account of the observed regularisation bias and it might be attributable to limited working memory (Ferdinand et al., 2013). Ferdinand et al. (2013) also suggest that a tendency to produce representative variants can lead to high-fidelity reproduction of the training set of variants under low memory constraints. This idea is very relevant to the scenarios that we aim to model, in which there is production and reproduction of a discrete number of variants with high-fidelity. In addition, it allows us to model memory as a cognitive feature that constraints regularisation on the basis of frequency learning.

The effects of content and coordination biases on variant propagation were studied by Tamariz et al. (2014), who constructed a parameterised model of cultural variant transmission to analyse the patterns of variant spread obtained in an experimental study by Fay et al. (2010). Fay et al. (2010) had groups of eight individuals playing a Pictionary-like communication game in pairs. During each game, the *director* produced a drawing to represent each of 16 meanings, one at a time. For each of these, the *matcher* tried to guess which of 20 possible meanings the director was trying to communicate. Participants played this game six times with each partner, with roles reversing for each game, so each participant drew and matched each meaning three times with each partner. For a given meaning, directors could invent their own ways to depict the meaning (that is, produce a novel variant) or produce a variant that was produced by a partner or by themselves in a previous game. After six games, participants swapped partners within their group and played another six games with their new partner. This partner-swapping was repeated until the populations were fully connected, with every participant having played with every other participant. This meant that, by the end of the experimental simulation, for each meaning, a particular variant could spread to the whole population. In other words, the population could converge on a single variant for each meaning. Tamariz et al. (2014) found that this convergence was best explained by a combination of egocentric bias and content bias, where participants would stick with variants they had produced previously, unless they encountered a better variant, in which case they would switch to that.

2.1.2 Population structure, network dynamics and specific social learning scenarios

Demographic factors also add selection pressures (e.g. J. Henrich, 2004; Mesoudi, 2011b; Mesoudi et al., 2006; Richerson & Boyd, 2008; Shennan & Wilkinson, 2001; Vaesen, 2012): The degree of adaptiveness, complexity and cumulative cultural evolution of cultural and communicative variants can be affected by population size (Cuskley, Loreto, & Kirby, 2018; Derex, Beugin, Godelle, & Raymond, 2013; J. Henrich, 2004; Kempe & Mesoudi, 2014; Kline & Boyd, 2010; Kobayashi & Aoki, 2012; Shennan & Wilkinson, 2001), by the degree of contact and migration between populations (Creanza, Kolodny, & Feldman, 2017; Muthukrishna, Shulman, Vasilescu, & Henrich, 2014; Powell, Shennan, & Thomas, 2009) and by the structure of the social network (Gong, Minett, & Wang, 2008; Lee, Stabler, & Taylor, 2005; Lupyan & Dale, 2010; Mueller-Frank, 2013; Olfati-Saber & Murray, 2004). An additional demographic variable, namely the connectivity between individuals within or across populations, also enhances adaptiveness and complexity because it affects the degree of diversity each individual has access to (J. Henrich, 2004; Kobayashi & Aoki, 2012; Powell et al., 2009; Shennan & Wilkinson, 2001). However, a recent study suggests that access to diversity is not the only variable at play, and that if we take into consideration the potential for an innovation to be adopted and spread, then an intermediate degree of connectivity may be optimal for cumulative culture, as too much connectivity stifles innovation, whilst too little cannot maintain complex traits (Derex & Boyd, 2016; Derex, Perreault, & Boyd, 2018).

In the evolution and history of human populations, the structure of the population might have played an important role in cultural change in ancestral and historical periods (Derex & Boyd, 2016). Critically, it should be noted that population fragmentation and cultural isolation have been identified as crucial factors to explain the spread of cultural variants such as high quality ideas (Björk & Magnusson, 2009), technology (Hovers & Belfer-Cohen, 2006) or research (March, 2005). Furthermore, it has been suggested that inter-population connectivity may be more than just a simple reflection of cultural accumulation, and that it may be a critical driver of cultural change (Creanza et al., 2017).

Important efforts to model the effects of social structures and network dynamics on the spread of cultural variants have been made in recent years. Particularly interesting for the present study is the work that follows Axelrod's model of dissemination of culture (Axelrod, 1997). This model was based on the assumption that people are more likely to interact with others who share the same cultural variants, and this in turn tends to increase the number of variants they share. These mechanisms, named homophily and influence, have been shown to be prominent explanations for the persistence of cultural diversity. Expanding on Axelrod's (1997) model, researchers have explored social learning and network connectivity effects by implementing new manipulations, which in general involve complex interactions of agents' cognition and social structures. Particularly relevant for us are agent-based models that contemplate an interaction between imitation choice rules and dynamic networks, such as the modelling of the co-evolution of networks and agent's preferences (Centola et al., 2007), network efficiency and conformity biases (Barkoczi & Galesic, 2016) and network connectivity and group performance (Lazer & Friedman, 2007). All these models manipulate specific features of the network topology and agent's cognition in order to predict the spread of cultural variants and the convergence on shared cultural variants. It is important to recognise that each of these agent-based models is based on a number of tractability assumptions that, depending on the focus of the study, lead to paying more attention to either macro processes or micro processes. The ability of many of these models to make global predictions takes focus away from micro-processes such as the composition of the pairings at the most elementary level, limiting the tractability of each of the agents that take part in the cultural process.

In the real world, the order in which individuals pair over time is sometimes determined by stochastic events that govern path dependencies, that is to say, a limited record of experiences that one has experienced in the past. When individuals in a community are carriers and transmitters of information, different orderings can yield different levels of sub-population isolation at different times, which in turn affects the probability of dissemination of information in social groups. For example, the specific pair composition within a micro-society determines the number of agents that can potentially share the same variant at different times. To be more specific, the present study can be applied to a number of socio-cultural scenarios that are governed by turn-based interventions (Sacks, Schegloff, & Jefferson, 1978), in which the organisation of turn taking might play an important role in, for example, the formation of sides during jury deliberations (Manzo, 1996; Stone, 1969), computer mediated communication (Garcia & Baker Jacobs, 1999), the evolution of communicative conventions in the lab (Bloom, Russell, & Wassenberg, 1987; Fay, Garrod, & Roberts, 2008; Fay et al., 2010), or communication in group decision-making (Bormann, 1996; Stasser & Taylor, 1991). This might include certain group dynamics in the work place, job selection processes or strategies in team games. What these scenarios have in common is that they usually

(a) occur in relatively short period of time;

(b) take a relatively small population;

(c) consist of a relatively small number of interactions between actors;

(d) are systematically structured by a prefixed scheduled algorithm; and

(e) have a tendency to require convergence as an outcome (e.g. jury verdict or final outcome of a decision-making process).

These properties make these socio-cultural scenarios sensitive to subtle changes in the network structure, which can potentially enhance the impact of the formation of short-term clusters (i.e. groups of agents with strong ties or who interact more frequently with one another) and individual biases on the outcome of the process. The formation of short-term clusters can lead to differential information concentration that can dramatically affect the spread of highly valued variants within the group (Axelrod, 1997; Centola, 2018; González-Avella, Cosenza, Klemm, Eguíluz, & San Miguel, 2007).

It is important to stress that given similar outcomes, differential temporal pathways towards convergence might have dramatic effects on specific social learning scenarios. When there is increasing pressure on actors to accelerate decision making, the importance of partial states increases. For example, time-dependent decision making models have shown partial outcomes to be crucial to determine the quality of research when scholars face the exploration-exploitation dilemma (Chavalarias, 2017). These time-dependent learning metrics can be amplified in multiple choice tasks in which opinion is based on one-shot experience per option, such as wine tasting (Mantonakis, Rodero, Lesschaeve, & Hastie, 2009), or even affect juror's decision making after a sequence of systematically organised sequence of speeches (Shteingart, Neiman, & Loewenstein, 2013; Stone, 1969). Although these scenarios can vary widely, in this study we explore how the time to convergence can be affected by connectivity dynamics.

Using the computer model described below we attempt to model social network structures such that they are characterized by a different schedule, which leads to the formation of differential short-term clusters that might serve as a reasonable proxy for scenarios as those mentioned above. While clustering and social influence in social networks has been studied empirically (Becker et al., 2017; Centola & Baronchelli, 2015), there are relatively few formal models that examine the combined effect of cluster formation and individual influence (Centola et al., 2007; Muthukrishna & Schaller, 2019). These models have successfully shown how the co-evolutionary dynamics of social influence and network structures can affect cultural change. In our model, we explore specific evolutionary dynamics of network structures and individual biases. We implement content biases, coordination biases and memory, three parameters that have been shown as crucial in the acquisition of variants under controlled conditions when using small-scale societies (Tamariz et al., 2014).

2.1.3 Connectivity dynamic and path dependencies

A key innovation of the present study is that we aim to address the effect of different orderings by focusing on a particular microscale manipulation of the social network that remains largely unexplored: The order in which connections between individuals unfold over time. We call this the network connectivity dynamic. The most basic network topology is a fully connected network, in which all nodes are interconnected. In a population of individuals, this means that, over time, each individual interacts with every other individual. However, the same fully connected network may follow different temporal patterns of connectivity, and this may have consequences for variant spread. For instance, a pattern may lead to temporary isolation of one or more sub-populations, generating path dependencies that affect the dissemination of information in the social network, in a similar way to how long-range ties might affect the population dynamics of beliefs, opinions and polarization (Turner & Smaldino, 2018) or stochastic events the evolution of cooperation (Smaldino & Schank, 2012). Importantly for the present study, Turner and Smaldino (2018) show that the order in which agents interact, all else being equal, has a dramatic influence on the long-term polarization of the population. This finding motivates us to propose a reevaluation of a specific small-scale cultural evolutionary model of dissemination of variants by systematically manipulating the effect of the temporal network connectivity dynamics on a range of various outcomes and how this interacts with other previously established factors, such as content-based bias, coordination bias and memory size. In such small sub-populations, drift can reduce diversity, disproportionately favouring variants that happen to be present in the population, and which are not necessarily adaptive (J. Henrich, 2004). A different connectivity pattern may never yield pockets of isolation, which would lead to different evolutionary dynamics. In this chapter, we address how the connectivity dynamic, alone and in combination with content and coordination biases, affects the spread of cultural variants in a population. In order to manipulate the connectivity dynamic, we draw on a well-known pairing methodology called round-robin tournament, which allows us to optimally control that the probability of contribution of each agent to the evolutionary process is the same. This strict control of the connectivity dynamic condition is important to then examine the robustness of our findings by experimenting with the manipulation of agents' cognition. Three different connectivity dynamics are used in this study, each of which yields a different level of temporal sub-population isolation at a given time (see Section 2.2).

To our knowledge, the specific manipulation that we propose here has not been taken into account in experimental work or models on cultural and language evolution, and it would be especially relevant to those researchers that use dynamic interactive micro-societies of agents switching partners over time (e.g. communicative games, cooperative games or tournaments: Baum, Richerson, Efferson, and Paciotti (2004); Byun, De Vos, Roberts, and Levinson (2018); Caldwell and Smith (2012); Fay et al. (2008, 2010); Mesoudi and Whiten (2008); Raviv, Meyer, and Lev-Ari (2019b); Tamariz et al. (2014). In most cases, experimental designs of micro-societies of interacting actors only include one pair composition out of all the possible combinations of pair shuffling, and therefore, outcomes are related with only one specific population connectivity dynamic, potentially affecting the accuracy of the generalizations made by these studies.

We use the experimental design of Fay et al. (2008, 2010) and Tamariz et al. (2014) as a starting point for constructing our model. Tamariz et al.'s (2014) study was designed to test whether the observed variant distributions obtained by Fay et al. (2010) were best explained by neutral drift or showed evidence of selection and adaptation mediated by content and/or coordination biases. Their results indicated an interaction of both biases: Participants displayed egocentric bias and content bias; they tended to produce the variants they had previously produced themselves except when they encountered a better (simpler, cleverer, etc.) variant (through mutation/innovation or via a partner), in which case they tended to adopt it. Additionally, participants seemed to have *short memory* and tended to produce mostly variants that they had seen or produced in the preceding couple of games. Although the model in Tamariz et al. (2014) was useful to show an interaction between content and coordination biases and depict the number of data structures explained by the biases in the experiment, predictions on the specific weight of each bias on the spread of variants were not determined due to the explanatory nature of the model implemented, which took as input the history of the representational variants that the participants had used or seen during the experiment. The interactions between memory and the cognitive biases examined were not explored either.

For the present study, we extend Tamariz et al.'s (2014) study in two important ways. First, we add a new manipulation: Population connectivity dynamic, in addition to content, coordination biases and memory size. We will evaluate the importance of individual parameters and their interactions on the evolution of convergence. Our model does not take experimental data as input, but will be initialised with a random computationally generated distribution of cultural variants. Importantly, we use our model to produce data that can be later used to make predictions or compared to the experimental datasets that have been mirrored. Second, our model aims to be a valid tool to find the best-model fit of connectivity conditions associated to existing experimental datasets that use interactive pairwise micro-societies based on round-robin tournaments applied to the dissemination of cultural variants. This is done by exactly imitating the schedule algorithm of existing experimental designs with human participants. These two innovations allow us to establish causal links between properties at the level of individual agents (content and coordination biases, memory size) and of the population (connectivity dynamics and population size) on one hand, and properties of the culture (evolution and adaptiveness of variants) on the other.

In light of previous research, two important questions are addressed in this chapter. On the one hand, to what extent do the interactions between content bias, coordination bias and memory affect the evolution of cultural diversity in round-robin interactive micro-societies? These analyses replicate and extend previous work. On the other hand, do connectivity dynamics affect cultural diversity in a population, and do they interact with the cognitive biases and memory size, modifying the rate of convergence and the adaptiveness of cultural variants during cultural evolution?

2.2 Methods

We constructed an agent based model that simulates micro-societies of agents who interact in pairs for a number of rounds. We look at the spread of n competing variants, each of which is originally unique and produced by each agent in the simulation. We systematically manipulated the values of several global and individual parameters, including connectivity dynamics and cognitive biases (see Section 2.2.4), and quantified the resulting changes in the evolution of the convergence and diversity of cultural variants (see Section 2.2.5). We also provide comparison of our model by testing our simulated data against the real experimental data collected by Fay et al. (2010) and coded by Tamariz et al. (2014) (see Section 2.4).

2.2.1 Purpose

The purpose of the model is to understand how the interaction between cognitive biases, memory and the order in which agents pair with each other over time affect convergence. It also aims at evaluating the relative importance of each parameter combination and make predictions on the evolution of cultural diversity. Finally, it is tested against experimental data to find the best model-model fit associated to different connectivity conditions.

2.2.2 The model

We consider a simplified micro-society of agents, each of whom is characterized by a number of state variables as described in Table 2.1 on page 28. The micro-society initially contains N agents, who pair-up and interact in pairs for a number of rounds (R). Each interaction consists of a cultural variant exchange (Figure 2.1). We simulated pairwise interactive micro-societies of $N = 8, 16$ and 32 agents, allowing us to track all agent pairings during $N - 1$ rounds according to the pair composition illustrated in Figure 2.2. This type of scheduling algorithm is called round-robin tournament, and it allows every agent to be paired with every other agent for exactly one round. In our micro-societies, N is even, which means that in each $N - 1$ rounds, $N/2$ games (or pairwise interactions) can be run concurrently.

Different algorithms can be used in order to create a schedule for a round-robin tournament and all of them contain the same number of games $N(N - 1)/2$. In our model, agents interact in each game by exchanging information, therefore not all the schedules allow the same spread of adaptive information, because different schedules can pair-up agents forming different levels of sub-population isolation at different times. We implement three different connectivity dynamics that differ with regards to how many agents could potentially share the same variant at a given time. In other words, each connectivity dynamic examined (early, mid, late) differ from the others in how fast a cultural variant could potentially spread out among the agents of the micro-society (Figure 2.2).

In our model, each agent i has its individual cultural attributes defined as a matrix of cultural variants. The state of an agent i is a $2 \times r$ matrix of cultural variants, which we call A,

$$A = \begin{bmatrix} \sigma_{ie1} & \sigma_{ie2} & \cdots & \sigma_{ier} \\ \sigma_{ia1} & \sigma_{ia2} & \cdots & \sigma_{iar} \end{bmatrix} \tag{2.1}$$

where σ_{ier} corresponds to a cultural variant produced by agent i in round r, and σ_{iar} corresponds to a cultural variant observed by agent i in round r. Thus, the state of agent i is its record of produced and observed cultural variants, which are stored in a two-row and r-column matrix. Produced variants are stored in the memory as egocentric, designated as e. Observed variants are stored in the memory as allocentric, designated as a.

2.2.3 Model dynamics

The dynamics of the model are defined by the following steps. In the initial state each agent i is randomly assigned a cultural variant σ_i selected from a pool of n distinct cultural variants without repetition. Each variant originally assigned to each agent is unique, and it is the first variant they produce in their first interaction at round $0, r_0$. The initial pool of variants in the micro-society can be defined as a V vector of length N, that contains one cultural variant for each agent $(\sigma_{i0}, \sigma_{j0}, \ldots, \sigma_{n0})$, where σ_{i0} is the variant initially assigned to agent i, σ_{j0} the variant initially assigned to agent j, and so on.

At the beginning of each round r, agents are paired using one of the three connectivity dynamics (*early, mid, late*) represented below (see *population connectivity dynamics*), each of which describes a different schedule to pair-up agents. Each connectivity dynamic examined differ from the others with respect to how many agents can potentially share the same cultural variant at one given round.

Once agents are paired, at each round r, they interact by presenting and observing one cultural variant. Within each pair (Figure 2.1), each agent in turn samples its memory to produce a variant (Figure 2.1, top) according to the probabilistic function defined in Equation (2.2) on page 27. At this point, there is a small probability of innovation, in which case, the variant produced will be randomly sampled from one of the n variants the population was initialised with. Then, both agents add both variants to their memories (Fig. 2.1, bottom). That is to say, at round r, when agent i and agent j interact, i produces variant σ_i and j variant σ_j. Both agents store the two variants in memory. Agent i stores variant σ_i in its egocentric memory as σ_{ier} and σ_j in its allocentric memory as σ_{iar}. Agent j stores variant σ_j in its egocentric memory as σ_{jer} and σ_i in its allocentric memory as σ_{iar}.

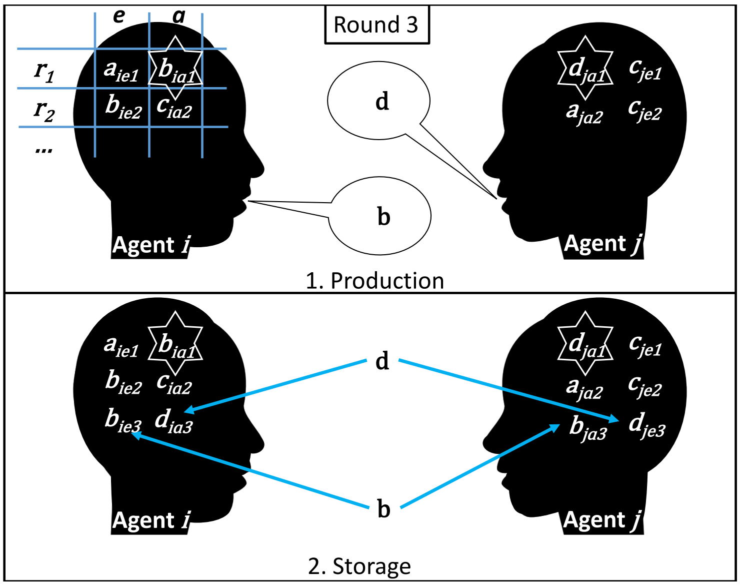


Figure 2.1: Illustration of the interaction between a pair of agents at one round of the simulation. At round 3, agents have already some variant tokens in memory, which they have stored in previous rounds. For each variant in memory, the letter represents the type (e.g. **a**, **b**). The first index represents the agent's identity; the second index represents whether the variant token is stored as allocentric (variant produced by other agent) or egocentric (produced by self). (Note that if an agent produces a token of a variant type that in the past was stored as allocentric, the token is stored as egocentric in the present round.) The third index indicates the round at which the variant was stored. During the round depicted (round 3), first, one variant token in each agent's memory is selected for production (in the figure, the tokens surrounded by a star). Once both agents have produced a new token of the selected variant, they proceed to store them in their memory. Their own variant is marked as egocentric, and as produced at round 3; the other agent's variant is marked as allocentric and also as produced at round 3.

2.2.4 State variables and probability distribution of variants

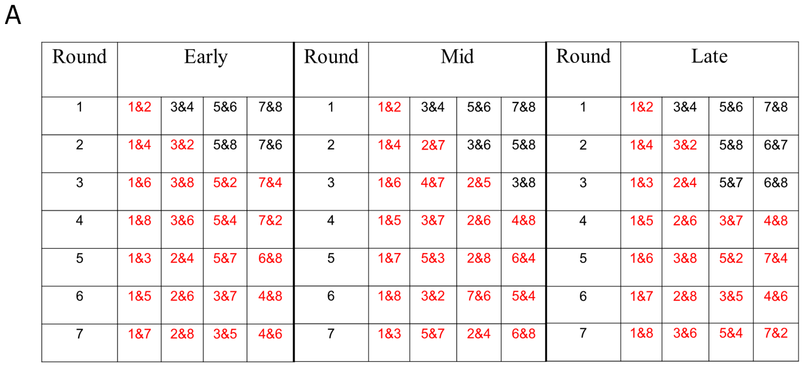
The model takes several parameters as described below:

(a) Number of agents (N): we simulate micro populations of $N = 8$, $N = 16$ and $N = 32$ agents.

(b) Number of rounds (R): This depends on the number of agents. A run of the simulation includes $N-1$ rounds, which allows every agent to be paired with every other agent for exactly one round. A single round is designated as r.

(c) Population connectivity dynamics. This reflects the order in which the agents are paired with each other. Different orderings yield different levels of sub-population isolation at different times (Figure 2.2). For example, in our 8-agent micro-societies, three different connectivity dynamics can be described with reference to how many agents could potentially share the same variant by round 3: In the early connectivity condition, all 8 agents could share the same variant by round 3; in the mid connectivity condition, 6 agents could share same variant by round 3; in the late connectivity condition, 4 agents could share the same variant by round 3. We examine 3 connectivity dynamics: Early (E), mid (M), late (L), in 8, 16 and 32-agent micro-societies. Due to the wide range of possible permutations for mid connectivity in the larger micro-societies, in this case, for each run, we use a random selection among possible permutations.

(d) Coordination bias (c) captures the extent to which an agent has a preference for self-produced variants or partner-produced variants. It fixes the likelihood of a variant being produced depending on whether it originates in others or the agent itself. It takes values from 0 (fully egocentric: Preferring self-produced variants over other-produced variants) to 1 (fully allocentric: Preferring other-produced variants over self-produced variants). The strength of the coordination model is determined by the combined mathematical complement of the content bias and the innovation rate. That is, for each agent and round, variants are selected for production using the probability of the complement of each of the other cognitive features. When coordination bias is 0.5, we have a neutral coordination model where variants in the egocentric and the allocentric memory are equally likely to be produced. Coordination bias values from 0 to 1 in steps of 0.1 are examined.



Round	Early				Round	Mid				Round	Late			
1	1&2	3&4	5&6	7&8	1	1&2	3&4	5&6	7&8	1	1&2	3&4	5&6	7&8
2	1&4	3&2	5&8	7&6	2	1&4	2&7	3&6	5&8	2	1&4	3&2	5&8	6&7
3	1&6	3&8	5&2	7&4	3	1&6	4&7	2&5	3&8	3	1&3	2&4	5&7	6&8
4	1&8	3&6	5&4	7&2	4	1&5	3&7	2&6	4&8	4	1&5	2&6	3&7	4&8
5	1&3	2&4	5&7	6&8	5	1&7	5&3	2&8	6&4	5	1&6	3&8	5&2	7&4
6	1&5	2&6	3&7	4&8	6	1&8	3&2	7&6	5&4	6	1&7	2&8	3&5	4&6
7	1&7	2&8	3&5	4&6	7	1&3	5&7	2&4	6&8	7	1&8	3&6	5&4	7&2

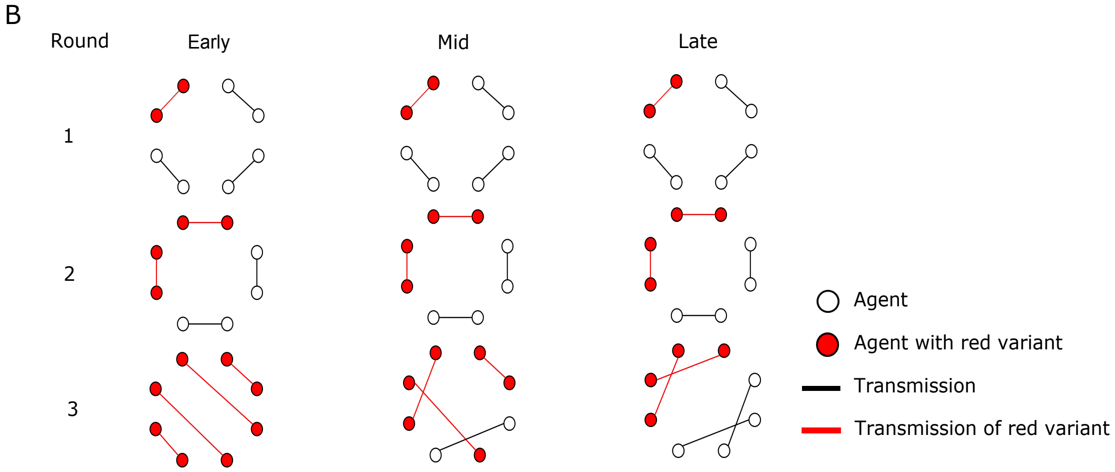


Figure 2.2: A. Examples of agent pairings for each type of connectivity dynamic, in a micro-society of 8 agents. In red, agents who may, in the current round, potentially have in their memory the variant produced by agent 1 in round 1. B. Three different network connectivity dynamics can be described within a pairwise interaction account for a dynamic fully connected network of 8 agents. By round 3, potentially 8 (in the early connectivity dynamic), 6 (in the mid connectivity dynamic) or 4 (in the late connectivity dynamic) agents share the red variant.

(e) Content bias (β) and variant quality (s) are two closely related parameters that determine variants selection due to their intrinsic value. Parameter s corresponds to the intrinsic value of each communicative variant and it indicates to what extent the variant is preferred over the other variants (s is 1 if the signal is preferred over the others, 0 otherwise). Variant quality is a binary trait and the population is initialized with one preferred variant, which is determined at random at the initial state of each simulation. This is motivated by the types of scenarios we aim to simulate, in which, in general, there is a tendency to converge on one or few shared variants. We assume all the agents have the same initial preference. However, this assumption does not mean that agents always choose high quality variants, since variant choice is defined by a probability distribution function which, in addition to content biases, is determined by other cognitive parameters and the frequency of variants in agents' history (see Equation (2.2) on the facing page). In the present study, we assume that the preferred variant is adaptive, that is, it has (by definition) higher probability of being produced. In real life, adaptiveness may be due to a variant being easy to produce, to memorise or to process, attractive, effective or efficient for a desirable function, etc. For adaptiveness measures, the frequency of high quality variants ($s = 1$) is considered (see Section 2.2.5). Parameter β captures agents' sensitivity to variant quality (s), and ranges from 0—not sensitive at all—to 1—fully sensitive—in steps of 0.1. Parameter β is operationalised only if the target variant σ_i has been produced or seen at least once—in other words, one cannot prefer to re-produce a variant due to its quality until one has been exposed to that particular representation. Thus, given a variant σ_i, the product $\beta \times s_i$ determines the probability that variant σ_i is selected due to its intrinsic value, and this product ranges from 0 to 1, in steps of 0.1. Note that, at each round, as explained in the model dynamics, each agent stores two variants in its memory (the one that was produced by the agent itself and the one that was produced by its partner). Thus, content bias does not affect the storage process in our model (agents store in memory all the variants they encounter), it only affects agents' production. In simple words, the higher the value of β, the higher the probability of producing a cultural variant due to its quality. When content bias is 0, we have a content drift model. Content bias values from 0 to 1 in steps of 0.1 are examined.

(f) Memory size. We manipulate memory size (m) by limiting agents' access to their memory store. It corresponds to the maximum amount of history that can affect the variant choice. At each round, when an agent has to produce a variant, it can only choose between the variants that were stored in the preceding rounds. A memory size of m means that the model remembers at most the last m egocentric entries $h \mid E,m$ and the last m allocentric entries $h \mid A,m$ from any history h. The relative frequencies of variants in $h \mid E,m$ define the egocentric initial distribution $f(h \mid E,m)$ and in $h \mid A,m$ the allocentric distribution $f(h \mid A,m)$.

(g) Innovation rate (μ). Agents can generate novel variants. We fix the innovation probability at 0.02 by using a flat distribution $\phi(x)$ weighted by innovation rate, which means that 98% of variant choices reflect the probability distribution yielded by all the parameters described above, while 2% is a random choice among all 8 initial variants. We have chosen this value for two main reasons. First, for consistency with the innovation rate found in Fay et al. (2010) experimental data by Tamariz et al. (2014). Second, because in the present study we attempt to find the best model fit associated to that same dataset.

For each round in the simulation, for each agent, the model yielded a probability distribution of variants (x) for a given history (h) of previous rounds, according to the following equation. We use the apostrophe ($'$) to denote the probabilistic complement: $a' = 1 - a$.

$$\Pr(x \mid h) = \mu' \beta' c' f(x \mid h_{|E,3}) + \mu' \beta' f(x \mid h_{|A,3}) + \mu' \beta s + \mu \phi(x) \tag{2.2}$$

We run the simulation with 1452 different parameter value combinations. For each parameter combination, we ran the simulation 1000 times. For each model run we assume that all agents have the same connectivity dynamic, biases and memory sizes. The results below show the average and standard deviations of the number of runs of each parameter combination examined. All parameters and state variables can be found in Table 2.1 on the next page.

Model parameters				
Entity	Parameter	Symbol	Number of levels	Value(s)
Agent	Content bias	β	11	0.0 to 1.0 in steps of 0.1
	Coordination bias	c	11	0.0 to 1.0 in steps of 0.1
	Memory	m	4	1, 3, 5, 7
	Innovation	μ	1	0.02
	Variant quality	s	2	[0,1]
	Agents' sensitivity to variant quality	b	11	0.0 to 1.0 in steps of 0.1
	Variant in agent's memory record	d	2	[0,1]
	Flat distribution of variants	$\varphi(x)$		
	Egocentric entries for a given history	$h_{\|E,m}$		
	Allocentric entries for a given history	$h_{\|A,m}$		
	Distribution of egocentric variants for a given history	$f(h_{\|E,m})$		
	Distribution of allocentric variants for a given history	$f(h_{\|A,m})$		
	Probability distribution of variants for a given history	$P(x\|h)$		
	Agent state (agent cultural attributes)	A		
	Cultural variant	σ		
	Agent ID	$i,j,\ldots,n$		
	Produced variant by agent *i*	σ_i		
	Stored variant in egocentric memory by agent *i*	σ_{ie}		
	Stored variant in allocentric memory by agent *i*	σ_{ie}		
Global	Number of agents per group	N	3	8,16,32
	Initial pool of variants	V	3	8,16,32
	Number of competing variants	n	3	8,16,32
	Number or rounds	$R=N-1$	3	7,15,31
	Round	r		
	Initial round	r_0		
	Number of games per round	$N/2$	3	4,8,16
	Number of games per run	$N(N-1)/2$	3	28,120,496
	Connectivity dynamic	*Early (E)* *Mid (M)* *Late (L)*	3	E, M, L

Table 2.1: Parameters, state variables and scales.

2.2.5 Quantifying convergence and adaptiveness

At its most general level, evolution is defined as a change in the frequencies of different variants in a population over time. We are therefore interested in the composition of the pool of 8, 16 or 32 variant tokens produced by the agents at each round, and how it changes over rounds. We examine the level of convergence in this pool, or the extent to which agents used the same variant. Following others (e.g. Fehér, Wonnacott, & Smith, 2016; Smith & Wonnacott, 2010), we quantify convergence using the information-theoretic notion of entropy (H) (Shannon, 1948):

$$H(V) = -\sum_{v_i \varepsilon V} p(v_i) \log_2 p(v_i) \tag{2.3}$$

where V corresponds to the set of variants, and $p(v_i)$ is the probability of i^{th} variant in that set. Entropy is a well-established alpha diversity index that has been used to measure cultural diversity. High entropy corresponds to low convergence.

Evolution, even by drift, may increase convergence (and decrease entropy), as random sampling at each round gradually eliminates variants from the pool (and our low level of innovation is not enough to compensate for that). For example, at round 0, where each agent produced its own unique variant, the probability distribution over the 8 variants was flat (each variant had a probability of $1/8$) and the entropy was maximal ($H = 3$ bits). Over time, as agents converged, entropy would decrease; if, say, by round 7, the probability of 1^{st} variant was 0.75, the probability of 2^{nd} variant was 0.25 and the probability of the remaining variants was 0, the entropy would be 0.811 bits.

To better understand the mechanics of evolutionary algorithms, a number of studies have investigated time to convergence (TC). The number of rounds until convergence has been used to analyze convergence properties of genetic algorithms in studies about population sizing, network structures and theory of convergence (e.g. Mueller-Frank, 2013; Olfati-Saber & Murray, 2004; Pelikan, Goldberg, & Cantú-Paz, 2000). This additional measure is important because it allows us to predict more accurately the moment at which one population will reach convergence under different conditions. Additionally, it gives us more information about how relevant agents' choices were in the first rounds. Therefore, in some analyses we will also present time to convergence (TC) or the number of rounds it takes for the population to reach full convergence (defined as $H = 0$ bit) for the first time.

Researchers on cultural evolution have developed models that link demography and cultural adaptiveness, using a variety of mathematical approaches (J. Henrich, 2004; Mesoudi, 2011b; Shennan & Wilkinson, 2001). These models describe how a trait changes in frequency over time. We calculated the adaptiveness (A) of the cultural system at each round (t) as the frequency of high-quality variants (quality is measured by s, see above) at that round,

$$A(t) = \frac{n_h(t)}{N(t)} \tag{2.4}$$

where $n_h(t)$ is the number of high quality variants at round t and $N(t)$ is the total set of variants produced in round t.

Two considerations apply to our study when we calculate average change in high quality variants. First, we use relative fitness equations to account for the adaptiveness of high quality variants, that is to say, we only consider the frequency of cultural variants having $s = 1$. Second, we assume that cultural variants are distributed, and therefore transmitted at each round, according to our parametrized model, as defined in Equation 2.2. Thus, in our case, the change in the adaptiveness (ΔA) of high quality variants due to selection follows immediately from our definition of adaptiveness. This tractability assumption simplifies our equation considerably, because now the change in adaptiveness equation reduces to:

$$\Delta A = A(t+1) - A(t) \tag{2.5}$$

where change in adaptiveness (ΔA) takes the difference between the adaptiveness in the subsequent round $A(t+1)$ and the adaptiveness in the earlier round $A(t)$. Therefore, a change in adaptiveness above 0 ($\Delta A > 0$) indicates that the fitness of high quality variants produced by agents increased from one round to the next. When $\Delta A = 0$, variant frequency was stable from round to round.

2.3 Results

In this section we offer a summary of the results of two selection models (content bias and coordination bias) against a drift model, and how they interact with each other (Section 2.3.1). Next, we show the effects of memory limitations (Section 2.3.2). Figures for these analyses can be found at the end of each section. Additional analyses on conditional entropy distributions can be found in Appendix A. In Section 2.3.3, we focus on the effects of population connectivity dynamics on entropy, time to convergence and change in adaptiveness of the cultural system from round to round. We also pay special attention to the interplay between connectivity and two strong drivers of convergence: Content bias and population size. We use linear and non-linear regressions to fit models to our data to establish the relationships between variables (see Appendix A). However, following White, Rassweiler, Samhouri, Stier, and White (2014), we use frequentist statistical models only to calculate effect sizes in our multifactorial simulations, but we do not report *p*-values, which can be meaningless when applied to simulation model output. In the following analyses we show mean values and standard deviations (Mean ± SD).

2.3.1 Cognitive biases

We ran simulations manipulating the level of content bias (β). When compared with a neutral content bias model ($\beta = 0$), content bias increased convergence (decreasing entropy). Mean entropy was greatest when $\beta = 0$ (2.451 ± 0.448 bits) and lowest when $\beta = 1$ (1.020 ± 1.136 bits). Similarly, when keeping coordination bias at a neutral level level (coordination bias = 0.5), mean entropy was greatest when $\beta = 0$ (2.277 ± 0.478 bits) and lowest when $\beta = 1$ (1.015 ± 1.137 bits).

When we considered coordination bias (c) alone ($\beta = 0$), egocentric bias reduced convergence. Mean entropy was greatest when $c = 0$ (strongest egocentric bias) (2.013 ± 1.060 bits) and lowest when $c = 0.5$ (neutral coordination) (1.768 ± 0.956 bits). Similarly, when keeping a neutral content bias ($\beta = 0$), mean entropy was greatest when coordination bias was fully egocentric, 0 (2.906 ± 0.171 bits) and lowest when coordination bias was neutral, 0.5 (2.277 ± 0.478 bits).

Both β and c had effects on convergence that differ from a drift model; see Figure 2.3. The effect of c on entropy was different for each level of β, revealing an interaction. Average entropy was highest when $\beta = 0$ and c was also 0 (weakest content bias and strongest egocentric bias) (2.906 ± 0.171 bits) and lowest when $\beta = 1$ and $c = 0.5$ (1.015 ± 1.136 bits). When agents' behavior was strongly content-biased, the rate of convergence increased, masking the effect of c. Conversely, weaker content biases allowed c to show its effect on convergence, which can be characterized by a distinctive asymmetric distribution. The slowing effect of c on the rate of convergence becomes hidden as β rises.

Our main result here, which establishes a baseline for posterior analyses, is that the population reaches more convergence when agents are more sensitive to the intrinsic value of variants. It is obvious that a content bias, as implemented in our model, dramatically affects the spread of variants in a population and that this result is consistent with previous studies showing the crucial effects of the individuals' direct assessment of the value of traits on adoption and transmission (Fay et al., 2010; J. Henrich & McElreath, 2007; Stubbersfield, Tehrani, & Flynn, 2015; Vale et al., 2017); for a review see R. L. Kendal et al. (2018). It is also interesting to note how the effect of β is modulated by c, which plays an important role in maintaining variation. There are two ways in which c acts against convergence:

(1) A strong egocentric bias preserves sign variation at the level of the population by inhibiting the adoption of others-produced variants (i.e. fully egocentric agents keep producing their original egocentric variant indefinitely, unless innovation occurs). This increases diversity globally because each agent tends to over-produce their own individual variant, reducing the probability of local convergence.

(2) A strong allocentric bias preserves sign variation at the level of the population by inhibiting the production of variants stored as egocentric (i.e. fully allocentric agents keep producing other-produced variants). This increases diversity globally because each agent tends to over produce variants that are stored in their allocentric memories.

Here, it is important to note that an intermediate level of coordination bias facilitates convergence by allowing agents to align more easily on common shared representations. Too allocentric or egocentric scenarios make these conventions unlikely. However, fully egocentric agents are more efficient than allocentric agents at preserving variation. This is because the allocentric memory of allocentric agents contains more shared variants with other agents (there are seven other agents) than the egocentric memory of egocentric agents (there is only one ego). Consequently, the probability that allocentric agents converge is higher than the probability that egocentric agents converge, which explains the asymmetry of the coordination model (Figure 2.3).

However, as β increases, it overrides the capacity of egocentric and allocentric biases to preserve diversity. This effect is important as it highlights both the importance of coordination when the adoption of variants is not based on competition for variant quality, and its irrelevance in scenarios that encourage variant adoption based on the intrinsic properties of the variants. Our model, therefore, assigns different weights and roles to each selection pressure, with β as the main driver of convergence—encouraging selection owing to intrinsic variant quality, and *c* as a modulating pressure—favouring variation when β is weaker. Given these baseline dynamics, in the next sections we will analyse how model outcomes are modulated by agents' memory size and network connectivity dynamics.

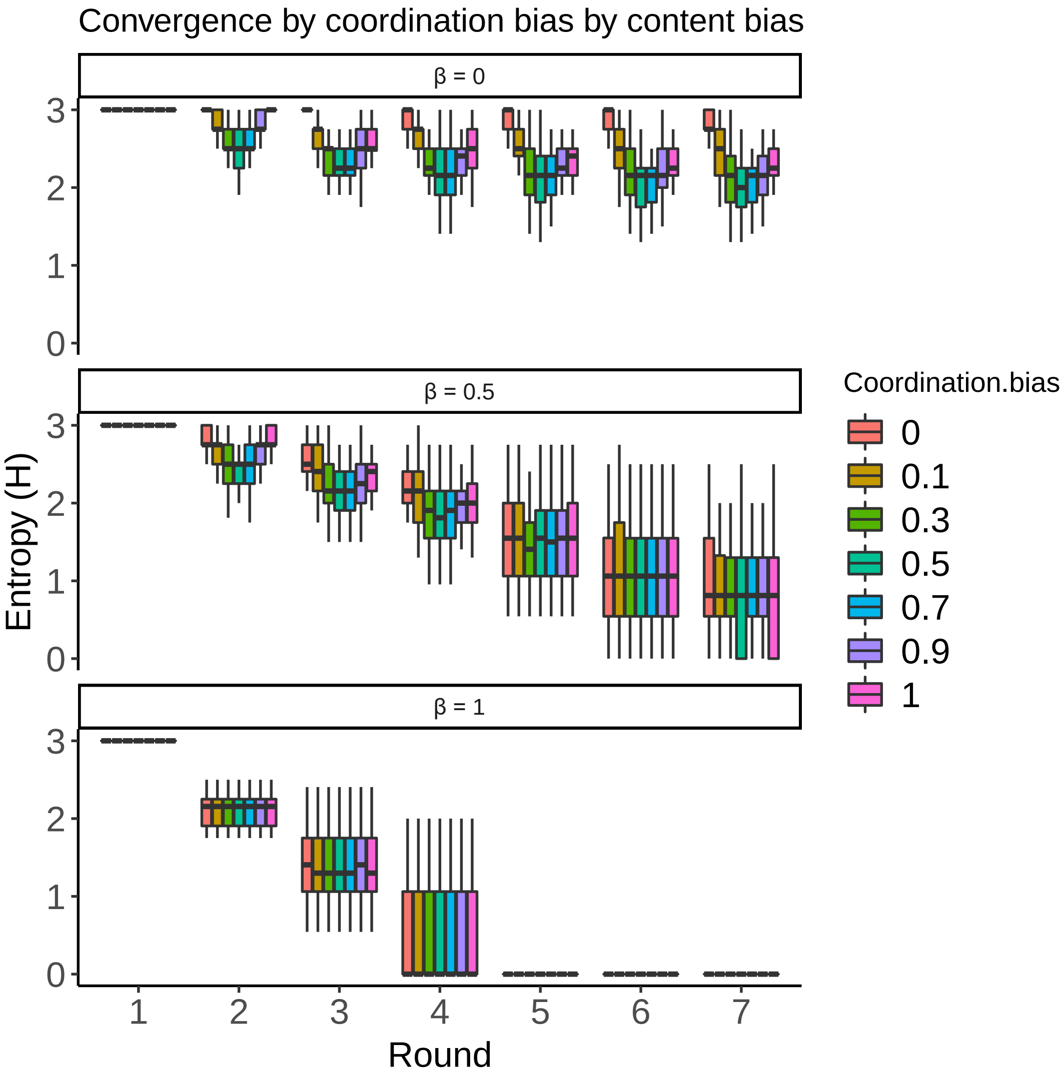


Figure 2.3: Convergence (measured as Sahnnon entropy H) by round by each combination of biases. Examples for content bias 0, 0.5 and 1. A drift model has a content bias of 0 and a coordination bias of 0.5. X-axis represents rounds from 1 to 7, Y-axis represents entropy in bits. In this and subsequent boxplots: middle line is median, 50% quantile; lower hinge, 25% quantile; upper hinge, 75% quantile; lower whisker is smallest observation greater than or equal to lower hinge $- 1 * IQR$; upper whisker is largest observation less than or equal to upper hinge $- 1 * IQR$.

2.3.2 Memory

Memory size increased entropy and therefore decreased convergence. Average entropy was greatest in the absence of memory limitations, when agents kept in memory all the variants they had been exposed to $(1.935 \pm 0.920$ bits$)$. In contrast, when we limited the agents' memory to the most recent 5 rounds $(1.920 \pm 0.930$ bits$)$, 3 rounds $(1.834 \pm 0.990$ bits$)$ or 1 round $(1.675 \pm 1.081$ bits$)$, entropy decreased noticeably.

This effect of memory on entropy was more noticeable for intermediate values of content bias ($\beta = 0.5$). When content bias was strongest, memory effects tended to be masked by a floor effect at the lower end of the entropy distribution in the last rounds. On the other hand, when content bias was neutral ($\beta = 0$), memory effects were greater for neutral values of coordination bias ($c = 0.5$), when compared with strongly egocentric or allocentric bias (Figure 2.4).

In our analysis of how memory interacted with cognitive biases and influenced convergence, our initial expectation was that, in general, memory limitations would reduce variation. These predictions were based on studies showing that memory limitations can lead to regularisation in concurrent frequency learning tasks, which is possibly attributable to limited working memory (Ferdinand et al., 2013; Hudson Kam & Chang, 2009; Hudson Kam & Newport, 2005; Kareev, Lieberman, & Lev, 1997). Our model is consistent with those findings, suggesting that a larger memory size decreases convergence by delaying social learning and, in turn, the spread of variants with high intrinsic value. Interestingly, however, our results show that a reduction in variation seems to be true in all scenarios except when the population is fully allocentric in a null content-biased model. This pattern of inversion suggests that allocentric agents with larger memory sizes might increase the probability of convergence on shared variants in scenarios where variant adoption is not encouraged on the basis of intrinsic variant quality.

As with coordination bias, the effect of memory fades when β rises to levels close to 1. This is because when the population is fully biased towards the current value of a given variant, memory size became less important as a selection mechanism for variants adoption. In contrast, as β decreases, agents begin to increasingly activate coordination as a selection mechanism, by paying more attention to whether variants are stored in egocentric and allocentric memories. The effect of memory size, thus, becomes crucial to determine which variant is chosen by each agent at each time step since it determines the size of the pool of variants that each agent can track back in time in its memory from a given moment. Since variation in convergent processes is in general greater in larger pools, a reduction in memory size yields higher levels of convergence in our model, suggesting an increased reliance in social learning, which helps variants with high intrinsic value to propagate.

As we have shown in the previous section, extreme coordination biases preserve variation by inhibiting the adoption of variants produced by others or by the agent itself. This explains why convergence is lower when $\beta = 0$, regardless of the level of memory implemented. However, the effect of memory is more noticeable for intermediate values of β. This is simply because in this intermediate scenario in which both β and c coexist, β is high enough to substantially increase convergence (yielding a large range of possible mid-convergence outcomes) but not strong enough to eliminate the effect of memory.

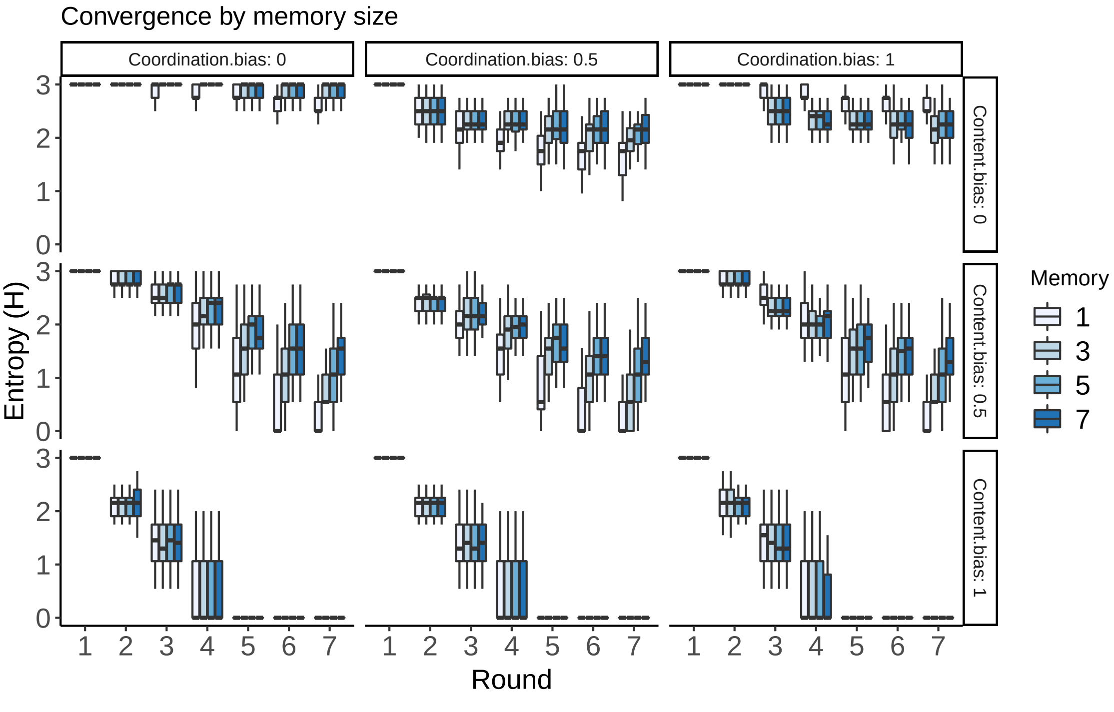


Figure 2.4: Convergence (measured as Shannon entropy H) by round by memory. Examples for content bias 0, 0.5 and 1 and for coordination bias 0, 0.5 and 1. X-axis represents rounds from 1 to 7, Y-axis represents entropy in bits. In this and subsequent boxplots: middle line is median, 50% quantile; lower hinge, 25% quantile; upper hinge, 75% quantile; lower whisker is smallest observation greater than or equal to lower hinge $- 1 * IQR$; upper whisker is largest observation less than or equal to upper hinge $- 1 * IQR$.

2.3.3 Connectivity dynamic of the population

A key finding in this study is that population connectivity dynamic affected the spread of variants. Convergence was delayed in populations that took longer to reach full connectivity. But this socio-structural effect was only manifested under certain conditions related to the cognition of individual agents. When running simulations using a null content bias model, entropy remained similar for all levels of connectivity (Figure 2.5). Interestingly, however, increasing content bias in the agents revealed a substantial effect of the connectivity dynamic on convergence. Mean entropy differences between conditions were greatest at round 3: 0.805 ± 0.730 bits under late connectivity, 0.464 ± 0.498 bits under mid connectivity, 0.133 ± 0.286 bits under early connectivity. Contrary to what happens with memory and coordination bias, the modulating effect of the network connectivity dynamic was stronger for high-moderate content bias. This is because the spread of variants with high intrinsic value is facilitated by the assessment of variants value based on the nature of the information itself but restricted by the conditions of accessibility to that same information imposed by the network connectivity dynamic. Thus, connectivity dynamics are almost irrelevant in scenarios with null or very low content bias, while they yield substantial effects on the pathways towards convergence for moderate and high content bias, with late connectivity delaying convergence and maintaining variation in the population for a longer time period. The effect of connectivity dynamic was more pronounced in larger populations, where we can observe that the delay due late connectivity dynamic lasts for a larger number of rounds when we increase population size (Figure 2.5). In content-biased populations, as population size increases, mean entropy differences between conditions remained significantly high for longer time periods (e.g. when $\beta = 0.8$, the relative difference between conditions remained above 10% for more than 6 rounds in 16-agent micro-societies, and for more than 11 rounds in 32-agent micro-societies).

Summarising, both content bias (by strengthening the selection of high quality variants) and population size (by lengthening the time to convergence) amplified the effect of the connectivity dynamic, and this in turn resulted in a deep alteration of the evolutionary trajectory of convergence. In these scenarios, late connectivity populations clearly show periods of rapid convergence followed by periods of relatively slower convergence, resembling punctuational evolutionary dynamics. In contrast, convergence in early connectivity populations was not affected by these evolutionary bursts and tended to be shaped by a monotonic sigmoid curve.

Although the effects of network connectivity dynamics examined here tend to vanish in the long run, it is important to characterize how fast a group of agents reach the state of convergence. Time to convergence (TC) is a widely used performance indicator in processes that require convergence as an outcome. Differential temporal pathways towards convergence between conditions might have dramatic effects on social learning scenarios in which outcomes are

(a) either time-dependent or based on intermediate states (e.g. strategic or argumentative scenarios in which the first consensus is the one that is considered), or

(b) affected by outcome primacy (e.g. the jurors' decision after a sequence of argumentative speeches where earlier data have more impact on behavior than later data).

Although these scenarios can vary widely, we hypothesised that TC can inform how connectivity dynamics may alter the outcome of such social processes. We examined the effect of connectivity on the TC in those simulations where full convergence ($H = 0$) was reached. Figure 2.6 depicts how fast micro-societies converged on a shared cultural convention in those simulations. The time required for a population to reach full convergence was longer in the late connectivity condition when compared to the mid and early connectivity conditions. This shows that the effects of the network connectivity dynamics examined here are strong enough to alter convergence performance in scenarios in which a common consensus is reached in the population. In other words, the outcome of social learning processes that are strongly restricted by time is substantially affected by these changes in the network connectivity dynamic. This is particularly true for moderate-high levels of β. However, coordination bias and memory do not seem to interact with the connectivity dynamic when it comes to explaining time to convergence (Appendix A, Figures A.12 and A.13). This is because the effect of these cognitive parameters on convergence is very limited compared to that of content bias, which is the main driver of convergence.

Similarly, the adaptiveness (A) of the cultural system increased more rapidly in populations with early connectivity. The change in adaptiveness of high quality variants remained above 0 across rounds, indicating that the proportion of high quality variants always increased from round to round.

However, changes in adaptiveness followed different patterns in populations with early, mid and late connectivity (Figure 2.7). Populations with late connectivity evolved in punctuated bursts of change followed by periods of slower change. For instance, in 8-agent micro-societies, at least 2 rapid bursts of change in the proportion of high quality variants can be observed before the population became a fully connected network in 7 rounds. As above, these patterns can be better observed when we increase population size. Bursts of rapid change are related to the evolutionary moments in which the pockets of isolated agents created by the late connectivity become connected. On the other hand, populations with early connectivity dynamics followed a single-peak evolutionary dynamic. This is due to the fact that high-quality variants could spread in the system continuously (and without any additional restriction imposed by the connectivity dynamic) until the population reached its equilibrium.

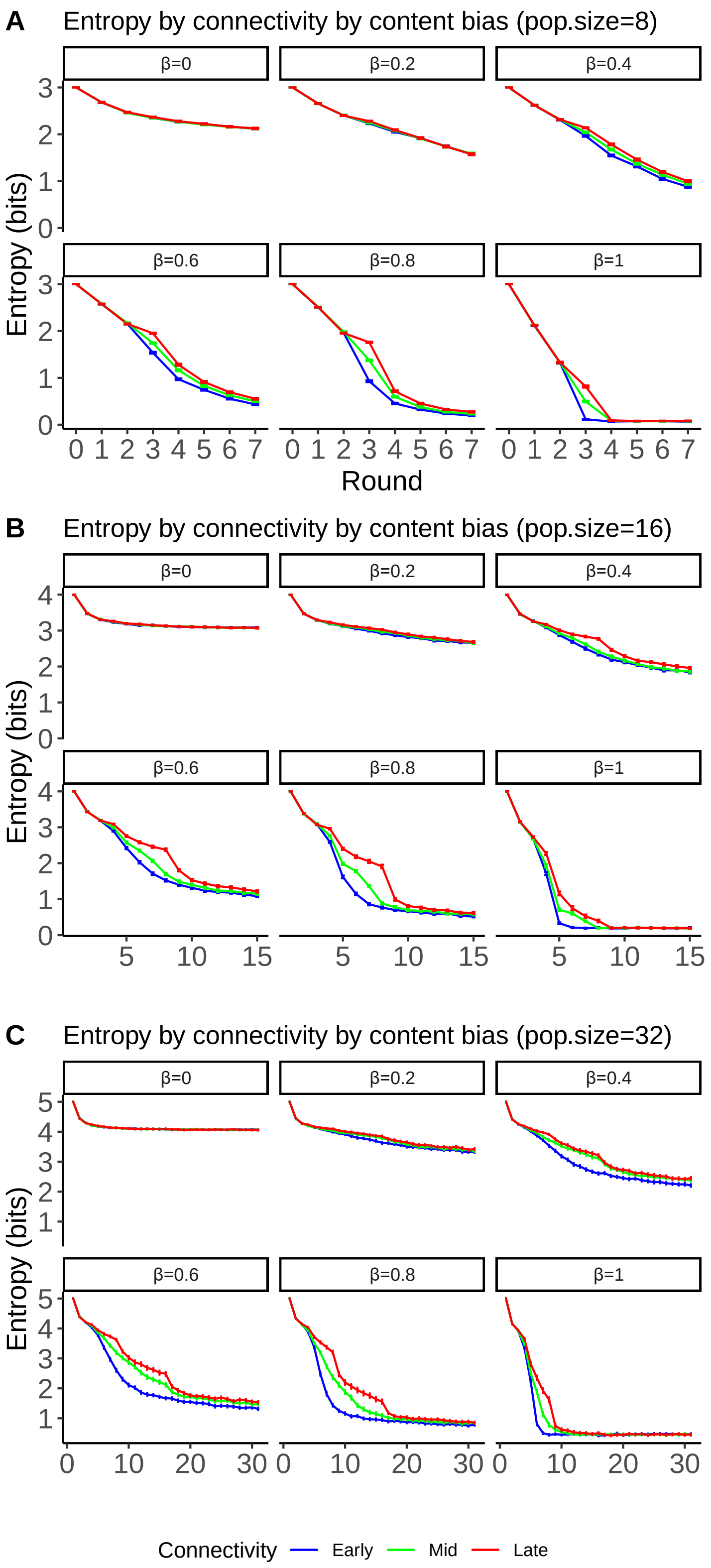


Figure 2.5: Convergence (measured as Sahnnon entropy H) averaged over each level of connectivity and content bias, for population size = 8 (A), population size = 16 (B) and population size = 32 (C). The x-axis represents rounds, and the y-axis represents entropy in bits. Drift models are shown in the top-left ($\beta = 0$). We ran 1000 simulations for each parameter combination. Error bars indicate 95% CIs.

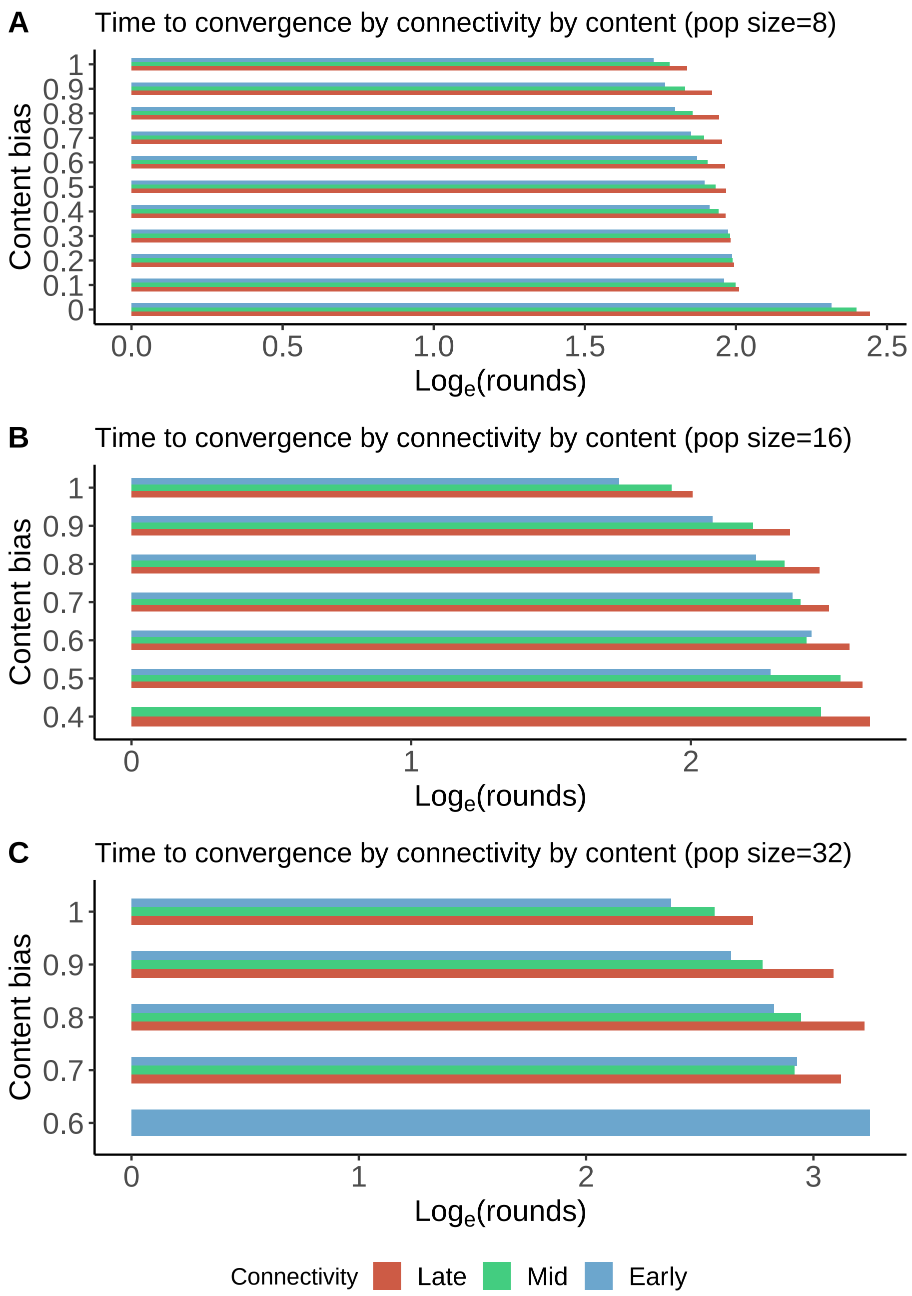


Figure 2.6: Time to convergence—measured as the natural logarithm of the number of rounds it takes for the population to reach full convergence (defined as $H = 0$ bit) for the first time—averaged over each level of connectivity and content bias, for population size = 8 (A), population size = 16 (B) and population size = 32 (C). We only considered micro-societies that reached full convergence. Scenarios like this did not occurred during the round-robin simulation when content bias was relatively low and population size was 16 or 32 agents. Late and mid connectivity delay time to convergence when compared to early connectivity. This is particularly true for high-moderate levels of content bias. We ran 1000 simulations for each parameter combination.

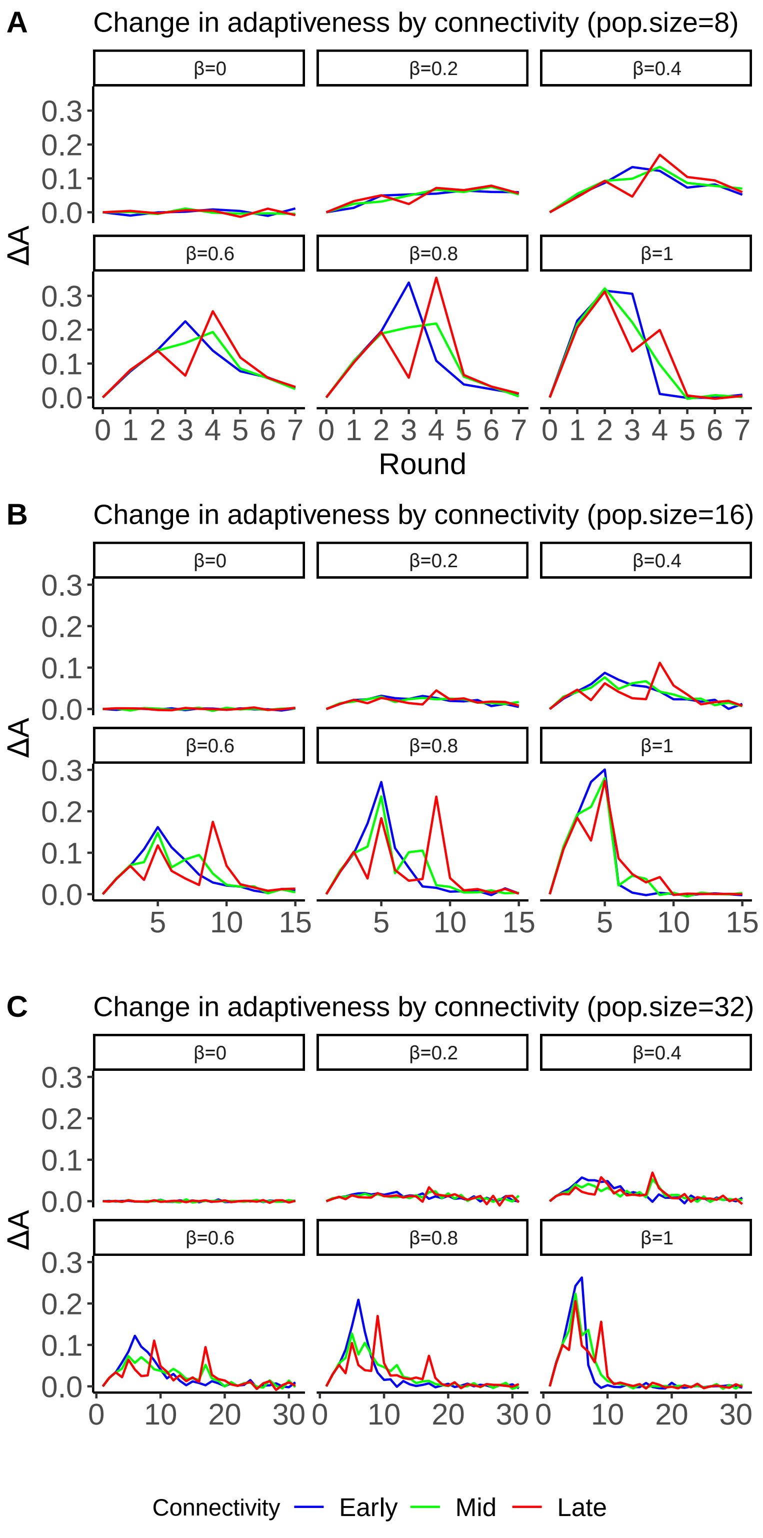


Figure 2.7: Change in adaptiveness ΔA of high quality variants by round, averaged over each level of connectivity and content bias, for population size = 8 (A), population size = 16 (B) and population size = 32 (B). Results above 0 indicate that the proportion of high quality variants increased relative to the previous round. When $\Delta A = 0$, variant frequency was stable from round to round.

2.4 Model comparison against experimental data

We collect simulated data that includes the parameter combination of the best fit-models for the 64 data structures coded by Tamariz et al. (2014). The coded data structures correspond to the experimental data collected by Fay et al. (2010) from four distinct 8-person communities and 16 concepts used in a Pictionary-like game, yielding a total of 64 data structures. A verification of the model and its explanatory power is provided in Tamariz et al. (2014). The data coded in Tamariz et al. (2014) for each generation is equivalent to the data produced by the simulator just after the interaction of the agents in each round, that is to say, once the agents have in memory the variants produced by their partners. This can be easily verified by observing the relatively high levels of convergence $(2 < \text{entropy} < 3)$ in the first generation of the graphical examples of data structures provided in Tamariz et al. (2014). For this reason, the first generation of Tamariz et al.'s data is equivalent to the first round of ours, and so on for subsequent generations. The initial state of our data ($r = 0$) (where all variants are unique) is therefore excluded from the model comparison. Absolute entropy was calculated for both datasets (simulated and experimental) using the information-theoretic notion of entropy (H) (Shannon, 1948). For a detailed explanation of the coding process see Tamariz et al. (2014).[1]

We run 1000 simulations under three different conditions: early (E), mid (M) and late (L) connectivity, and test them against the experimental data coded (T). Our model with early connectivity exactly mirrors the pair composition and connectivity dynamic used in Fay et al. (2010). Thus, we predict to find the best fit model for the 64 data structures in the early connectivity. According with Tamariz et al. (2014), most data structures (95%) were best accounted for without $\beta = 0$, and using a wide combination of content-biased models. Thus, using this model fit as a reference for our analysis, we also predict that a distribution of entropy yielded by models including a range of content bias parameters will fit the experimental data set better than models using extreme configurations of biases. Our simulated data structure includes 1452 different parameter value combinations in each round. In this analysis we choose to evaluate the predictive power of our model applied to Fay et al. (2010) by testing, against the experimental data, simulations using (A) the best-fit parameter combinations in Tamariz et al. (2014) (without $\beta = 0$), (B) all parameter combinations, (C)

---

[1] For reproducibility reasons, the experimental entropy data set used for the present analysis can be found at: https://github.com/jsegoviamartin/network_connectivity_dynamics_model.

a content biased model of $\beta = 1$ and, (D) drift model. As a benchmark for illustration, a graphical representation of these model tests against the experimental data (T) can be found in Figure 2.7. Table 2.2 on the following page summarizes model abbreviations.

EA model was associated with better model fit, in particular from round 2 onwards, when the three different connectivity dynamics begin to diverge. Mean distance between *EA* simulated data and experimental data was $M = 0.056, SD = 0.14$. Figures 2.8 and 2.9 show that (i) our model reproduces the experimental patterns of convergence, and (ii) the version of the model with parameter combination *EA* that most closely matches the assumptions of the experiment is the best fit model of the versions examined.

These results suggest that *EA* model qualitatively reproduces the empirical data-pattern for the acceptable range of accuracy under the set of experimental observations. As predicted, the model with early connectivity, which mirrors the connectivity dynamic and pair composition used in Fay et al. (2010), outperforms its alternatives. Additionally, our results are consistent with Tamariz et al. (2014), who found that 95% of data structures were best accounted for within a range of content biases. We show that as long as we add $\beta = 0$ to the models (see models B in Figure 2.8) they lose predictive power, because they underestimate the decrease of entropy in the experiment. Similarly, but due to an overestimation of entropy decline, models with only $\beta = 1$ (models C in Figure 2.8) are far from the reality shown by Fay et al. (2010). Thus, we think that there is substantial evidence to conclude that content bias (in a wide range of levels) was an important driver of the spread of the culturally transmitted variants in the experiment. Our results are also consistent with a variety of studies about the role of content biases. In particular, two studies (Fay & Ellison, 2013; Fay et al., 2008) have used the same corpus of data collected by Fay et al. (2010) to examine the transmission of the communication systems that we model in this chapter. These studies found functional adaptations of the selected variants that are exactly the same that would be predicted if a critical level of content bias were operating on the communication system (Tamariz et al., 2014). We know that content biases have a strong impact on the adoption of variants. Crucially, for the purpose of the present study, connectivity dynamics, amplified by content biases, may have important consequences for the evolution of cultural variants in populations. Failing to take into account the role of connectivity dynamics in experiments and in real life may preclude a full understanding of the data observed.

	Simulated best-fit (without $\beta = 0$)	Simulated (all parameter combinations)	Simulated ($\beta = 1$)	Simulated (Drift model $\beta = 1$)	Experimental data
Early connectivity	EA	EB	EC	ED	T
Mid connectivity	MA	MB	MC	MD	–
Late connectivity	LA	LB	LC	LD	–

Table 2.2: Model abbreviations

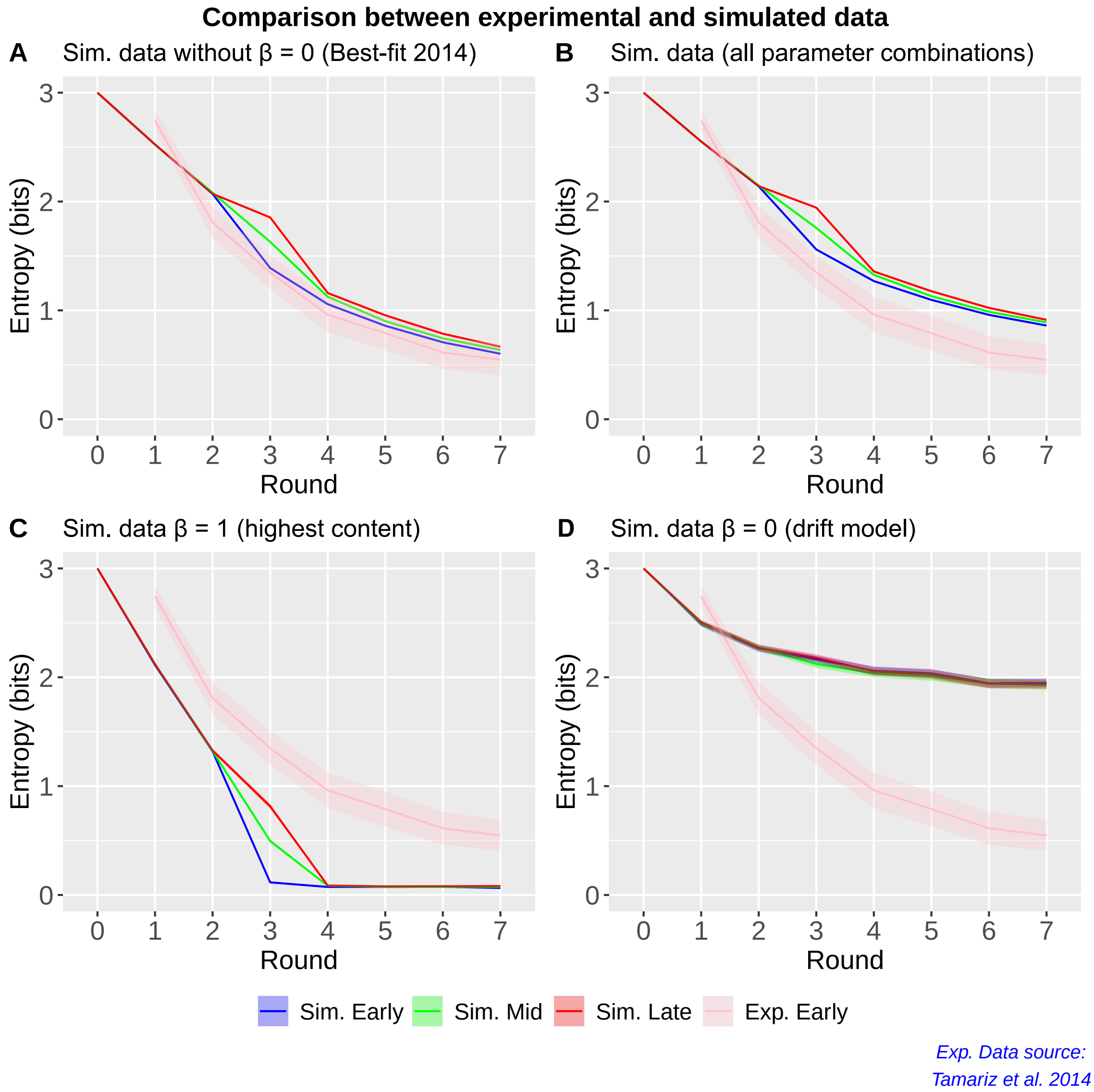


Figure 2.8: Comparison between experimental and simulated data. Simulated data from 16 models (4 parameter combinations with 3 connectivity dynamics). Entropy (H) (y-axis) is plotted against time in rounds (x-axis). Simulated data using early connectivity (E) (blue), mid connectivity (M) (green) and late connectivity (L) (red) is tested against the experimental data collected by Fay et al. (2010) and coded by Tamariz et al. (2014) (pink). The line plots here show the evolution of entropy, with increasing convergence, over rounds. The simulated early connectivity (EA) model was associated with better model fit, in particular from round 2 onwards, when the three different connectivity dynamics begin to diverge. Models with $\beta = 0$ (B) lose predictive power when compared to A, because they underestimate the decrease of entropy in the experimental data. Models with only $\beta = 1$ (C) underestimate the decrease of entropy. Lines indicate mean entropy and ribbons indicate 95% CIs.

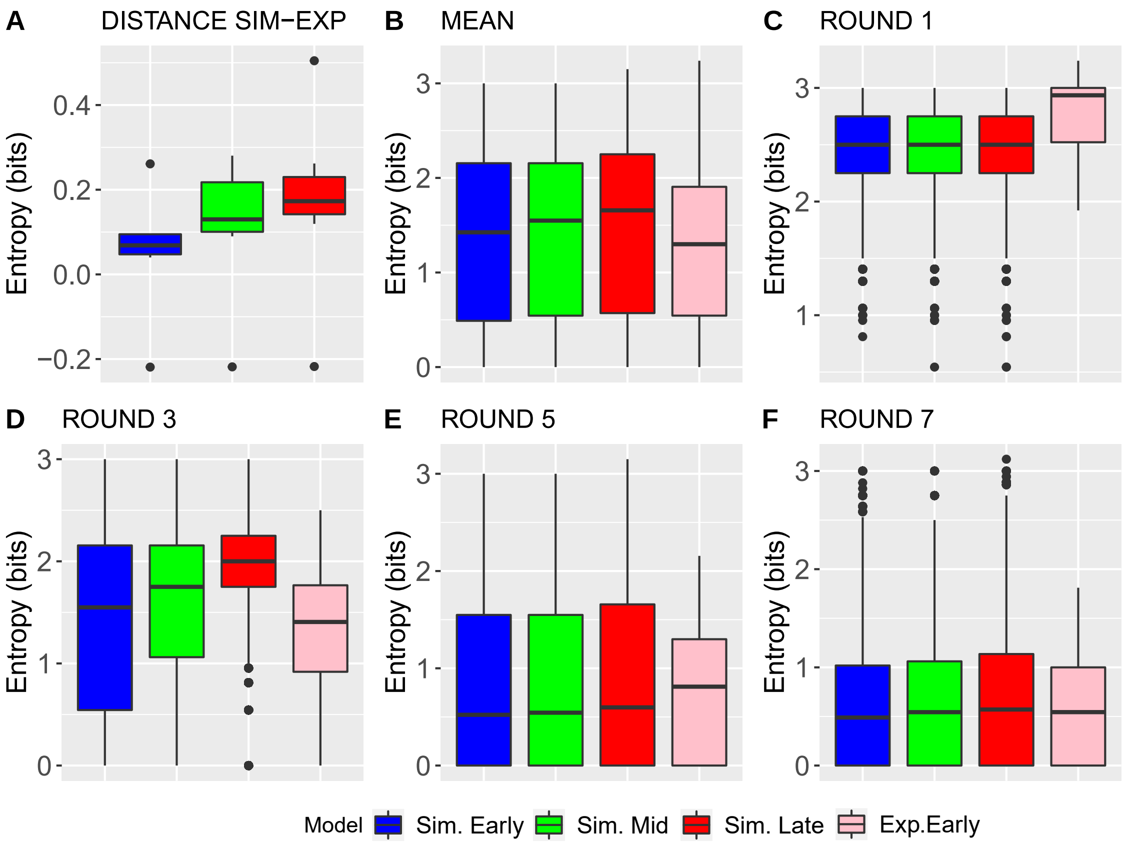


Figure 2.9: Mean distance between simulated data and experimental data. Mean distance between simulated early connectivity (EA, in blue) and experimental data (T, in pink) was $M = 0.056$, $SD = 0.14$. B. Mean entropy by model. C, D, E and F. Mean entropy by model by round. Error bars indicate 95% CIs.

2.5 Discussion

Using a computational approach, the present study extends formal and experimental findings about learning in social networks by simulating pairwise interactive micro-societies based on round-robin tournaments, where individual cognitive biases, memory constraints and population connectivity dynamics are systematically manipulated. Our results extend previous studies on social learning showing that content biases are important drivers of convergence. We also show that connectivity dynamics affect the time-course of the spread of variants in moderate-high content biased populations: When populations take longer to reach full connectivity, convergence onto a single cultural variant is slowed. Content bias accelerates convergence and amplifies the effects of connectivity dynamics. Larger memory size and coordination bias, especially egocentric bias, are also shown to slow down convergence, especially in moderate-low content biased populations. Finally, connectivity dynamics are shown to affect the frequency of high quality variants (adaptiveness), with late connectivity populations showing bursts of rapid change in adaptiveness followed by periods of relatively slower change, and early connectivity populations following a single-peak evolutionary dynamic.

While content bias is the main predictor of convergence, in some circumstances its effect can be modulated by the population connectivity dynamic (Figure 2.5 and Figure 2.6). This means that adding connectivity dynamics may improve the predictive power of models based on cognitive biases and social networks, especially in cases of strong content biases. Population convergence on shared cultural conventions is driven by the agents' content bias, and the time required to reach a certain degree of convergence (or time to convergence) can be deeply affected by the specific order of interactions between agents, that is, by the population connectivity dynamic: In general, the less connectivity the more time is needed to converge. Furthermore, the effects of these different dynamics in the order of interactions of the agents can be observed even if we maintain the same network topology, in our case a fully connected network.

It is important to note that even though previous work in the field has frequently used fully connected networks (Fay et al., 2010; Komarova, Niyogi, & Nowak, 2001; Tamariz et al., 2014), this type of network topology is unrealistic because it restricts the interaction between agents to a particular pattern of interconnectedness, reducing the complexity of the system. Therefore, population connectivity dynamics might play a different role in other networks, such as scale-free networks (Barabási, Albert, & Jeong, 1999; Barabâsi et al., 2002). Furthermore, it is well established that people do not contribute equally to group discussions, leading to different degrees of network connectivity (Fay, Garrod, & Carletta, 2000; Stasser & Taylor, 1991). Both factors might potentially motivate future extensions of our current investigation.

Our results also identify a general tendency for adaptiveness to change over time and for cultural variants to converge on high quality variants, in such a way that it is possible to identify causal links between connectivity dynamics and evolutionary trajectories. In this way, in late connectivity dynamics several punctuational bursts occur in the course of a complete cycle of interactions between agents. In contrast, early connectivity dynamics follow a single-peak evolutionary trajectory. These computational results extend a number of studies that, under a variety of assumptions, have proposed punctuational or rapid bursts of change as a feature of cultural and language evolution (Atkinson, Meade, Venditti, Greenhill, & Pagel, 2008; Dixon & Robert Malcolm Ward, 1997; Fitch, 2008; Janda & Joseph, 2003; Sabherwal, Hirschheim, & Goles, 2001). Punctuational changes in our model may provide insight into processes underlying the human ability to adapt quickly to cultural variants introduced by new agents (e.g. due to migration), showing that these changes can be induced merely by manipulating the order of interactions in a population.

In high content-biased populations, the effect of the connectivity dynamic is amplified (Figure 2.5 and Figure 2.6), while coordination bias and memory size effects are masked (Figure 2.3 and Figure 2.4). Interestingly, in low content-biased populations the effect of these parameters became visible: when compared with a drift model, egocentric and allocentric biases both reduce convergence (Figure 2.3). This is because cultural diversity can more easily be maintained over time in the presence of behaviours that maximise the occurrence of either self-produced signals—in the case of egocentric agents—or partner-produced signals—in the case of allocentric agents. The effect of egocentric bias is stronger than that of allocentric bias. This is due to the fact that fully egocentric agents stick to their own variant, which is always the same in the egocentric memory, unless there is an innovation. At the population level, this means that each agent has a different variant, returning maximum entropy—which can only decrease through mutation. Fully allocentric agents, in contrast, always adopt variants produced by others. High variation is thus maintained, but to a lesser extent than in the egocentric case because allocentric agents choose variants from among all the variants stored in their allocentric memories—variants produced by current or previous partners. This sometimes leads to more than two agents converging on the adoption of a variant, and therefore reducing entropy.

Memory also shows its effect more markedly when content bias is low or intermediate (Figure 2.4). The longer the memory span, the more variation is maintained, as variants from earlier rounds, that might not appear at one round, are kept in memory and may reappear. Our study is consistent with previous literature showing that memory limitations lead to a reduction in variation (Ferdinand et al., 2013; Tamariz & Kirby, 2015). Nevertheless, we show that this reduction could be masked in high content-biased populations, when agents have a strong preference for signals with high intrinsic value.

Our results also agree with recent studies showing that population structure and population interaction can be strong predictors of cultural evolution (Creanza et al., 2017; Derex & Boyd, 2016; Derex et al., 2018). In addition, our model shows that cognitive biases and population connectivity dynamics may interact in important ways. When content-biased populations evolve in high isolation (late connectivity dynamic), convergence is slower than when they evolve in moderate (mid connectivity dynamic) or low isolation (early connectivity dynamic). This is because more isolated subpopulations cannot benefit from wider cultural exchange during the first rounds, those in which agents are acquiring the basis of their culture and storing it into their respective memories. This suggests that population structure and, in particular, the connectivity dynamics of the population, can have important effects on cultural convergence and should be taken into account when it comes to research on the interactions between cognitive biases, network structures and cultural evolution.

2.6 Implications

Our results are relevant to social learning scenarios governed by turn-based interventions in which convergence on shared conventions is crucial (e.g. Bloom et al., 1987; Fay et al., 2008, 2010; Manzo, 1996; Shteingart et al., 2013; Stone, 1969). Our computational model can be used to fit real data obtained from turn-based cultural processes and might be helpful to improve the organisation of the turn taking by mitigating undesirable effects linked with one particular connectivity.

In light of evolutionary theory applied to cultural evolution (Atkinson et al., 2008; Fitch, 2008), our results also suggest that, in some scenarios, bursts of change in the cultural system may be partly explained by the order in which individuals interact over time. This demonstrates, for the first time, a direct connection between convergence, adaptiveness and population connectivity dynamics for a fixed range of combinations of individual cognitive biases.

Our findings are consistent with scientific models and with theoretical and experimental studies of human communication showing that convergence is driven by content biases (Gong et al., 2008; Tamariz et al., 2014), and also agree with studies on rational learning in social networks showing that the level of convergence is partially determined by the degree of connectivity in the social network (Barkoczi & Galesic, 2016; Centola & Baronchelli, 2015; Centola et al., 2007; Lazer & Friedman, 2007; Mueller-Frank, 2013; Olfati-Saber & Murray, 2004). The specific manipulation of the network connectivity dynamic that we have studied here has not been taken into account in previous experimental work and it would be especially relevant to those researchers that use dynamic interactive microsocieties of agents switching partners over time (e.g. communicative games, cooperative games or tournaments; Baum et al., 2004; Byun et al., 2018; Caldwell & Smith, 2012; Fay et al., 2008, 2010; Mesoudi & Whiten, 2008; Raviv et al., 2019b; Tamariz et al., 2014). In most cases, experimental designs of microsocieties of interacting actors only include one pair composition out of all the possible combinations of pair shuffling, and therefore, outcomes are related with only one specific population connectivity dynamic, potentially affecting the accuracy of the generalisations made by these studies. Our results suggest that this type of research would benefit from experimental designs that control the probability of occurrence of each possible connectivity dynamic.

Our agent-based model is a simplification of a specific problem. Each agent is characterized by a combination of biases towards the quality and origin of a set of variants. The network topology is complete and organised in dyads. Thus, following Rubinstein (2006), in the dilemma of responding to reality, we regard our model as a very limited set of assumptions which is inevitably inapplicable to many contexts.

In this chapter we have presented an innovative small-scale simulation project and a comparison of the simulations against experimental data collected from existing experiments, which allowed us to run a high resolution test of the specific parameters examined within a small but highly controlled dataset. However, given that previous experiments only considered one specific connectivity dynamic (i.e. early connectivity), our attempt to provide a comprehensive evaluation of the predictive power of our model is still missing a complete dataset with at least two connectivity dynamics to compare with. In Chapter 3, with the aim of verifying the assumptions and conclusions of the present study, we will conduct an experiment in the lab to collect data from early and late connectivity dynamics.

To conclude the present chapter, properties of populations can be important predictors of cultural evolution, and our model has shown that convergence can be altered by the connectivity dynamic. This may help improve the experimental design of ongoing research in the field of cultural evolution and better explain the interactions between network topologies, cognitive biases and cultural transmission.

Chapter 3

Testing early and late connectivity dynamics in the lab

Abstract

How does the order of individuals' interactions affect the emergence of shared conventions at the population level? The answer to this question is relevant for a number of fields, such as cultural evolution, linguistics, cognitive science or behavioral economics. In this chapter we investigate experimentally how two different network connectivity dynamics affect the evolution of the diversity of cultural variants of the communication system. We report an experiment in the lab in which participants engage in a Pictionary-like graphical communication task as members of a 4-participant micro-society, interacting in pairs with the other three members of the community across 4 rounds. The experiment has two main goals: First, to evaluate the effect of two network connectivity dynamics (early and late) on the evolution of the convergence of micro-societies on shared communicative conventions under controlled conditions. Second, to compare the predictions of the agent-based model described in the previous chapter against experimental data, and calibrate the model to find the best-fitting parameter setting. Our experimental data shows that, as predicted by the model, an early connectivity dynamic increases convergence and a late connectivity dynamic slows down convergence. We found significant differences between conditions in round 3 and round 4. We estimate the best-fit parameter combination for the 96 data structures coded. Medium to high content bias, neutral to egocentric coordination bias and memory size of 3 rounds was associated with a better model fit. In the light of the model evaluation and the experiment outcome, we discuss the impact of our predictions on social influence research and possible factors that might help to improve model precision.

3.1 Introduction

Cultural conventions shape most of the aspects of our social life. The emergence of these patterns of collective behaviour has been investigated in different fields (Christiansen & Kirby, 2003; Garrod & Doherty, 1994; Lass, 1997; Lewis, 2008; Young, 1993). However, more research is still needed to clarify the extent to which subtle variations in the patterns of interaction between individuals combined with biased learning give rise to the emergence of these social conventions (Barkoczi & Galesic, 2016; Boyd & Richerson, 1988; Derex & Boyd, 2016; Mason & Watts, 2012). For example, simple changes in the network structure of human communities can drive the dynamics of norm formation (Centola & Baronchelli, 2015), health behaviour (Centola, 2010), group estimates (Becker et al., 2017), beliefs and polarisation (Turner & Smaldino, 2018). Using graphical communication tasks, researchers have also studied the emergence of communicative conventions in dyadic and small communities (Branigan, Pickering, McLean, & Cleland, 2007; Fay et al., 2010; Galantucci, 2005; Garrod & Anderson, 1987; Garrod & Doherty, 1994; Selten & Warglien, 2007). These prior experimental studies have nonetheless been limited to studying a limited number of pair compositions, which made difficult to evaluate the extent to which a simple change in the order in which connections between individuals unfold over time (so-called *network connectivity dynamic*) can lead to different rates and dynamics of convergence. As a result, the quantification of the state of specific features of individual cognition (e.g. cognitive biases and memory) and how they are affected by the network connectivity dynamic over time remains restricted to a limited number of experimental observations (Tamariz, Ellison, Barr, & Fay, 2014).

In this chapter, we investigate how a simple manipulation of the order in which connections between individuals unfold over time (network connectivity dynamic) affect the cultural diversity of graphical variants in a communication system. In particular, how two different network connectivity dynamics (early and late) affect the convergence of micro-societies on shared communicative conventions. First, we review the experimental and computational work analysing cultural and communicative accounts of convergence and network structures. Tereafter, we describe a graphical experimental design that permits a direct test of our theoretical hypothesis. Next, we use an agent-based model (ABM) that mirrors the logic behind the experiment to simulate convergence over time. Finally, we report the results of the experiment and calibrate our model to find the best-fit parameter combination for the data structures coded.

3.1.1 Experiments

Theories of collective behavior suggest that diversity can be affected by social influence (Page, 2008). For example, experiments in the lab have shown that convergence on shared interpretations increased when individuals observed other's beliefs, which resulted in a fall in diversity (Lorenz, Rauhut, Schweitzer, & Helbing, 2011). Diversity of shared conventions can also be affected by the structure of the communication network. For example, an experiment where participants were asked to estimate numerical quantities in a number of visual stimuli (e.g. caloric content on a plate of food, or number of candies in a container) showed that social influence in decentralised networks (i.e. where all individuals have the same number of ties) reduced diversity of estimates and improved individual accuracy when compared to centralised networks (i.e. where some individuals have more ties) (Becker et al., 2017). Interestingly, the emergence of spontaneous globally shared social conventions from local interactions can be caused by simple changes in the network connectivity. Centola and Baronchelli (2015) show that an increase in network connectivity can accelerate convergence on shared conventions. Crucially, these global conventions can emerge even though participants do not have access to global information.

When it comes to explaining the evolution of communicative conventions, both iterated learning and collaborative accounts play important and distinguishable roles (Fay et al., 2010). Iterated learning accounts predict that prior biases combined with vertical and unidirectional transmission (i.e. parent to offspring) lead to the emergence of regularities in the communicative system that reflect the individuals' learning biases. This effect has been shown in a variety of experiments and linguistic levels (e.g. Kalish, Griffiths, & Lewandowsky, 2007; Kirby, Cornish, & Smith, 2008; Kirby, Tamariz, Cornish, & Smith, 2015). By contrast, collaborative accounts assume that communicative conventions are an emergent property of horizontal and bidirectional transmission (i.e. the agents' local and vertical interactions). This interactive alignment has also been examined in a variety of experiments and linguistic levels (e.g. Branigan, Pickering, McLean, & Cleland, 2007; Garrod & Doherty, 1994; Raviv, Meyer, & Lev-Ari, 2019a; Steels, 2006).

Fay et al. (2010) showed in an experiment that different micro-societies converged on different communicative conventions. However, they only used one possible network connectivity dynamic in their experimental communities—one possible order in which connections between individuals unfold over time. Since a different network connectivity dynamic can give rise to pockets of isolated pairs that do not share information with the rest of the community during a number of rounds, we hypothesised in a previous study that an increase in the network connectivity dynamic might accelerate the rate of convergence on communicative conventions, especially under moderate and high content biased populations (see Chapter 2). Furthermore, by contrasting the specific network connectivity conditions that we propose, we can evaluate whether convergence exceeds the maximum convergence that horizontal alignment alone can explain.

3.1.2 Formal and computational models

The evolution of cultural diversity has been typically studied through formal and computer modelling across different disciplinary traditions. In the 1950's and the 1960's researchers developed formal models of social influence—the agents' tendency to adopt cultural variants from influential agents—that showed that there is a tendency towards convergence on a shared cultural convention in fully connected networks (e.g. French, 1956; Harary, 1959). Later on, a number of models that took into account geographic isolation (Schelling, 1978), polarisation (Latané, Nowak, & Liu, 1994; Nowak, Szamrej, & Latané, 1990) and the agents' capacity to retain original cultural variants (e.g. Friedkin & Johnsen, 1990; Marsden & Friedkin, 1993), showed that full convergence is not always the expected outcome. In 1997, Axelrod proposed an innovative agent-based model that included homophily—the agents' tendency to interact more frequently with similar agents. Axelrod's results showed that the combination of influence and homophily produces changes in the network connectivity that may favour local convergence and in turn a global diverse equilibrium. These models use a broad definition of culture where cultural variants can be thought of as a proxy for opinions, beliefs, artistic tastes (music, painting) or communicative variants.

In these early models, however, the network of social interactions remained fixed, and the agents' cognitive evolution was limited to a number of evolving traits. More recently, researchers on social influence have addressed many of these issues by modelling new interactions between the agents' imitation rules and social dynamic networks: For example, by analysing network efficiency and conformity biases (Barkoczi & Galesic, 2016), network connectivity and group performance (Lazer & Friedman, 2007) and the co-evolution of social networks and the agents' ties in the presence of cultural drift (Centola et al., 2007). These models have successfully shown the effect of a number of interactions between specific features of the network topology and the agent's cognition.

A different and more specific tradition of agent-based computer simulations has been used to investigate the evolution of convergence of variants within a communication system (Gong et al., 2012; Tamariz et al., 2014). These models use a combination of previously identified selective pressures to break down communicative influence into several parameters, such as content bias, coordination bias, prestige bias or variant value. Broadly speaking, these models show that in order to understand the spread of variants in a micro-society, it is necessary to take into account the interaction between different, sometimes antagonist, cognitive biases and the social conditions underpinning their relative importance for explaining diversity. For example, if variants show heterogeneity of content values, transmission error can delay diffusion and help preserve cultural diversity (Gong et al., 2012). In another study, a model that was used to explain experimental data showed that content bias consistently increased convergence, and interacted with coordination (Tamariz et al., 2014).

In the previous chapter, we combined these two ABM's traditions to construct a micro-scale model of dissemination of culture that took into account network connectivity dynamics, content and coordination biases, memory size and innovation rates. Our goal was to construct a comprehensive ABM to simulate convergence over time for all the possible value-parameter combinations included in the model. Using this ABM we examined the effect of the order in which connections between individuals unfold over time (network connectivity dynamic), and how it interacts with the rest of the individual parameters. Three different network connectivity dynamics (early, mid, late) were examined according to how many agents could potentially share the same variant system at a given time. We showed that connectivity dynamics can affect the time-course of variant spread, with lower connectivity slowing down convergence of the population onto a single cultural variant. Furthermore, we also showed that, compared to a neutral evolutionary model, content bias accelerated convergence and amplified the effect of the connectivity dynamic, whilst larger memory size and coordination bias, especially egocentric bias, slowed down convergence.

Although we contrasted our model against the experimental data collected by Fay et al. (2010), these data corresponded only to one of the modelled network connectivity dynamics (i.e. early connectivity). Thus, little has been said so far about the real effect of the connectivity dynamic on the evolution of cultural diversity, and on how valid our model predictions are to fit an experimental case in which two different conditions of connectivity are examined. Our model predicts that, when compared to an early connectivity, late connectivity delays global convergence in small-scale micro-societies

3.1.3 Purpose of the study

In the present chapter we report an experiment in which micro-societies of 4 participants play a Pictionary-like game in pairs. Participants were allocated to one of two conditions: early connectivity dynamic and late connectivity dynamic.

We also constructed a version of our ABM mirroring the network connectivity dynamic of the experiment, and used it to find the parameter combination associated with better model fit.

The present chapter has two main goals: First, to test directly, under controlled experimental conditions, the hypothesis that a late connectivity dynamic delays the convergence of participants on shared variants when compared to an early connectivity dynamic. We will also evaluate the rate of convergence on shared communicative conventions before agents have access to complete information in the late connectivity condition. This will tell us to what an extent convergence on shared conventions occurs without horizontal transmission. Second, to test the robustness of the predictions of our model by running simulations that closely mirror the experiment and finding the best-fit parameter combination. In this manner, we propose the most likely combination of parameters that might have been operating during the experiment.

3.2 Experiment

3.2.1 Task and procedure

The experimental task is based on the graphical communication task developed by Fay et al. (2010). In our task, participants play a Pictionary-like game in pairs. In each pair, one participant acted as the director and its partner as the matcher. In each trial, the director's task was to communicate to its partner a specific concept from an ordered list of 20 concepts (16 target items and 4 distracters). The matcher's task was to identify the referent of each drawing from an unordered list containing the same 20 concepts. We used the same concepts as Fay et al. (2010). The complete list of concepts can be found in Table 3.1.

Places	People	Entertainment	Objects	Abstract
Art Gallery	Arnold Schwarzenegger	Cartoon	Computer Monitor	Homesick
Parliament	Brad Pitt	Drama	Microwave	Loud
Museum	Hugh Grant	Sci-Fi	Refrigerator	Poverty
Theatre	Russell Crowe	Soap Opera	Television	Sadness

Table 3.1: Set of concepts used in the experiment.

Participants played 4 rounds. In each round, they played the game 6 times. Each game consisted of 20 trials—the same number of trials as concepts. Throughout the 20 trials, partners alternated roles every 10 concepts. Each participant acted as the director on half of the trials and as the matcher on the other half. Participants used the same concepts on each game. The order of the list of concepts was randomised on each game.

The task was conducted on computer terminals using the open-source virtual whiteboard tool GUESS-A-SKETCH (Lonely Star Software, 2017). Each participant sat at a computer terminal in an independent booth, so they could not identify their partners. Drawing input was made via a standard mouse. Concepts to be drawn by the director and to be guessed by the matcher were presented on separate sheets. Following Fay et al. (2010), director drawing was restricted to black ink, while the matchers were allowed to ask for further information by circling a part of the drawer's illustration using green ink. Participants were also allowed to erase parts of their drawings during the course of each trial by using an erase button. All the drawing activity was displayed simultaneously on the shared whiteboards. When the matchers considered they had identified the director's intended referent, they typed 'Got it' in an integrated board chat and made their selection. At this moment, directors initiated the next trial by clicking a clear board button and typing the number of the trial to be played. The integrated chat was only used by the director to communicate trial number and by the matcher to communicate 'Got it'. Director and matcher roles were switched each 10 concepts. During the experiment, participants received no feedback on their communication success. Whiteboard activity was recorded using HYPERCAM 4.0 by Hyperionics and Solveig Multimedia.

3.2.2 Participants and design

Forty undergraduate students participated in exchange for payment. Participants were recruited via MYCAREERHUB at the University of Edinburgh. Ethical approval was granted by the Research Ethics Committee of the University of Edinburgh.

Participants were randomly allocated to one of two conditions:

(i) early connectivity dynamic; and

(ii) late connectivity dynamic.

Participants were organized into groups of 4 participants with testing lasting approximately 2 h and 30 min.

Participants played six consecutive games with their partner in each round. Then, participants switched partners and played a further six games with the new partner. In the two conditions, each participant interacted once with two other members of the micro-society and twice with their first partner. In the early connectivity dynamic, each participant re-encounters its first partner in the last round, while in the late connectivity dynamic, they re-encounter their first partner in the second round. The pair composition in each condition can be found in Table 3.2.

Round	Pair composition (early connectivity)	Pair composition (late connectivity)
1	1&2 3&4	1&2 3&4
2	1&4 2&3	1&2 3&4
3	1&3 2&4	1&4 2&3
4	1&2 3&4	1&3 2&4

Table 3.2: Pair composition in each laboratory condition. Within each 'Round' participants played six consecutive games of the communicative graphical task with their partner.

In the early connectivity dynamic, participants could potentially share the same communicative variant system by round 3, while in the late connectivity dynamic participants could potentially share the same variant system one round later. For example, in the early connectivity condition, if participant 2 adopts the variant system of participant 1 during the first round, and participant 3 is influenced by participant 2 during round 2, then participant 3 and participant 1 will share the same variant system at some moment during round 2, despite never having interacted in the past. If we also assume that participant 4 adopts the variant system of participant 1 in round 2, then everyone could share the same variant system by round 3. Thus, in the early connectivity dynamic, rounds 1–2 constitute prealignment partner interactions, whereas rounds 3–4 are postalignment partner interactions. In contrast, in the late connectivity dynamic, rounds 1–3 are prealignment partner interactions and the remaining one (round 4) the only postalignment partner interaction. We contrasted variant diversity in the early connectivity condition against the late connectivity condition.

3.2.3 Video coding

Among the 10 experimental micro-societies, 2 micro-societies experienced connection problems during the task, and in 2 other groups one participant did not consistently follow the order of the concepts to be drawn. These groups were excluded by necessity, leaving us with 6 distinct 4-person groups. The data to be evaluated are structured by micro-society—6 in total: 3 groups in the early connectivity condition and 3 groups in the late connectivity condition—and by concept—16 target concepts, totalling 96 data structures. Each data structure includes 16 drawings, one drawing per participant in each of the four rounds.[1]

Because participants used a variety of distinct variants to communicate each meaning, and because different micro-societies used different variants to communicate the same meaning, a unique coding scheme was developed for each data structure. Firstly, it was established the initial state of the variant system, labelling the variants produced in round 1 as unique. Secondly, the spread of each variant type was tracked across the subsequent generations. The coding scheme was specified as follows:

1. Only the relevant traits of each variant (the ones that remained in the drawing across rounds) were traced back.
2. New relevant traits and ambiguities were treated as innovations.
3. If a participant combined or drew two relevant traits in the same drawing, we traced back those traits and labelled the drawing as the variant where the most preserved trait appeared.
4. Analogies (traits that look alike but do not have a common evolutionary origin within the experiment) were labeled as the variant where the trait first appeared.
5. For each round, the first drawing of each participant was considered.

The 96 data structures were coded by myself. A second coder (MT) independently coded 16.6% (one micro-society) of the data structures in order to assess inter-coder reliability as quantified by Cohen's kappa coefficient (Cohen, 1960). Inter-coder reliability was $\kappa = 0.883$, which corresponds to good reliability as reported by Carletta (1996, p. 252). An illustrative data structure can be found in Figure 3.1.

---

[1] Coded data and code to run the analyses described in this paper are available at: https://github.com/jsegoviamartin/network_connectivity_dynamics_model.

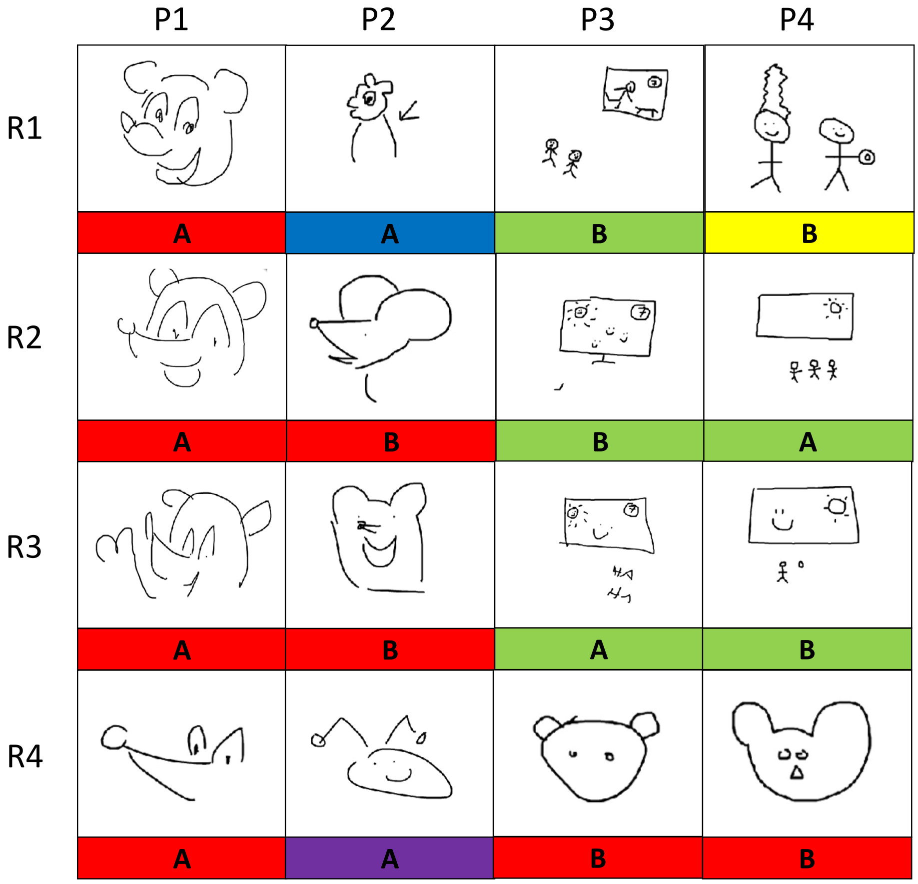


Figure 3.1: Cultural evolution of the signs used to represent CARTOON in a 4-person micro-society. Columns correspond to participants (P1–P4) and rows to rounds (R1–R4). Capital letters (A, B) indicate the different participant pairings in a given generation, and colours indicate the different variant types. When participants played with their first partner (round R1), they used a variety of different signs: At the beginning of round R1, participants tended to use a unique variant. As they interacted with the other members of their micro-society, some traits of the red variant ('animal funny face with big ears and pointed nose') propagates until everyone use a refined version of this sign (either 'big ears', 'pointed nose' or both). An ambiguity arose when tracking the evolutionary origin of the drawing produced by participant 2 in round 4, and it was classified as a new variant (innovation). Note that participants tend to retain their initial variant until they encounter the red variant, and sometimes even after. According to (Tamariz et al., 2014) this suggests a strong content bias for the red variant such that it was more likely to be adopted by participants compared with its competitors in this particular micro-society.

3.3 Quantifying cultural diversity in the communication system

We are interested in the composition of the pool of variant tokens produced by the participants at the beginning of each round, and how it changes over rounds. We examine the level of convergence in this pool. As in Chapter 2, we quantify convergence using the information-theoretic notion of entropy (H) (Shannon, 1948). However, we normalise entropy by $\log_2 n$ to obtain $H_n(V) \in [0,1]$:

$$H_n(V) = -\sum_{v_i \varepsilon V} \frac{p(v_i)\log_2 p(v_i)}{\log_2 n} \tag{3.1}$$

where V corresponds to the set of variants, $p(v_i)$ is the probability of the i^{th} variant in that set, and n is the number of variants. Entropy is a well-established alpha diversity index that has been used to measure cultural diversity. High entropy corresponds to low convergence. In the following analyses we show mean values and standard deviations (Mean $\pm$ SD).

3.4 Mirror model predictions

We constructed a version of the model described in Chapter 2 mirroring the network structure of the experiment. We simulated 4-agent micro-societies with the same network connectivity dynamics as those described in section Section 3.2.2: Early connectivity and late connectivity. For our measurements, we averaged across 1000 runs for each condition.

The best-fitting parameter values for the data collected by Fay et al. (2010) and coded by Tamariz et al. (2014) included a range of content biases (0.1 to 1) (Segovia-Martín et al., 2019). Here, we ran the model with the same parameter combination to predict the evolution of the cultural diversity of the communication system in 4-person micro-societies. Model prediction using this parameter combination can be found in Figure 3.2. It is possible that this prediction does not accurately describe the observed convergence of the experimental micro-societies on shared communicative conventions, but it might provide a general forecast under a reasonable set of assumptions of connectivity and content bias. Simulated late connectivity decreased convergence when compared to simulated early connectivity. Our predictions show a maximum difference in entropy between conditions at the beginning of round 3—mean normalised entropy 0.35 in the simulated early connectivity condition and 0.65 in the simulated late connectivity condition. This difference between conditions is predicted to be reduced in round 4—mean normalized entropy 0.25 in the simulated early connectivity condition and 0.34 in the simulated late connectivity condition. Since the first structural difference between both connectivity dynamic conditions occurs in round 2 (see Table 3.2 on page 61), the model does not predict differences in entropy levels either at the beginning of round 1 or in the first production of variants of round 2.

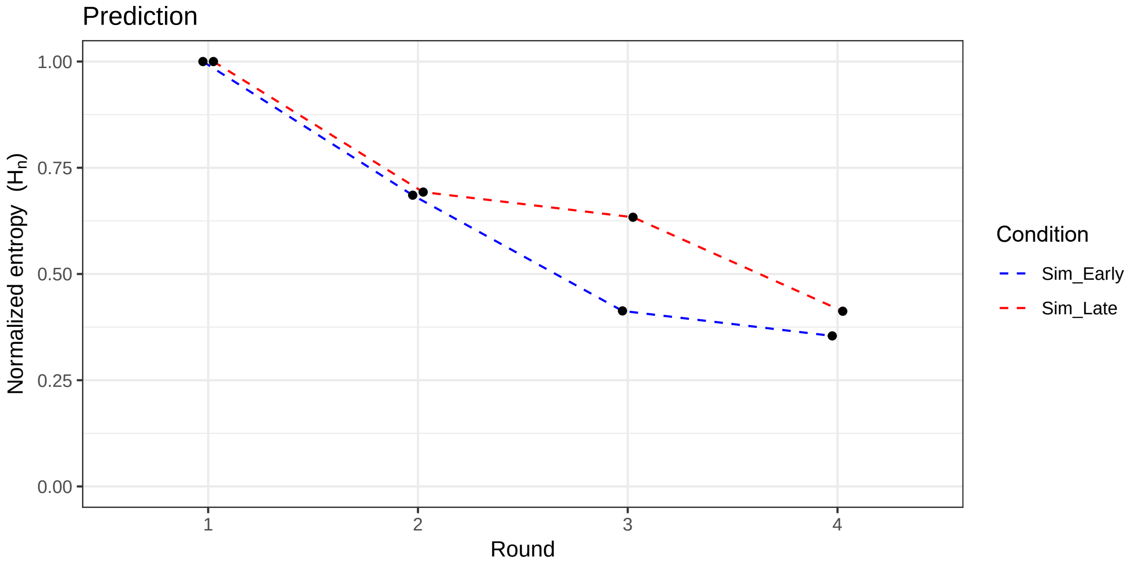


Figure 3.2: Predicted entropy (H_n) averaged over each level of connectivity and round. Parameter combination used: Content bias range $= [0.1, 1]$, coordination bias $= 0.5$, memory size $= 3$, innovation rate $= 0.02$.

3.5 Experimental results

The cultural diversity levels of the communication system at each round show that micro-societies in the experimental late connectivity condition indeed decreased convergence, and micro-societies in the experimental early connectivity condition increased convergence, both in round 3—mean normalized entropy 0.35 in the experimental early connectivity condition and 0.45 in the experimental late connectivity condition—and round 4—mean normalized entropy 0.24 in the experimental early connectivity condition and 0.35 in the experimental late connectivity condition; see Figure 3.3.

We ran a mixed-effects linear model predicting entropy levels with fixed effects of condition—early connectivity dynamic or late connectivity dy-

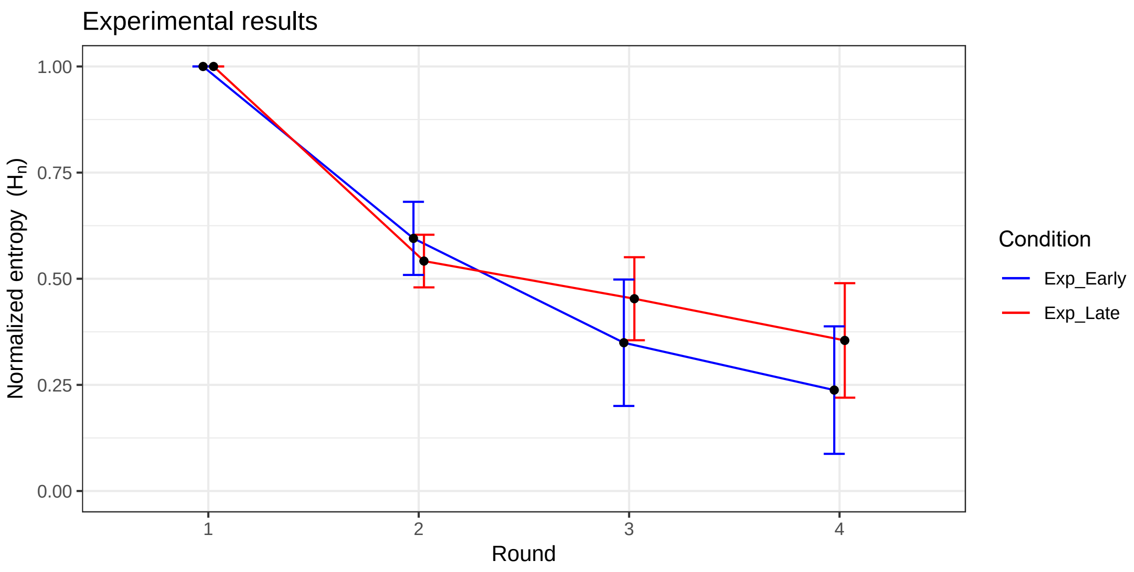


Figure 3.3: Observed entropy (H_n) averaged over each level of connectivity and round. Error bars indicate 95% CIs.

namic—and round (1 to 4), with random intercepts per concept and micro-society. Condition was sum-coded and Round was transformed so that the model intercept corresponds to mean entropy across conditions and Round 1. Likelihood ratio testing confirms that our model performs significantly better than the null model with random effects only ($\chi^2(2) = 396.81, p < 0.001$). Normalized entropy levels were higher by 0.02 (s.e. 0.010) in the late connectivity condition relative to the model intercept of 0.91 (s.e. 0.028). Comparisons to reduced models with each fixed effect removed confirm a main effect of condition ($\chi^2(1) = 4.766, p < 0.029$) and round ($\chi^2(1) = 395.19, p < 0.001$).

In order to assess the differences in cultural diversity of the communicative system from the moment the two connectivity conditions diverge, we ran mixed-effects regressions on the round 3 and round 4 data only. Using round 3 data only, we ran a mixed effects linear model predicting entropy levels with fixed effects of condition and random intercepts per concept and micro-society. Likelihood ratio testing confirms that our model performs significantly better than the null model with random effects only ($\chi^2(1) = 7.7397, p = 0.005$). Normalized entropy levels were higher by 0.052 (s.e. 0.018) in the late connectivity condition relative to the model intercept of 0.401 (s.e. 0.038). Using round 4 data only, we also ran a mixed effects linear model predicting entropy levels with fixed effects of condition and random intercepts per concept and micro-society. Likelihood ratio testing confirms that our model performs significantly better than the null model with random effects only ($\chi^2(1) = 6.8081, p = 0.009$). Normalized entropy levels were higher by 0.058 (s.e. 0.022) in the late connectivity condition relative to the model intercept of 0.296 (s.e. 0.044).

3.5.1 Model evaluation

We used the experimental 96 data structures coded for the present study to obtain best-fitting parameter values in our simulations. A total of 968 parameter combinations were evaluated—11 levels of content bias, 11 levels of coordination bias, 2 levels of memory, 2 levels of connectivity and 2 levels of value across variants; see Table 3.3.

Parameter	Number of levels	Values
Content bias	11	0 to 1 in steps of 0.1
Coordination bias	11	0 to 1 in steps of 0.1
Memory	2	1 & 3
Variant value distribution	2	$[1,0,0,1]$ & $[1,0,0,0]$
Connectivity	2	Early & Late
Innovation rate	1	0.02

Table 3.3: Parameters examined.

Bayesian factors have become increasingly common for model selection (e.g. Baele, Li, Drummond, Suchard, & Lemey, 2012; Suchard, Weiss, & Sinsheimer, 2001; Tamariz et al., 2014; Xie, Lewis, Fan, Kuo, & Chen, 2010). In the present study, the strength of evidence for a particular parameter combination was evaluated by estimating a Bayes factor using Bayesian Information Criteria (Wagenmakers, 2007): The likelihood ratio of the marginal likelihood of two competing hypothesis. The likelihood of the data was considered under both the null hypothesis—mean normalised entropy in the selected parameter combination and in the experimental condition are equal—and the alternative hypothesis—mean normalised entropy in the selected parameter combination and in the experimental condition are different. One of the benefits of this Bayesian approach is that it quantifies support for both the null (labelled BF_{01}) and the alternative hypothesis (labelled BF_{10}), thereby providing a clearer estimate of the strength of evidence in the data for model selection (Jarosz & Wiley, 2014). This contrast can be expressed as follows:

$$BF_{01} = \frac{\text{likelihood of data given } H_0}{\text{likelihood of data given } H_1} \tag{3.2}$$

where $BF_{01} = \frac{1}{BF_{10}}$. The higher BF_{01} the higher the strength of evidence in support of the null hypothesis, and the lower in support of the alternative.

Inversely, BF_{10} yields the strength of evidence in support of the alternative hypothesis.

Statistical analyses were conducted using the BAYESFACTOR package, version 0.9.12–4.2 (Morey, Rouder, Jamil, & Morey, 2015), using default priors of $\frac{\sqrt{2}}{2}$ for consistency with Rouder, Morey, Speckman, and Province (2012). Figure 3.4 shows Bayes factors of 4 relevant model comparisons between experimental and simulated data using:

A. Base line parameter combination.

B. Reference parameter combination.

C. Best fitting parameter combination with constant values.

D. Best fitting parameter combination with dynamic values.

Table 3.4 shows the values of these parameter combinations. For illustration purposes, Figure 3.5 shows line plots of experimental and simulated normalized entropy for each parameter combination selected.

Using normalized entropy levels in simulated and experimental early connectivity dynamics, we conducted two sample tests between groups for each round. Parameter combination D was associated with a better-model fit. In round 2, an estimated Bayes factor (null/alternative) of $BF_{01} = 6.4$ in favor of the null hypothesis suggested that the mean observed values are 6.4 times more likely to occur under a model with parameter combination C, rather than a model with any other parameter combination. In round 3, BF_{01} was 6 : 1 in favor of the null hypothesis, and in round 4, 6.4 : 1 in favor of the null hypothesis. According to Jeffreys (1961), these Bayes factors suggest ‘substantial’ evidence for the values of parameter combination D to produce entropy levels like those observed in the experimental data. A model including all the rounds yielded a BF_{01} of 12.31 : 1, suggesting ‘strong’ evidence for Parameter combination D. All these results together could be said to provide substantial to strong evidence that convergence in the experimental early connectivity condition might have been driven by a relatively strong content bias (0.8 during prealignment and 0.7 during postalignment), a slight egocentric bias (0.3), a dynamic variant value distribution (individuals went from two preferred variants $[1,0,0,1]$ during prealignment to one preferred variant $[1,0,0,0]$ during postalignment) and a memory size of about 3 rounds.

Using normalized entropy levels in simulated and experimental late connectivity dynamics, we conducted two sample tests between groups for

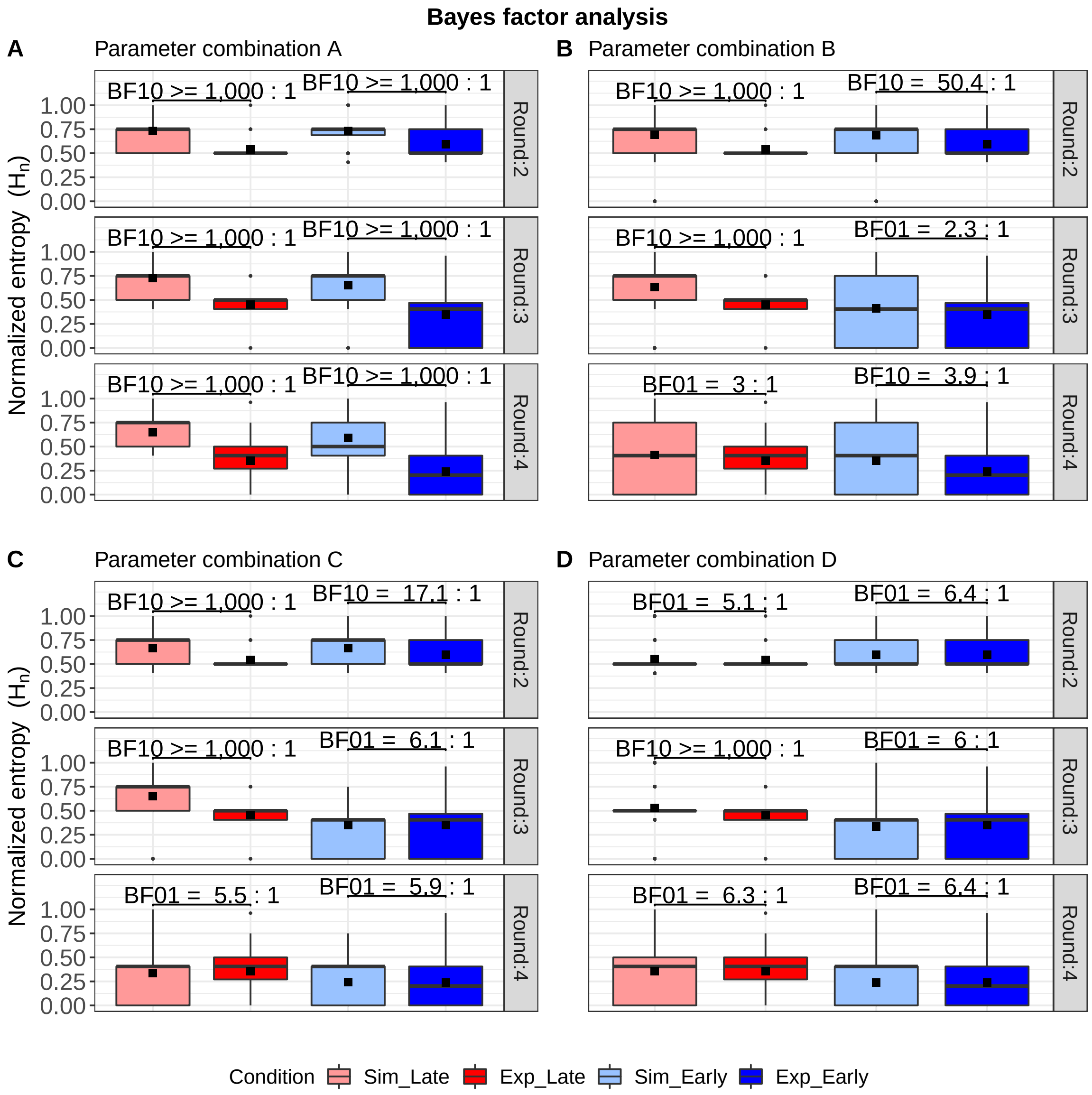


Figure 3.4: Comparison between simulated and experimental data. Diversity of communicative variants (measured as Normalized Shannon Entropy) over rounds and conditions. The lower and upper hinges correspond to the first and third quartiles—the 25th and 75th percentiles. Boxplot with whiskers with maximum 1.5 Interquartile Range (IQR). A. Base line parameter value combination against experimental data. B. Reference parameter value combination—best-fit model for Fay et al. (2010)—against experimental data. C. Best fitting parameter value combination (using constant values) against experimental data. D. Best fitting parameter value combination (using dynamic values) against experimental data

Parameter	Parameter combination A	Parameter combination B	Parameter combination C	Parameter combination D			
Connectivity	Both	Both	Both	Early		Late	
Value setting	Constant	Constant	Constant	Prealignment	Postalignment	Prealignment	Postalignment
Content bias	0	[0.1, 1]	0.7	0.8	0.7	0.9	0.6
Coordination bias	0.5	0.5	0.5	0.3	0.3	0.5	0.5
Memory	3	3	3	3	3	3	3
Variant value distribution	[1, 0, 0, 0]	[1, 0, 0, 0]	[1, 0, 0, 0]	[1, 0, 0, 1]	[1, 0, 0, 0]	[1, 0, 0, 1]	[1, 0, 0, 0]
Innovation rate	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Table 3.4: Values of selected parameter combinations in computer simulations. Parameter combination A corresponds to a base line model, with content bias 0 (null) and coordination bias 0.5 (null). Parameter combination B corresponds to the reference model, the one that was used to make the initial prediction. Parameter combination C corresponds to the best fitting parameter combination using constant values across rounds. Parameter combination D corresponds to the best fitting parameter combination using specific values for each connectivity and alignment phase.

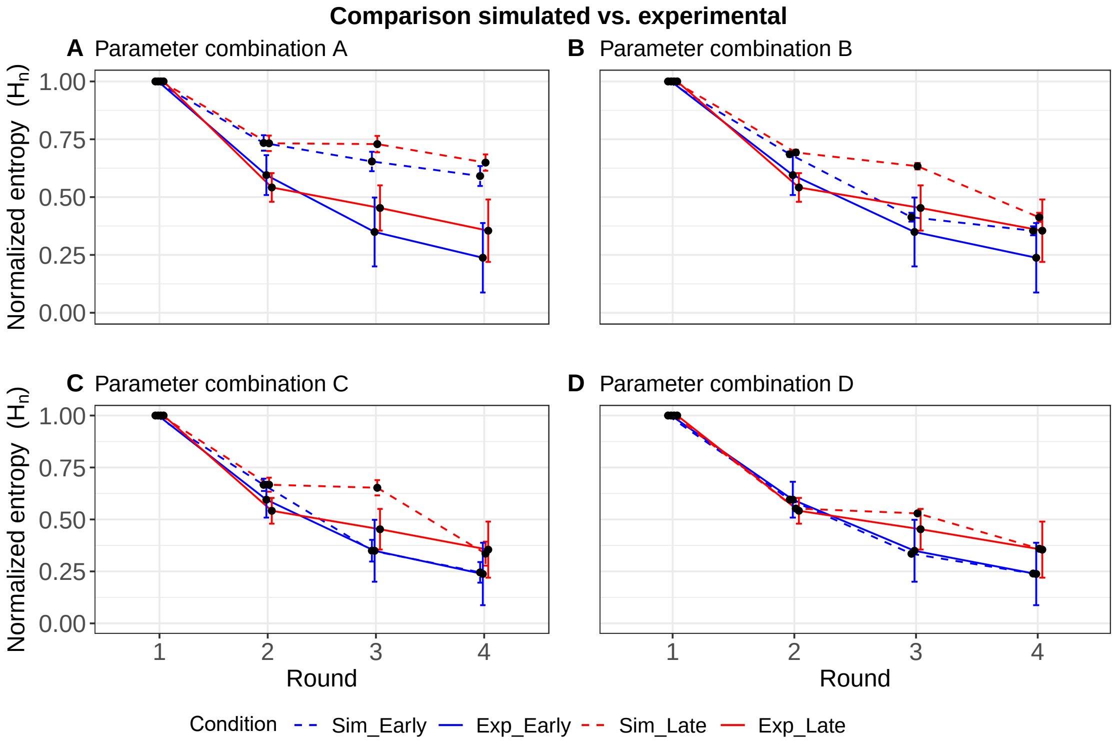


Figure 3.5: Comparison between simulated and experimental convergence trajectories over rounds and conditions. Error bars indicate 95% CIs.

each round. Parameter combination D was associated with a better-model fit. In round 2, an estimated Bayes factor (null/alternative) of $BF_{01} = 5.1$ in favor of the null hypothesis suggested that the data are 5.1 times more likely to occur under a model with parameter combination C, rather than a model with any other parameter combination. In round 3, Bayes factor was $BF_{10} \geq 1000:1$ in favor of the alternative hypothesis, and in round 4, $6.3:1$ in favor of the null hypothesis. These Bayes factors suggest ‘substantial’ evidence for the values of parameter combination D in round 2 and 4, and ‘very strong’ evidence for another parameter combination during round 3. A model including all the rounds yielded a BF_{01} of $6.6:1$, suggesting ‘substantial’ evidence for Parameter combination D. Broadly, all these results together could be said to provide substantial evidence that convergence in the experimental late connectivity condition might have been driven by a relatively strong content bias (0.9 during prealignment and 0.6 during postalignment), an intermediate coordination bias (0.5), a dynamic variant value distribution (individuals went from two preferred variants $[1,0,0,1]$ during prealignment to one preferred variant $[1,0,0,0]$ during postalignment) and a memory size of about 3 rounds. The model, however, did not accurately simulate the observed mean values of convergence in round 3 under late connectivity conditions.

In the next section, we discuss possible explanations for these results and suggest directions for future work.

3.6 Discussion

This study supports the hypothesis that the convergence of a small-scale society of human participants on shared communicative conventions can be affected by the order in which connections between individuals unfold over time. In particular, we have shown that a late connectivity dynamic can slow down convergence and an early connectivity dynamic can accelerate convergence when there is a preference for communicative variants with high adaptive value. Although our study differs from previous work on network connectivity dynamics in several respects (population size and experimental procedure), and it might be limited by the number of independent observations collected, our results are consistent with previous findings regarding some important points:

(i) Subtle changes in a population network structure can affect the formation of social conventions (Centola & Baronchelli, 2015).

(ii) Social influence seems to reduce diversity in networks where every agent has equal connectivity, when compared to networks where the agents' access to information is not equal (Becker et al., 2017).

Previous network experiments on larger groups (e.g. Judd, Kearns, & Vorobeychik, 2010; Kearns, Judd, Tan, & Wortman, 2009) have successfully addressed important questions related with these decision-making issues. However, those studies based on social-agreement tasks do not focus on the open-ended nature of human communication. In our experimental design, participants engaged in an open-ended, continuous, graphical communication task, allowing them to dynamically develop, negotiate and agree upon their own communicative conventions, much like human communication in the real world.

We also suggest that a key advantage of our experiment when compared to previous studies is that by testing the effects of a simplest network change in the smallest possible pairwise, interactive, fully-connected collaborative micro-society (4 participants, two pairs, 4 rounds with one repetition), we have been able to identify specific changes yielded by one specific and experimentally controlled change in the network connectivity dynamic. This small-scale experimental design allowed us to minimise the number of confounding variables, and facilitated the establishment of causality between the specific manipulation of the network connectivity dynamic and the convergence subsequently observed.

The basic structure of the agent-based model used in the present chapter was developed and described in Chapter 2. There, we compared our model against existing experimental data from one possible connectivity condition—early connectivity dynamic. In the work presented here, we have used the same model to exactly mirror the two network connectivity dynamics—early and late—implemented in the experiment. Mirror model predictions were useful to generate a testable hypothesis. To be sure, the prediction that an increase in global connectivity during the first rounds—early connectivity—would accelerate the formation of communicative conventions was proven to be statistically accurate for the experimental data analysed.

One of the most prominent explanations for the maintenance of cultural diversity is that individuals may have interests that are resistant to social influence (e.g. Friedkin & Johnsen, 1990; Marsden & Friedkin, 1993). Interestingly, contrary to model predictions, differences between conditions did not tend to reduce during postalignment. This suggests that subtle changes in the connectivity dynamic during the first interactions might affect the individuals' attachment to local variants in a fundamental way. Crucially, in the late connectivity condition, individuals interact with the same partner during two consecutive rounds. This isolation exposes individuals to the same conventions longer than in the early connectivity dynamic. By the time they encounter other individuals, they might have developed a stronger resistance to change their first conventions (even if the new observed variants have high adaptive value), which in turn decreases convergence and increases diversity in the form of polarization. In contrast, individuals under early connectivity conditions have been exposed to a higher diversity of conventions starting from an earlier round, which might weaken their resistance to adopt variants with high adaptive value. This might explain the maintenance of the differences in convergence between experimental conditions from round 3 to round 4 (see Figure 3.5).

We also present an attempt at evaluating the predictive accuracy of our model by conducting a Bayes Factor analysis. Our model evaluation shows that the model is able to accurately predict convergence under the set of experimental observations using an early connectivity dynamic. Notwithstanding, none of the parameter combinations examined under the late connectivity condition were able to accurately predict convergence across all rounds. Even if we successfully identified the best parameter combination from a pool of candidates, these parameters might be inadequate to generalise to other datasets or real situations. For example, our results agree with those reported by Tamariz et al. (2014) that a content bias and an egocentric bias were operating on the experiment. Also, that low memory size (3 rounds in our study, between 2 and 4 rounds in Tamariz et al. (2014)), was associated with better model fit. We nevertheless found differences in best-fit parameter values between these studies, with the present experiment generally showing a higher content bias. These differences might be due to individual cognitive variation combined with key differences in the experimental design of both studies (network structure, population size, sample size and coding method). Further experiments to collect data using different connectivity dynamics and samples that are representative of the whole population are needed.

Be that as it may, it is interesting to observe that at the beginning of round 3, under late connectivity, diversity falls bellow 0.5. Technically, this means that there is convergence between participants of different dyads before they interact. This suggests that at least part of the convergence is not driven by alignment between partners, but by either learning biases (Kalish et al., 2007; Kirby et al., 2008, 2015) or intrinsic properties of the variant representations that participants shared before the experiment. These intrinsic properties might be iconic affordances of the concepts used in the experiment for which participants already share a cultural basis: That is, common mappings between forms and meanings that are present in participants inventories of conceptual representations. Such a finding would suggest evidence of convergent cultural evolution, that is the emergence of similar forms across different communities (e.g. Caldwell & Millen, 2008; Dingemanse, Torreira, & Enfield, 2013; Mesoudi et al., 2006), although this result might be somewhat skewed by the application of the coding scheme, which is inevitably subject to a certain degree of subjectivity in the evaluation of analogies. Note that in the present study we performed a functional analysis, that is, two variants with similar form and function were considered as the same variant, even if that implies convergent evolution prior to the participant's first encounter. Otherwise, we could not have discerned whether two agents converged or continued with their own variant after interacting, because the variant would have had a similar form prior to the participants' first encounter anyway.

Here, we have shown that researchers should not neglect the role of the micro-scale connectivity dynamics when doing microcosm studies, as it may influence the spread of variant outcomes. Our research might be of interest to those researchers investigating the effects of network topologies, social networks or collaborative behavior. In particular, to those that use dynamic interactive micro-societies of agents switching partners over time (e.g. Baum et al., 2004; Caldwell & Smith, 2012; Fay et al., 2010; Mesoudi & Whiten, 2008).

In Chapters 4 and 5, we will build on our agent-based model to develop more sophisticated reinforcement mechanisms between agents' cognition and their social environment. These models will primarily be focused on the inclusion of institutional mechanisms and additional individual biases that affect social influence, along with a more realistic implementation of distributions of value across cultural variants in the population. By doing so, we will try to overcome some of the limitations of the current state of the ABM.

Chapter 4

Fixed institutions, homogeneity, hegemony and content bias

Abstract

Cultural diversity is crucial to maintaining healthy societies and is related to economic growth, social cohesion, productivity and biodiversity indices. A number of studies of cultural evolution have shown that agents' choices and value systems coevolve affecting diversity. But little is still known about how the interactions between individual cognitive biases, the distribution of values across variants and institutional reinforcement affect cultural diversity under global connectivity. High hegemony of cultural systems—where one variant is strongly preferred over others—and homogeneity of value systems among agents in the population may accelerate the extinction of cultural traits and thus detract from cultural diversity, as may institutions that impose or reinforce certain value systems. This chapter explores these and related hypotheses using agent-based models to simulate micro-societies where cognitive biases, value systems and institutional reinforcement are systematically manipulated. Our findings indicate that both hegemony and homogeneity of the value system accelerate the extinction of cultural variants, and that content-based transmission bias amplifies those effects. In certain regions of the parameter space, institutions that do not reinforce original value systems tend to be effective preserving cultural diversity, especially for intermediate levels of content bias. We extend previous work that shows that, under certain cognitive constraints, external pressures may dramatically and non-monotonically affect the way in which cultural evolution proceeds. We discuss a number of potential implications of these findings in the context of a globalised society.

Keywords: cultural evolution; cultural diversity; variant value; homogeneity; heterogeneity; content bias; agent based models

4.1 Introduction

4.1.1 Cultural diversity and selection pressures: a brief recap

Cultural diversity, the variety of cultural traits in a society, is the outcome of cultural transmission and evolution. As people use and interact with culture, certain variants of a tradition persist, while others die out, and this often results in sets of cultural variants that are shared by groups of individuals in a population at a given time. Although some processes, such as conformity, tend to limit diversity to make society functional (Newson, Richerson, & Boyd, 2007), cultural diversity is essential on many dimensions to maintain healthy societies, and it has been largely correlated with economic growth (Ager & Brückner, 2013; Ashraf & Galor, 2011), development (Montalvo & Reynal-Querol, 2005), productivity (Bellini, Ottaviano, Pinelli, & Prarolo, 2013) and biodiversity indices (Loh & Harmon, 2005; Maffi, 2005). Cultural diversity can also be perceived as a threat to social cohesion and trust (e.g. Huntington & Dunn, 2004; Twigg, Taylor, & Mohan, 2010), which has historically been a topic of preoccupation for governments and policy makers. However, some of these negative effects are not widely accepted (Letki, 2008), and have fuelled the debate on how to manage cultural diversity (Verkuyten, 2006). Certainly, the management of cultural diversity and cultural convergence, that is, the balance between the preservation of differences and equality over time, requires the development of cultural evolutionary models that systematically manipulate group identities and the role of internal and external selection pressures in order to make more useful predictions about the spread of behavior.

In recent times, we have witnessed considerable interest in the research of these dynamics of culture in different domains. There is an increasing number of studies seeking to identify the key parameters of diffusion using different approaches: for example, diffusion research tradition (Rogers, 2010), which focuses on the spread of innovations, dual-inheritance theories (J. Henrich & McElreath, 2007), which focus on gene-culture coevolution, and social network analysis (Centola, 2018), which focuses on how changes in people's social networks alter their behaviors. However, diversity is difficult to quantify because: (i) it can refer to different scales (local vs. inter-local) or dimensions (richness, evenness and distance) (Van Parijs, 2011; Whittaker, 1960), and (ii) different cultural traits spread following different dynamics (e.g. Boyd & Richerson, 1985).

Different cultural traditions have been explained by distinct evolutionary models. As mentioned in Chapter 2, some cultural traits are best explained by neutral (drift) models. These include patterns, decorations, pottery and baby names (Bentley, Hahn, & Shennan, 2004), dog breeds (Herzog, Bentley, & Hahn, 2004) and a number of diachronic changes in language (Komarova & Nowak, 2003; Kroch, 1989). Drift models are appropriate null models against which other models can be tested (Hahn & Bentley, 2003; Lipo, Madsen, Dunnell, & Hunt, 1997; Neiman, 1995; Reali & Griffiths, 2009; Shennan & Wilkinson, 2001).

Other cultural traits are best modelled by selection models. These models involve internal and external factors. Internal factors include cognitive biases that affect the individuals' likelihood of adopting a given variant. For example, *content-based biases*, also termed *direct biases* by Boyd and Richerson (1988), refer to the individuals' preference for intrinsic properties of cultural traits; as a consequence of this, some variants have higher probability of being adopted than others (J. Henrich & McElreath, 2007; Vale et al., 2017). If a method to solve a problem is particularly easy, or simple, or effective, all else being equal, people will tend to adopt this method over others. For example, violin sound hole shapes that maximise air-resonance power efficiency (Nia et al., 2015), or Edison's innovations to early telephone models towards maximising effectiveness (Ziman, 2000) have come to prevail. *Model biases* are related to the individuals' preference for cultural variants due to the prestige, skills, success or similarity to self of the individual displaying a behaviour (J. Henrich & McElreath, 2007). For instance, we preferentially buy the products or adopt the habits used or endorsed by celebrities, experts or people that we identify with. Other biases related to the individuals' cognition, such as frequency biases, payoff-biases, coordination bias and conformist biases, have been used applied to the the evolution of social learning rules (Acerbi et al., 2016; Boyd & Richerson, 1985; J. Kendal, Giraldeau, & Laland, 2009).

External factors that exert selection pressures include population size (J. Henrich, 2004; Kline & Boyd, 2010; Shennan & Wilkinson, 2001), degree of contact between agents in a population (Derex & Boyd, 2016; Muthukrishna et al., 2014; Powell et al., 2009), social network structure (Lee et al., 2005) and network connectivity dynamic (Centola & Baronchelli, 2015, Chapter 2). Environmental conditions and genes also play a part in the spread of cultural variants, with culture being an important mechanism to modify sources of natural selection by humans in their environment, through processes such as niche construction and gene-culture coevolution (Laland et al., 2000; Mesoudi, 2011a). Crucially, in human environments, social institutions (broadly defined as a system of rules, such as organizations (Hodgson, 2015)), have been suggested to constitute a prominent external mechanism that controls behavior, transcending individuals primary intentions by constraining agent's actions and beliefs at each particular moment (Giddens, 2013). In this chapter, we are particularly interested in the external influence exerted by hypothetical institutions that affect how much individuals' interactions shape value systems.

4.1.2 Content bias and dissimilar values

Selection models of dissemination of culture capture how these internal and external factors affect cultural diversity and its opposite, convergence. A number of models have done this for cases in which all agents in a population share the same *content bias* and the same distribution of value across cultural variants in each simulation run—e.g. Gong et al. (2012); Segovia-Martín et al. (2019); Tamariz et al. (2014) show how content bias shared across the population operates against diversity. However, in real life not everyone shares the same values, and different sub-populations may develop dissimilar *value systems* (e.g. opposing interests), which in turn might affect the spread of cultural variants. Using a computational approach, researchers on social behavior have shown that a *complete conflict of interest*, that is, when agents have different preference orderings over acts, can affect the maintenance of communication in signaling games (Godfrey-Smith & Martínez, 2013; Martínez & Godfrey-Smith, 2016).

The idea that different value systems can lead to different distributions of variants in a population and affect cultural diversity has already been explored in the past. For example, Axelrod's (1997) model of dissemination of culture was based on the assumption that people are more likely to interact with others who share the same cultural variants, and this in turn tends to increase the number of variants they share. These mechanisms, which he named *homophily* and *influence*, are prominent explanations for the persistence of cultural diversity (Axelrod, 1997). When combined with dynamic co-evolving networks, they can lead to stable cultural diversity in the face of cultural drift (Centola et al., 2007). Building on Axelrod (1997), researchers have found that several factors affect the dynamics of cultural change and cultural diversity, including globalisation (Greig, 2002), technological innovation (Leydesdorff, 2001), mass media (González-Avella et al., 2007; Shibanai, Yasuno, & Ishiguro, 2001), political institutions (Bhavnani, 2003) and cultural drift (Klemm, Eguíluz, Toral, & San Miguel, 2005).

The mathematical modelling of culture is a rapidly developing and highly productive discipline. Notwithstanding, the complex network of interactions between value systems, population biases and other types of reinforcement (such as institutional reinforcement) has received comparatively little attention. While the goal of Axelrod's model was to study why cultural diversity persists even though agents approach each other through a self-reinforcement mechanism of more interaction, our goal here is to investigate how the spread of cultural variants in a population is affected by three main factors: value systems in the population, agent content biases and institutional reinforcement of value systems. Specifically, we aim at modelling the extent to which cultural diversity can be best maintained under enforced global connectivity, and make predictions about the effectiveness of external reinforcement in different scenarios.

4.1.3 Value systems

In our study, an agent's *value system*, S, is operationalised as the set of values that the agent assigns to each cultural variant. Value systems can remain fixed, or they can evolve depending on the agents' choices or external reinforcement. Value systems encompass two variables: The number of independent value systems in use in a population, and the pattern of dominance between variants.

Regarding value systems, populations can be either *homogeneous*, when all agents share the same variant inventory and also the same value system—e.g. a population of mathematicians, who share the same mathematical conventions, and assign the same value, or meaning, to each of them—or *heterogeneous*, when the agents share the same variant inventory but there are two or more value systems—e.g. agents from different groups such as employers and employees may assign different value to variants such as *flexibility* and *precarity*, even when they refer to the same payoff matrix. Readers should therefore not confound *heterogeneity* of value systems (number of independent S) with the *heterogeneity* of the population (number of traits, q) as defined in related literature (Centola et al., 2007).

Regarding *hegemony*, or the relative dominance between variants in a value system, different scenarios include one variant being preferred over all the others—high *hegemony*, e.g. there is only one correct way to address the Queen; two variants that are preferred over all the rest of variants—e.g. bilingual speakers are happy to use either of the languages they are proficient in, but not other languages; or all variants being equally likely to be produced—low *hegemony*, e.g. if we are happy to greet our friends in any way, by shaking hands, saying hello, embracing, kissing, and so on.

We predict that high hegemony (one preferred variant over the others) and homogeneity (one value system shared by the whole population), will accelerate the extinction of cultural traits, lowering cultural diversity, while low hegemony (two or more preferred variants) and heterogeneity (two or more independent value systems across the population) will contribute to maintaining cultural diversity. We aim to explore the effect of value systems S, in combination with the degree of individual (content) bias in a population, on cultural diversity.

4.1.4 Institutions

Institutions have been defined by researchers on institutionalism as 'integrated systems of rules that structure social interactions' (Hodgson, 2015, p. 501). Similarly, Boyd and Richerson (2008) define social institutions as norms and conventions that give durable structure to social interactions within a population. Institutions are not only mere providers of goods and services, they also influence the evolution of values, tastes, and personalities (Bowles, 1998; Frank, Gilovich, & Regan, 1993; J. Henrich et al., 2005).

In the present chapter, we will use the notion of *institutional reinforcement*, which models the effort on the part of an institution to favor a particular value system, and may lead to changes in the value systems of individuals. Examples of institutional reinforcement include the approval or condemnation of particular usages of grammar and vocabulary by national linguistic academies and the prescriptions of a moral code that encourages people to do what is right and not to do what is wrong.

Theories of collective behavior suggest that institutional mechanisms (e.g. the introduction of collective incentives) reduce cultural diversity by facilitating the formation of social conventions (David, 1994; Kearns et al., 2009; Voigt & Kiwit, 1998; Young, 2001), although it has been shown that social conventions can also emerge without the intervention of institutional mechanisms (Becker et al., 2017). Institutions are not always group-beneficial (Boyd & Richerson, 2008); regarding this issue, important progress was made in the past as researchers developed agent-based simulations to explore under which circumstances co-evolutionary processes generate high levels of group-beneficial behaviors along with institutions (Bowles et al., 2003). However, how the evolution of variant values in the presence of institutional reinforcement affects the emergence of social conventions in a population when individuals have different degrees of sensitivity towards variant value has remained relatively under-explored.

In the current chapter we will use simulated micro-societies of agents in which we systematically manipulate the degree of preference for variants with high value (content bias), the number of value systems in the population (homogeneity/heterogeneity), the degree of relative dominance between variants (hegemony) and the degree of institutional reinforcement of value systems. We aim to establish the effect of these variables on the cultural diversity of a population.

4.2 Methods

We address the research aims presented above by analysing the spread dynamics of n competing variants of a cultural trait within a micro-society. Next we compare alpha diversity levels in micro-societies that differ in the properties of their values systems and the level or institutional reinforcement. In the next following we describe the probabilistic model.

4.2.1 The probabilistic model

The present study uses an extended version of the agent-based model we developled in Chapter 2[1], in which 8 agents play recurring games in a pairwise interactive micro-society. A description to obtain the mathematical model can be found in Appendix B. The initial state consists of 8 agents, and each agent i is randomly assigned a cultural variant σ_i selected from a pool of variants without replacement. The dynamics of the model are described as follows:

At the beginning of each round r, agents are assigned a partner. Each pair interacts, and an interaction involves the following: First, each agent produces a variant sampled from its memory according to equation Equation (4.2) on page 86. There is a small probability of mutation, in which case the variant is not sampled from memory, but it is a novel one sampled from the pool of possible variants. For example, when agent i and agent j interact, i produces variant σ_i and j variant σ_j. Next, both agents store the two variants in their memory stores. At each round, agents switch partners so that by the end of the final round every agent has been paired with every other agent. Each agent is initialised with a value system S, which is a vector of floating points that indicate the value s assigned to each variant and indicates to what extent the variant is preferred over the other variants —s takes values ranging from 0 (null value) to 1 (highest value)—.

The model takes several parameters as described below:

Heterogeneity: In the heterogeneous condition, the population is divided into two sub-populations (formed by agents 1–4, and agents 5–8, respectively). Each sub-population is assigned a different value system. In the homogeneous condition, all agents are initialised with the same value system.

Hegemony: It is operationalised as the distribution of value across variants, and has three levels:

(i) One-takes-all (OTA) simulates a fully hegemonic cultural system. One preferred variant (σ_1) has value = 1, the others, 0.

(ii) Competition (C), where two competing variants (σ_1 and σ_2) have intrinsic value 1, and the rest, 0.

(iii) Pseudo-random (PR), where values were assigned to variants in the set $\{\sigma_1, \dots, \sigma_8\}$ using random sampling without replacement from a list containing values from 0 to 1 in steps of 0.1.

---

[1] Code available at https://github.com/jsegoviamartin/network_connectivity_dynamics_model

Institutional performance (ε): Is a parameter identifying to what extent the institution reinforces the agents' S. It reflects the extent to which institutions limit how much the agents' interactions can shape their own value systems (S). Technically, for each agent i and round r, if a variant $\sigma_i \in h$ (where h stands for the agent' history), then the value assigned by an agent i to a variant σ_i at round $r+1$ is a function of the current variant value s_{ir}, choice $|\sigma_{ir}|$ and insitutional performance ε. At each round, $|\sigma_{ir}|$ is 1 if σ_{ir} is produced by an agent i, 0 otherwise. Two conditions of institutional performance were examined:

(i) *Reinforcement* (Rf), where a global institution favours a fixed value system ($\varepsilon = 1$) (which is the original system the agents are initialised with).

(ii) *Responsiveness* (Rp), where the global institution weakens the original value system ($\varepsilon = \frac{1}{2}$), allowing value systems to evolve according to the agents' choices.

The value assigned by a typical agent i to a cultural variant σ_i at each round r is updated according to the following equation:

$$s_{ir+1}(s_{ir}, \sigma_{ir}, \varepsilon) = \begin{cases} s_{ir}\varepsilon + |\sigma_{ir}|(1-\varepsilon) & \text{if } \sigma_i \in h \\ s_{i0}, & \text{otherwise} \end{cases} \tag{4.1}$$

where s_{ir+1} stands for the value of variant σ_i at round $r+1$.

Content bias (β): Identifies the degree of preference for variants with high value. It encompasses two parameters (b,d). Parameter b is the agents' sensitivity to variant value (s), and ranges from 0 (not sensitive at all) to 1 (fully sensitive) in steps of 0.1. Parameter d specifies whether the variant is in the agent's memory record, and equals 1 if the variant is in memory, and 0 otherwise. Parameter β is equal to $b \cdot d$. Thus, content bias (β) assigns a value from 0 to 1 to each variant. When content bias is 0, we have a model of drift, or neutral evolution. Content bias values from 0 to 1 in steps of 0.1 were examined.

For each parameter combination examined we ran 1000 simulations. In each simulation, the pair composition was randomized without repetition. All networks generated were fully connected networks. Our model includes two other parameters, which, for the purposes of this study, take fixed values: First, memory size (m) represents the agents' access to their memory store, and it was limited to 3 rounds in the memory ($h_{|M,3}$). Memory sizes between 2 and 4 were associated with better model fits in Tamariz et al. (2014). Second, mutation rate (μ), or the probability that when agents produce a variant, they do not sample from their memory store, but from the initial set of 8 variants in the micro-society. We fixed the innovation probability at 0.02 by using a flat distribution $\phi(\sigma)$ of variants weighted by mutation rate. This reflects the innovation rate fitted by Tamariz et al. (2014) for Fay et al. (2010) experimental data (Tamariz et al., 2014).

The model described above can be formalized as follows:

$$\Pr(\sigma_{ir} \mid h_{ir}) = \mu'\beta' f(\sigma_i \mid h_{|M,3}) + \mu'\beta s_{ir} + \mu\phi(\sigma) \tag{4.2}$$

where $Pr(\sigma_{ir} \mid h)$ yields a probability of production of variant (σi) by a typical agent i at round r for a given history h_{ir} of previous rounds. The over-bar denotes the probabilistic complement (e.g. $a' = 1 - a$).

4.2.2 A numerical example

Let $P = (p_{i,r})$ be a matrix where each element in the matrix represents the probability of production of a variant σ_i at round r by a given agent:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{18} \\ p_{21} & p_{22} & \cdots & p_{28} \\ \vdots & \vdots & \ddots & \vdots \\ p_{81} & p_{82} & \cdots & p_{88} \end{bmatrix} \tag{4.3}$$

Thus, P_{*3} refers to the column vector that contains the probability of production of each variant at round 3 by such typical agent:

$$P_{*3} = \begin{bmatrix} p_{13} \\ p_{23} \\ \vdots \\ p_{83} \end{bmatrix} \tag{4.4}$$

Here, p_{13} is the probability that the agent produces variant σ_1 at round 3, p_{23} is the probability that the agent produces variant σ_2 at round 3, and so on.

In order to calculate S, we need information about the frequency and the value of each variant for that agent. Regarding frequencies, assume that the current relative frequency of variants in the agent's memory is defined by the following vector:

$$F_1 = \begin{bmatrix} f_{13} \\ f_{23} \\ \vdots \\ f_{83} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.25 \\ \vdots \\ 0.25 \end{bmatrix} \tag{4.5}$$

where f_{13} stands for the number of times the agent has added σ_1 to its memory divided by the total number of variants in memory by round 3, and so on. In the present example, the agent has stored σ_1 twice in its memory, which means that the relative frequency of σ_1 is 0.5. Our typical agent has also stored σ_2 and σ_8 once, which means those are all the variants the agent has in memory by round 3. Note that at each round, agents store two variants in memory, the one they produce and the one they observe.

By round 3, let's consider that the value system of the agent is defined by the following vector:

$$S = \begin{bmatrix} s_{13} \\ s_{23} \\ \vdots \\ s_{83} \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.2 \\ \vdots \\ 0 \end{bmatrix} \tag{4.6}$$

where s_{13} stands for the relative value assigned by the agent to σ_1 at round 3, s_{23} stands for the value assigned to σ_2 at round 3, and so on. Variant values are determined in the previous round (Equation (4.1) on page 85). In our example, if we consider a situation in which content bias is quite strong (e.g. $\beta = 0.9$), then we have:

$$P_{*3} = 0.98 \cdot 0.1 \begin{bmatrix} 0.5 \\ 0.25 \\ \vdots \\ 0.25 \end{bmatrix} + 0.98 \cdot 0.9 \begin{bmatrix} 0.8 \\ 0.2 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0.0025 \\ 0.0025 \\ \vdots \\ 0.0025 \end{bmatrix} \tag{4.7}$$

Thus, the operation yields:

$$P_{*3} = \begin{bmatrix} 0.7571 \\ 0.2034 \\ \vdots \\ 0.0027 \end{bmatrix} \tag{4.8}$$

which corresponds to the probability of production of each variant as determined by the state attained in the previous round in the Markov chain.

Thus, at round 3, the probability of σ_1 being produced by the agent is $p_{13} = 0.7571$, while $p_{23} = 0.2034$ and $p_{83} = 0.0027$. The probability of the rest of the variants is 0.0025.

4.2.3 Quantifying cultural diversity

Cultural evolutionists have developed quantitative measures of cultural diversity by borrowing related alpha diversity indices from biology (Kandler & Laland, 2009). In the present study, we use two complementary measures, Shannon's entropy (H) which represents the distribution of variant tokens of each type in a population, and Simpson's Evenness (E), which is sensitive to the number of different types present in the population.

Entropy (H). This is the same metric used in Chapter 2. For the convenience of the reader we reproduce the formula below:

$$H(V) = \sum_{v_i \varepsilon V} p(v_i) \log_2 p(v_i) \tag{4.9}$$

where V corresponds to the set of variants, and $p(v_i)$ is the probability of i^{th} variant in that set. High entropy corresponds to high cultural diversity.

Evenness (E). We also use a common index of evenness called Simpson's evenness (E), to quantify how equal the distribution of variants are (Simpson, 1949). In order to calculate E, first we obtain dominance, which is defined as:

$$\text{Dom} = \sum_{i=1}^{S} p_i^2 \tag{4.10}$$

where (p_i) stands for the proportion of the entire community that variant i represents. Now, evenness (E) can be calculated by taking the reciprocal of dominance (Dom) and expressing it as a proportion the total number of variants present:

$$E = \frac{\frac{1}{\text{Dom}}}{S} \tag{4.11}$$

where S corresponds to the total number of existing variants in the system (richness). Evenness (E) takes a value between 0 and 1, with 1 being complete evenness. When E is complete, all the surviving variants are equally abundant in the system.

Extinction. We also borrow from biology a well established survival estimator to calculate extinction probability (Jolly, 1965; Seber, 1965). We define extinction probability as the probability that a cultural variant present in the community during round i is not present at some later round j. We first compute the number of variants observed in round i (R_i), and then the number of these variants still present in round j (R_j). Extinction probability is then calculated as:

$$1 - \hat{\phi}_{ij} = 1 - \frac{R_j}{R_i} \tag{4.12}$$

where $\hat{\phi}_{ij}$ is the complement of extinction probability and denotes the probability that a variant present in i is still present in j.

4.3 Results and discussion

We first present the results of simulation runs with institutional reinforcement, in which agents' value system are fixed (Rf model), in different scenarios of homogeneity, hegemony and content bias. Then we compare simulations with institutional reinforcement against those with an institutional responsiveness (Rp model) in the same scenarios.

At the beginning of the first round r_1, each agent produces a unique signal, so that at that point of the simulation, diversity indices are highest in all conditions. From then on, diversity decreases as agents begin to converge on a shared or conventional cultural variant. The rate of diversity loss will depend on the parameter combination selected.

4.3.1 Entropy

The entropy of the variants in the population is affected by homogeneity, hegemony and content bias (Figure 4.1). Highly hegemonic systems (one-take-all, henceforth OTA) lose entropy the fastest, followed by competition between two highly valued variants (C) and finally pseudo random ones (PR). Homogeneous systems lose entropy faster than heterogeneous ones. And content bias amplifies the effects of both homogeneity and hegemony. PR maintained diversity better than OTA and C.

While drift models remain indifferent to changes in the homogeneity of value systems in all the scenarios examined, content-biased models are highly sensitive to it. The higher the content bias the less diversity, and the more heterogeneity contributes to the maintenance of diversity levels.

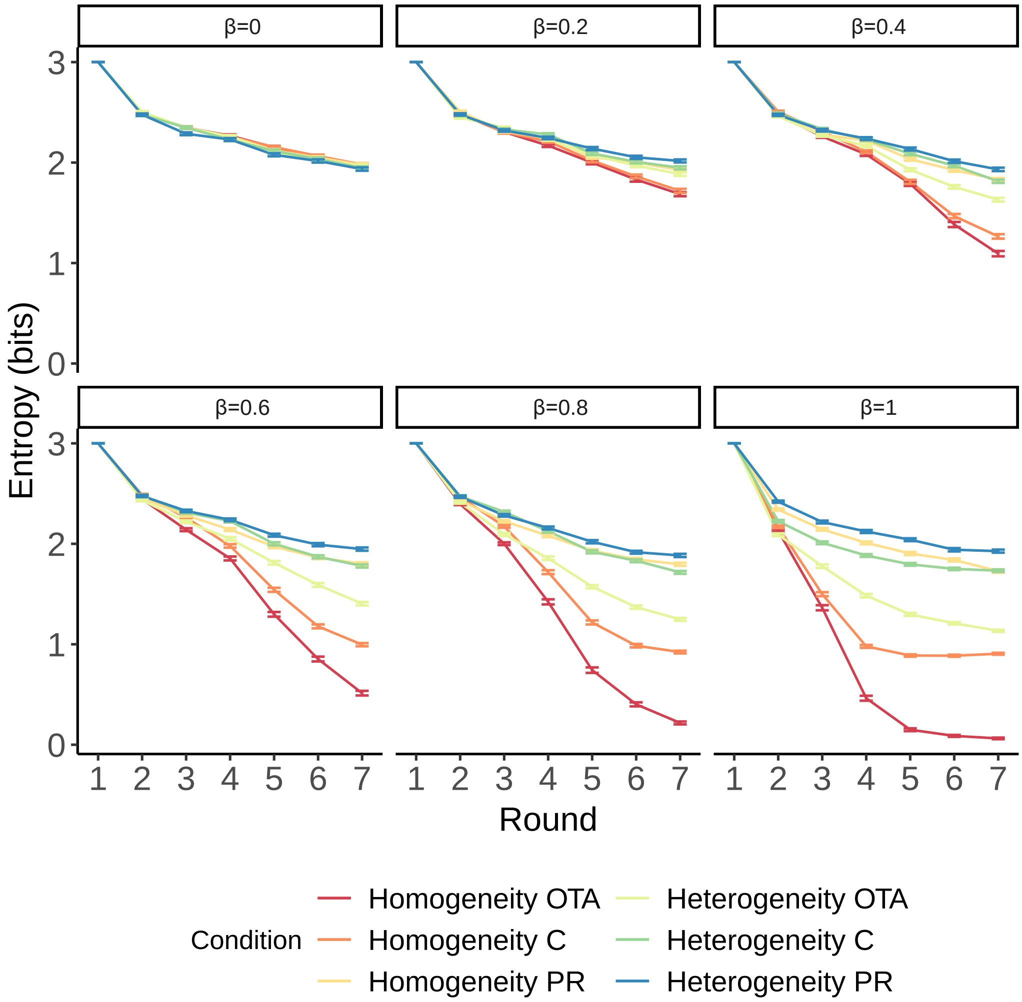


Figure 4.1: Entropy (H) of the distribution of cultural variants in the population averaged over each level of content bias (β), homogeneity and hegemony. Drift models ($\beta = 0$) are shown in the top-left throughout the study. Error bars indicate 95% CIs throughout the study.

With high hegemony of one variant over the others (OTA), the population tends towards local convergence in heterogeneous populations, and towards global convergence in homogeneous populations. The reason is that in heterogeneous increases, the different agent sub-populations converge on their own cultural convention, and this can lead to global polarization even under high hegemony. This is because agents within a sub-population share the same value system (S) but agents across sub-populations have different systems. Interestingly, these results are similar to the effects yielded by homophily in Axelrod's (1997) model of dissemination of culture, where homophily leads to local convergence and an increase in polarization. This suggests that institutional reinforcement in fully connected micro-societies plays a similar role to that of self-reinforcement in dynamic social networks when it comes to explaining the maintenance of cultural diversity.

4.3.2 Simpson's evenness

The evenness of the cultural variants E is affected by hegemony, in interaction with homogeneity and content bias (Figure 4.2). In high-hegemony (OTA) homogeneous conditions, E decreased in the first rounds. It increased later on as the population resettled an equilibrium over the shared surviving variants, which were fewer in number, but more equally distributed in the system than in the previous rounds. These effects are strongly amplified by content bias. At very high levels of content bias ($\beta = 0.8$ and $\beta = 1$), in the last rounds, E was higher in homogeneous than the heterogeneous condition. This is because when there are few competing variants, it is more likely that an equitable balance of variants will be reached if the whole population share the same value system. In high-hegemony (OTA), heterogeneous conditions, the effects on E are qualitatively similar, but weaker than in the homogeneous conditions.

In the competition hegemony condition (C), the effects of homogeneity and content bias are qualitatively similar, but weaker again. In the pseudo-random (PR) hegemony condition, homogeneity and content bias do not change the evolution of E over rounds. Interestingly, at high levels of content bias, OTA and C yield higher E levels than PR in both conditions in the last rounds. This is because in OTA and C the selective pressure is stronger than in PR, which leads to greater extinction, reducing the number of variants and therefore increasing the probability of an equitable distribution among the surviving variants.

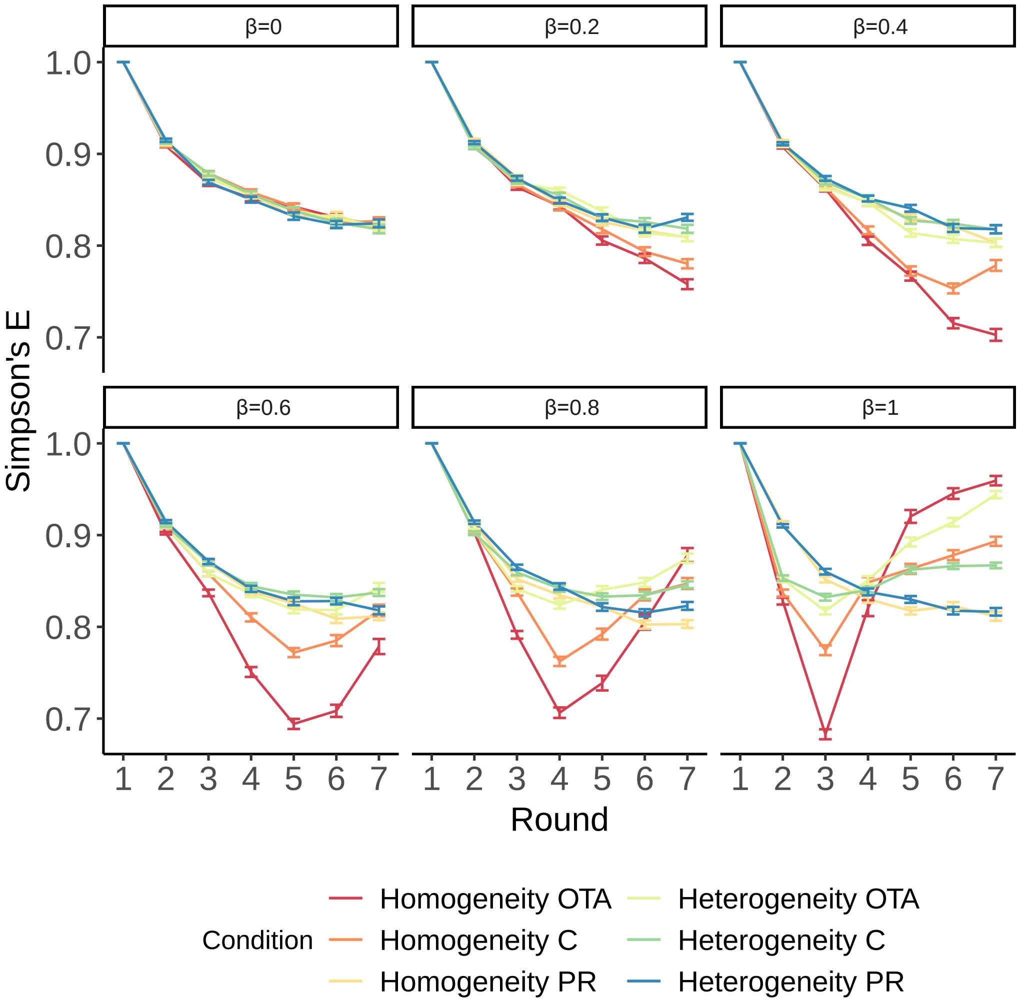


Figure 4.2: Simpson's evenness (E) of the distribution of cultural variants in the population averaged over each level of content bias (β), homogeneity and hegemony.

4.3.3 Extinction

Homogeneity, hegemony and content bias affect extinction rate, or the probability that a variant type is lost at each round. Figure 4.3 clearly shows two distinct periods in dynamics of extinction over time. In the initial period (r_1, r_2) we observe a high extinction probability. This is driven by the loss of variants with low intrinsic value, which only occasionally reappear later on with negligibly low frequencies due to mutation. During the late extinction period ($r_3, \ldots, r_7$), the surviving variants compete. Here, homogeneity in combination with high hegemony (OTA) accelerated extinction when compared with the other conditions. The subsequent decline observed after the peak in extinction happens when there is only one variant type left and extinction is not possible any more.

Summing up, our model shows that in highly hegemonic systems with high individual content biases (both pushing diversity down), the contribution of heterogeneity to entropy and evenness level is higher because it remains as the only force pushing diversity up. In competitive (C) or pseudo-random (PR) scenarios, heterogeneity is no longer the only pressure helping to maintain diversity, and this is why its contribution to diversity is lower. In the PR conditions, the persistent overabundance of cultural variants and the heterogeneity of S in agents actually prevent the formation of distinct cultural sub-populations with different distributions of variants, and thus cultural diversity must be explained by the relaxation of constraints on the agents' value system.

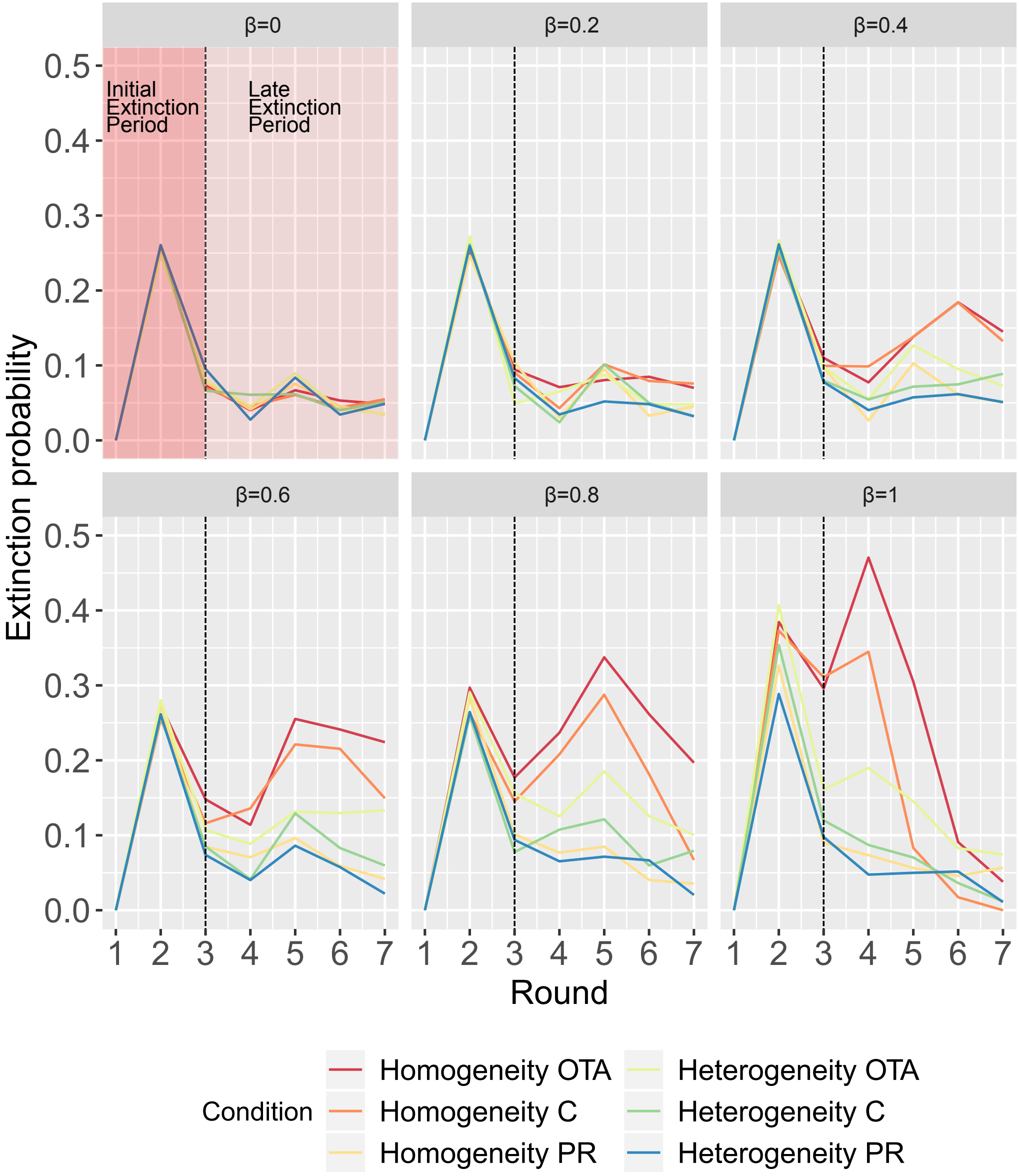


Figure 4.3: Extinction probability of cultural variants in the population averaged over each level of content bias (β), homogeneity and hegemony.

4.3.4 Institutional reinforcement

In the results presented so far, the value system S remained fixed, that is to say, regardless of the type of variant one agent chose to produce in one particular round, the original S for that agent remained the same. We interpret those as cases in which a prescriptive global institution reinforce the agent's original value system (or, in heterogeneous populations, systems). We label those cases as *reinforcement* (Rf).

We now consider an alternative scenario in which the agents' value systems (S) change depending on their behavior (the choices they make). We use for these cases the label *responsiveness* (Rp). Responsiveness can be interpreted as the outcome of a dynamic institution that promotes the values emerging from social interaction and use. Examples of emerging values include the changing attitudes towards gender violence or environmental protection and the shifting values attached to terms for disabled people. In this scenario, the agents' choices (behaviour) co-evolve both with the agents' value systems and with the level of reinforcement applied by the global institution. We address the question: How does institutional performance that responds to the agents' choices affect cultural diversity?

In the comparison of Rf and Rp we examine only Shannon entropy (H) as the measure of diversity and consider the three levels of hegemony separately for clarity. As seen in Figures 4.4, 4.5 and 4.6, cultural diversity is significantly influenced by the type of institutional performance. Compared with reinforcement, a responsive institution maintains higher levels of diversity, but this effect depends on the degree of hegemony of preferred variants, the level of context bias β and the level of homogeneity of value systems S in the population.

Figure 4.4 shows the results for the fully hegemonic condition (OTA). The effect of institutions is stronger under *homogeneity* than *heterogeneity*. Diversity (measured as entropy) differences between conditions (Rf vs. Rp) are higher in homogeneous than heterogeneous populations, particularly for intermediate values of β. Thus, non-interventionist, responsive institutions are more effective at maintaining cultural diversity in moderately biased populations where all agents share the same variant values. When $\beta = 1$, the difference across institutional performance is reduced. Here, agents are extremely biased towards the high value variant, and this reduces the probability that they change their behavior and choose another variant. In this scenario, since the agents' choices are strongly driven by their value system, choices remain mostly constant over time. Thus, even if agents are potentially able to evolve an alternative value system, they simply do not do it, which explains the limited effect under a responsive institution.

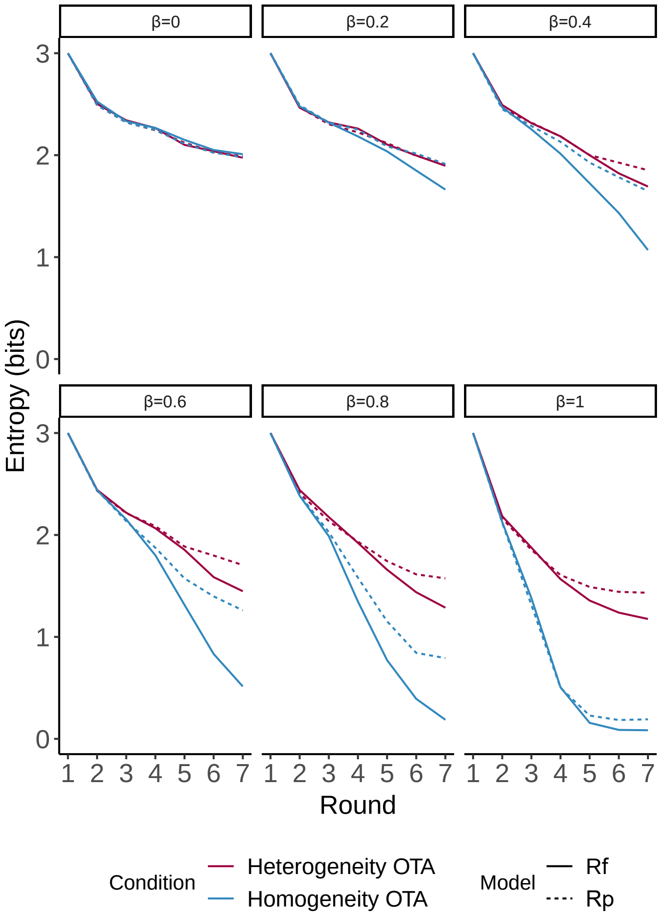


Figure 4.4: Cultural diversity (measured as Shannon Entropy) averaged over each level of content bias and homogeneity under OTA, comparing institutional reinforcement (Rf, solid line) and institutional responsiveness (Rp, dashed line).

Figure 4.5 shows that, in a competitive scenario (C), the results are similar as far as homogeneity is concerned. However, in contrast to OTA, here the loss of cultural diversity is very limited. This is due to the dispersion of value across the two competing variants: the less hegemony in the cultural system, the smaller the effect of the institutions. In other words, Rp is more effective in facilitating social conventions and Rf is more effective in increasing cultural diversity. The effect of $\beta = 1$ here is more pronounced under the competitive (C) than in the OTA hegemony condition. Since there are now two equally preferred variants, even if, under responsiveness, agents change their choice (which has a low probability in the first place), they will simply change from one preferred variant to the other. This means that their choices change very little, similarly to what happens under reinforcement, which maintains the original value distribution.

In contrast with OTA and C, the effect of the institutional model is not noticeable in the pseudo-random scenario (PR) (Figure 4.6). This indicates that a great dispersion of value across the variants in the initial state drastically inhibits the institutional effect implemented in the model and maintains high diversity, reducing the convergence of the population on a shared, conventional cultural variant. In the responsive scenario, in the absence of one or a pair of preferred variants, variants are lost more slowly and randomly across the population (some agents will lose a variant, other agents will lose another). This means the frequency distribution of variants across agents does not change. Therefore, the outcomes of responsiveness Rp and reinforcement Rf are very similar.

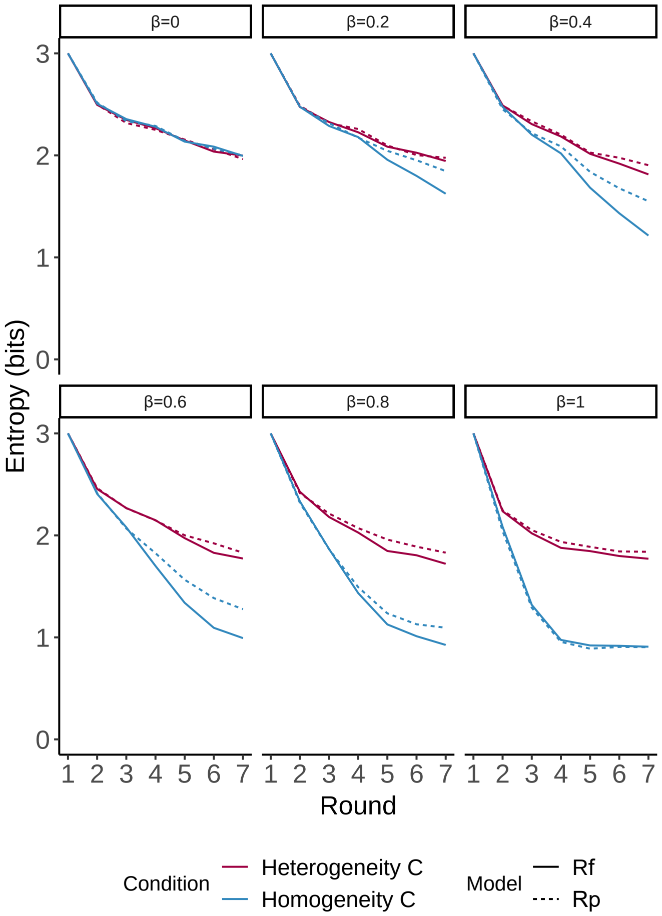


Figure 4.5: Cultural diversity (measured as Shannon Entropy) averaged over each level of content bias and homogeneity under C, with institutional reinforcement (Rf, solid line) and institutional responsiveness (Rp, dashed line).

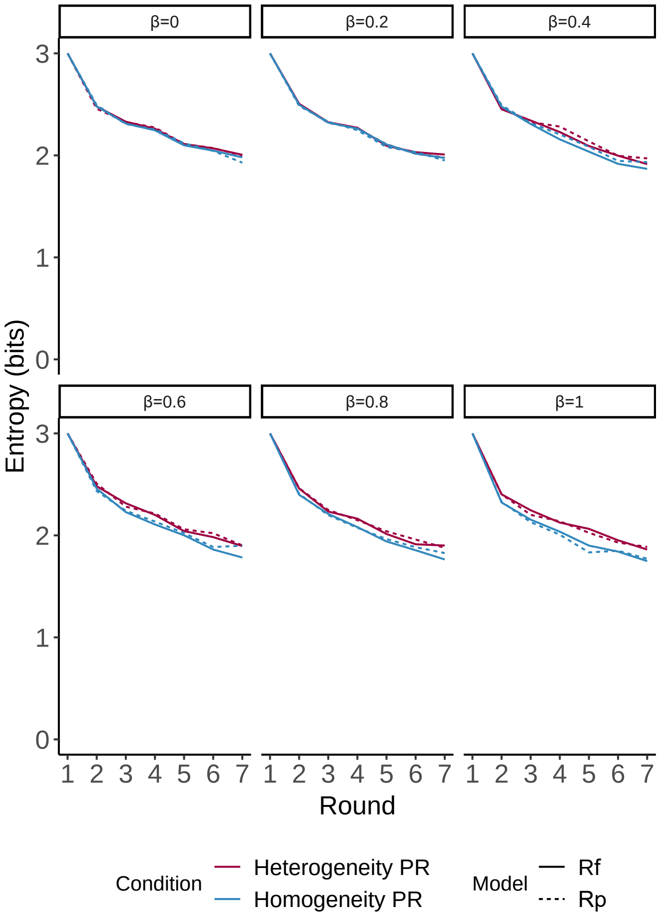


Figure 4.6: Cultural diversity (measured as Shannon Entropy) averaged over each level of content bias and homogeneity under PR, with institutional reinforcement (Rf, solid line) and institutional responsiveness (Rp, dashed line).

4.4 Conclusions

In this chapter we have extended an existing model of dissemination of cultural variants in a micro-society (see Chapter 2) to study the evolution of cultural diversity. The new model manipulates the hegemony of value systems (the degree to which a minority of variants are preferred over the rest), the agents' content bias (how sensitive agents are to hegemony), the homogeneity of value systems in the population, and whether institutionally promoted values are sensitive to the agents' choices.

Our simulation results show that diversity is lower in populations with homogeneous value systems and in highly hegemonic value systems. These effects are amplified by content bias. Regarding the effect of institutions, in scenarios with responsive institutions in which the value systems emerge from the choices of the population, diversity is maintained to a higher degree than in institutions that reinforce an existing value system. However, the effect of institutions interacted with hegemony, homogeneity and content bias: Diversity differences between conditions (Rf vs. Rp) are greater in high-hegemony value systems and homogeneous populations, particularly for intermediate levels of content bias. This latter point suggests that institutional intervention is more effective at maintaining cultural diversity when the agents' behaviours are not extreme. In addition, the effect of institutions may coexist with other local processes (e.g. coordination and transmission) and can vary over time and across levels of analysis (Bishop & Wößmann, 2004; Dacin, 1997).

Our results are consistent with existing theories of collective intelligence suggesting that institutions play an important role on the emergence on cultural conventions (David, 1994; Kearns et al., 2009; Voigt & Kiwit, 1998; Young, 2001), although we agree that conventions can also emerge without the intervention of institutional mechanisms that facilitate these processes (Becker et al., 2017). In our simulations, this was the case especially in the presence of extreme individual biases.

An obvious limitation of our model is that institutional performance remains fixed over time. This assumption can be relaxed by constructing a more comprehensive co-evolutionary model which includes a dynamic system of institutional values and agents' value systems. That will be precisely what Chapter 5 will be devoted to. We will construct an agent based model which includes a more realistic dynamic system of institutions, value systems and agents' choices. Additional individual biases that affect social influence (e.g. conformity and confirmation biases) will also be examined.

What we presented in Chapter 4, therefore, is the simplest possible model that allows us to examine the dynamics of cultural variant choice and value systems in different scenarios of institutional performance. Our model can be useful to inform how the interaction between institutional power and individual biases affect cultural diversity and human cognition.

Chapter 5

A co-evolutionary model of institutions, value systems and cognitive biases

Abstract

In the previous chapter we developed a model of the joint evolution of value systems and choice in different scenarios of institutional performance. Although such model can be useful to inform how the interaction between institutional power and individual biases affects cultural diversity and human cognition, an obvious limitation was that institutional performance remains fixed over time. In the present chapter we develop a co-evolutionary model of value systems, institutions and choice to explore how the dynamics of cultural diversity in populations with different levels of cognitive biases and institutional power evolve. To the best of our knowledge, our mathematical model is the first attempt to quantify the propagation of cultural variants by incorporating a comprehensive parameter combination of compliance, confirmation, content and frequency biases into the learning and production algorithm. Results show that, in some regions of the parameter space, institutional power facilitates the emergence of shared cultural conventions when compliance biases increase. In general, a compliance bias pushes diversity up when institutions are diverse, and pushes diversity down when institutions convey value systems with strong dominance of one or few cultural variants. Interestingly, in some scenarios, a decrease in institutional power and compliance bias allows the emergence of cultural conventions from the mutual reinforcement of local interactions and institutional values. We asses the robustness of these results by examining how sensitively they depend on different initial conditions of variant assignment, population sizes and alpha diversity indexes. **Keywords:** cultural evolution; co-evolution; institutions; value systems; cognitive biases; compliance bias; confirmation bias;

5.1 The co-evolution of institutions and value systems

One thing that makes humans special is their remarkable capacity to cooperate and build complex niches (Fuentes, Wyczalkowski, & MacKinnon, 2010; Kobayashi, Wakano, & Ohtsuki, 2019; Laland et al., 2000). As a result of social learning, which involves learning through copying and sharing information with others, humanity has succeeded to construct institutional environments which interact with the individuals' cognition, affecting evolutionary dynamics and the distribution of culturally transmitted traits in a population (Bowles et al., 2003). Yet there are very few models that explore the co-evolutionary dynamics of institutional diversity and population cultural conventions at individual, institutional and historical levels by incorporating a comprehensive set of relevant learning biases. Therefore, it remains unclear how populations with a different characterization of biases and institutions with different degrees of power and interests, all together, co-evolve social conventions and cultural diversity over time.

Although it has been shown that social conventions can emerge without the intervention of institutional mechanisms (Becker et al., 2017), prominent theories of collective behavior suggest that institutions reduce cultural diversity by facilitating the emergence of social conventions (David, 1994; Kearns et al., 2009; Voigt & Kiwit, 1998; Young, 2001). It is widely recognised that these variance-reducing institutions constitute niches, that is, modified environments which in turn are capable of altering sources of variation such as individual cognition and collective behaviours (Bowles et al., 2003; J. Kendal et al., 2011; R. Lewontin, 1965). Notwithstanding, the nature of the institution itself (i.e. whether the institution conveys diverse values or hegemonic values) and the institutional capacity to transmit values to the population might interact in ways that potentially amplify or reverse the mentioned sources of variation and the formation of social conventions over time. In this chapter, we seek to answer these questions by using a co-evolutionary model.

Evidence gathered over recent decades overwhelmingly indicates that cognitive diversity varies substantially across human populations (J. Henrich et al., 2005; J. Henrich, Heine, & Norenzayan, 2010; Levinson & Levinson, 2003). Understanding such source of variation implies understanding how transmission biases affect the spread of cultural variants during transmission processes (Boyd & Richerson, 1985; Feldman & Cavalli-Sforza, 1976). We know that biased reasoning strategies can be adaptive and yield fitness advantages when compared to unbiased ones. A number of these biases have been defined and modelled (Boyd & Richerson, 1985; J. Henrich & McElreath, 2003). However, despite the importance of biased transmission in cultural evolution, there is a lack of formal co-evolutionary models integrating a diverse range of cognitive tendencies for the exploration of the complex dynamics of value systems and institutions.

Related to biased adoption and transmission of variants, we investigate three relevant oppositions between biases which play an important role in the maintenance of cultural diversity, namely:

(1) A disproportionate tendency to copy the institutional value system (captured by an emergent institution) vs. a tendency not to conform to institutional values;

(2) Deeply-entrenched beliefs vs. weak beliefs; and

(3) A preference for content vs. a preference for variants that are more frequent in memory.

We operationalise these oppositions in terms of mathematical complements. The first opposition is characterised here as a bias to resist or conform to institutional values. The term *compliance* has been used in social influence studies to characterise the individuals' adherence to social norms, beliefs, acts and values in a broad sense; for a review see Cialdini and Goldstein (2004). A related factor used in social science is *authority bias*, but in this study we will use compliance because it does not imply that the institution is necessarily authoritarian. *Compliance bias* is also related to positive *frequency-based biases*, such as *conformist bias*—'follow the majority' (Acerbi et al., 2016; Asch, 1956; T. J. Morgan, Acerbi, & Van Leeuwen, 2019; T. J. H. Morgan & Laland, 2012). In scenarios where our agents share cultural values to a high extent, *compliance* is very similar to *conformity*; however, when different sub-populations have different value systems, the analogy does not apply.

We also examine the tendency to adhere to one's personal beliefs, which play an important role in the dissemination of culture (Del Vicario et al., 2017; Quattrociocchi et al., 2016). A value system can be thought of as a set of personal beliefs or hypotheses that affect behaviour. It can be operationalised as a set of values that agents assign to each cultural variant they encounter throughout their life and that they are capable to recall. When agents have deeply-entrenched beliefs they develop a tendency to recall information in a way that confirms their personal value system (henceforth *confirmation bias*), as opposed to agents that are more prone to consider new information to update their value systems. In this study, we operationalise *confirmation bias* as a tendency to process new information in a way that confirms one's prior beliefs (Nickerson, 1998). Therefore, a *non-confirmation bias* corresponds to informational influence, that is a tendency to copy and reproduce information from others (Centola & Baronchelli, 2015).

In a previous model (Chapter 4), we developed a model in which we simulated micro-societies of agents in which we systematically manipulated the degree of preference for variants value (*content bias*) and for the most frequent variants in memory (*frequency bias in memory*), a number of distributions of value systems in the population and the degree of institutional reinforcement of value systems. However, institutions were fixed and hence they were not part of a co-evolutionary model of cumulative culture. In the present study, we develop a co-evolutionary model of institutions, value systems and choice that integrates a much richer variety of relevant cognitive biases and social influence processes. We aim to establish the effect of these variables on the evolution of cultural diversity and the emergence of shared cultural conventions.

5.2 Methods

5.2.1 The model

We consider a simplified micro-society of agents, each of whom is characterized by a number of state variables as described in Table 5.1. The micro-society initially contains N agents, who pair-up and interact for a number of rounds (R). Each interaction consists of an exchange of cultural variants selected from an initial pool of variants (V) (Figure 5.1).

Agent pairings are scheduled using a method that takes the sequence of agents of the population after each round, shuffles the order of agents and then aggregates the agents into pairs, so that the order of scheduling was randomized at each time step. Each simulation begins with N agents, each initialised with a unique cultural variant and a value system. Agents' production of variants and their value systems evolve according to the model dynamics described in section 5.2.2. Population size was kept constant. Model runs proceeded in discrete time steps, called ”rounds”. For illustration purposes, a flowchart depicting relevant activity during one round is diagrammed in Figure 5.1.

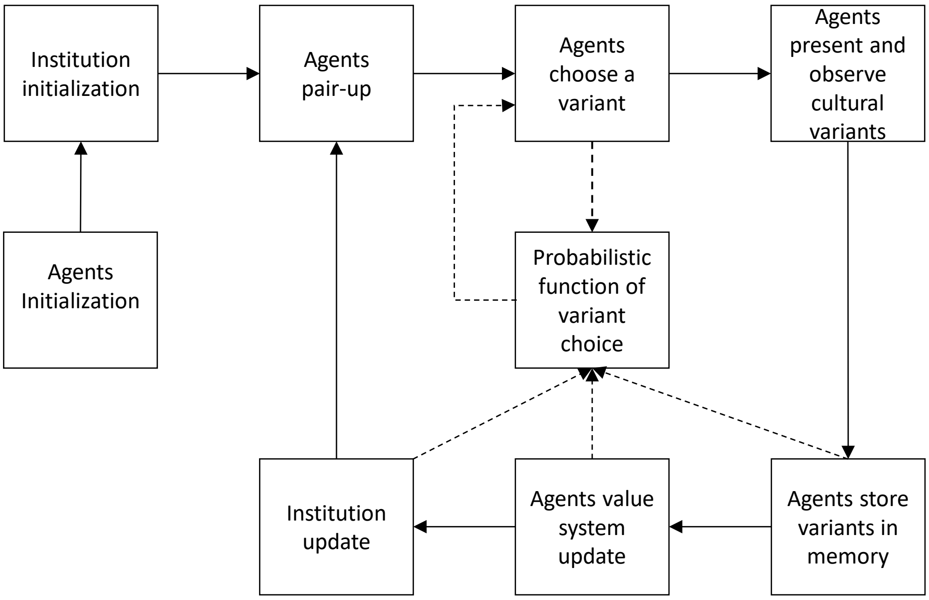


Figure 5.1: Flowchart depicting relevant activity during one round. Arrow direction represents the time-flow of events. Plain lines represent transition from one event to the following one. Dashed lines represent new data that is used to update agents prior information, affecting agents' variant choice over time.

5.2.2 Standard model dynamics

Let $X = \{x_1, x_2, \ldots, x_V\}$ be a vector of V cultural variants (each variant represents a different kind of property, taste or behavior of an agent regarding a particular trait (e.g., language, norm, religion, music choice, etc.) and takes its value from the combination of a range of parameters (see Equation 5.3)), and let $A = \{a_i, a_j, \ldots, a_N\}$ be the set of agents in a population. In the initial state each agent $n \in A$ is randomly assigned a cultural variant $v \in X$ selected from X with replacement, so that the model is not initialised with maximum diversity. Then, x_{nv0} is a random variant v assigned to an agent n at round 0 ($r = 0$). For example, at round 0, x_{iv0} is a random variant initially assigned to agent i, x_{jv0} is a random variant initially assigned to agent j, and so on.

At the beginning of each round r, agents are paired randomly. Once agents are paired, they interact by presenting and observing one cultural variant (Figure 5.2). Agents' choice of variant x_{nvr} to produce is sampled from its history (Figure 5.2, top) according to the probabilistic function defined in Equation 5.2. At this point, there is a small probability of innovation, in which case, the variant produced will be randomly sampled from one of the V variants the population was initialised with. Both agents add both the produced and observed variants to their memories (Figure 5.2, bottom). That is to say, at round r, when agent i and agent j interact, agent i produces variant x_{ivr} and agent j variant x_{jvr}. Thus, at each round r, the total set of produced variants X_r in the population is a vector of N (one per agent) cultural variants $(x_{ivr}, x_{jvr}, \ldots, x_{Nvr})$, where x_{ivr} is the variant produced by agent i, x_{jvr} is the variant produced by agent j, and so on.

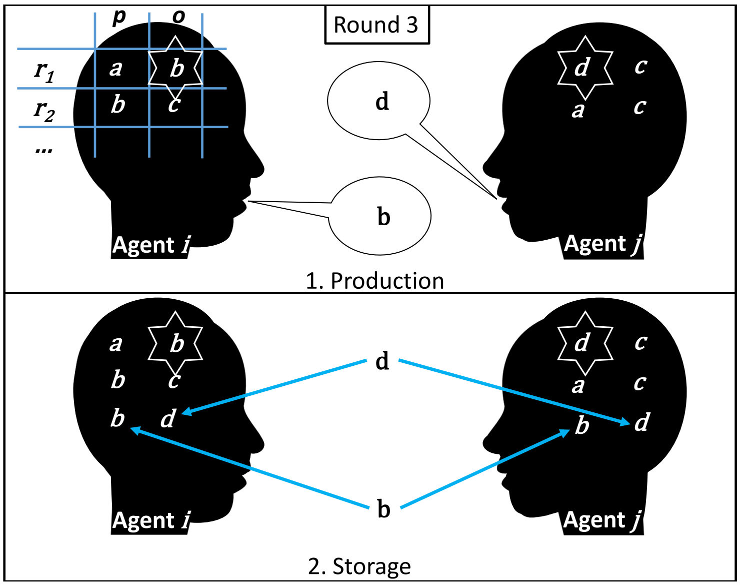


Figure 5.2: Illustration of the interaction between a pair of agents at one round of the simulation. At round 3, agents have already some variant tokens in memory, which they have stored in previous rounds. For each variant in memory, the letter represents the type (e.g. $a = x_1$, $b = x_2$,...). Index p represents variant tokens that were produced by self, while index o are variants that were observed by the agent (variant produced by other agent). Index r indicates the round at which the variant was stored. During the round depicted (round 3), first, one variant token in each agent's memory is selected for production (in the figure, the tokens surrounded by a star). Once both agents have produced a new token of their selected variant, they each proceed to store both in memory.

State variables and probability distribution of variants

The model takes several parameters as described below:

(a) Number of agents (N): We simulate micro populations of $N = 10$ and $N = 100$ agents. Each agent is initialised with a cultural variant randomly selected from a pool of V distinct cultural variants with replacement, where $V = N$ at the beginning of each simulation. We also assess the robustness of our findings with respect to the initial variant assignment by examining random assignment without replacement.

(b) Number of rounds (R): Model runs proceeded in 100 rounds r. At each round, the pairing is randomized in such a way that each agent pairs up with another agent ($N/2$ pairs are formed).

(c) Value system (S): Each agent n has a value system S, which is a vector of length V, which assigns a numerical value to each possible variant-choice. The state of the value system S of agent n at round r can be defined as a vector that contains one value for each V possible variant-choices, $(s_{n1r}, s_{n2r}, \ldots, s_{nVr})$, where s_{n1r} is the value assigned by agent

n to variant x_1 at round r, and s_{n2r} is the value assigned by agent n to variant x_2 at round r, and so on. Thus, s_{nvr} is the value assigned by a random agent n to a cultural variant x_v at round r. S assigns a number between 0 and 1 to each variant in V (that is, there is a preference on V represented by $s : V$). For example, if at round 0, S is $(1,0,\ldots,0)$, then, at round 0, agent n assigns value 1 to variant x_1, value 0 to variant x_2, and so on. We examine two conditions:

1. Pseudo-random (PR): In the initial state, each agent n is assigned a value system S so that the value of each variant is a random floating point number N such that $0 \leq N \leq 1$.
2. One takes all (OTA): Agents in the population are initialised with a hegemonic value system S so that the value assigned to one preferred variant is 1, and to the others 0.

(d) Institution (G): We assume that the micro-society is governed by a global institution G, which is a vector of length V. The state of the institution G at round r can be defined as a vector that contains one value for each V possible variant-choices, $(g_{1r}, g_{2r}, \ldots, g_{Vr})$, where g_{1r} is the value assigned by the institution to variant x_1 at round r, and g_{2r} is the value assigned by the institution to variant x_2 at round r, and so on. G assigns a numerical value between 0 and 1 to each possible variant-choice. For example, if at round 0, G is $(1,0.5,\ldots,0)$, then, institution G assigns value 1 to variant x_1, value 0.5 to variant x_2, and so on. At each round r, the institutional value g_{vr} assigned to each possible variant-choice x_{vr} is calculated as an arithmetic mean of the N values $s_{ivr}, s_{jvr}, \ldots, s_{Nvr}$, where s_{ivr} corresponds to the value assigned by agent i to variant x_{vr}, and s_{jvr} corresponds to the value assigned by agent j to variant x_{vr}, and so on. That is:

$$g_{vr} = \frac{1}{N} \sum_{nvr=1}^{N} s_{nvr} = \frac{s_{ivr} + s_{jvr} + \ldots + s_{Nvr}}{N} \tag{5.1}$$

(e) Institutional power (ε): It is a parameter that captures the capacity of the institution to effectively communicate its values to the agents. It takes values from 0 (null capacity) to 1 (full capacity).

(f) Compliance bias (κ): It is a parameter that identifies the agent's bias to conform to institutional values. Its complement (a non-compliance bias κ') identifies a preference for the agents' own choice and value system, which can be (or not) different than institutional preferences. It takes values from 0 (null compliance) to 1 (full compliance).

(g) Confirmation bias (γ): It captures the tendency of agents to give stronger weight to their prior beliefs or hypotheses, which are encoded in the current value system of each agent. Thus, γ identifies the tendency of an agent n to assign value to a cultural variant x_v according to its current value system s_{nvr}. Its complement γ' identifies a tendency of agent n to assign value to a cultural variant x_v according to its current choice, regardless of the current value system s_{nvr}. That is, when γ is low, agents that are more prone to consider new information to update their value systems; γ takes values from 0 to 1.

(h) Content bias (β) encompasses two parameters (b, d). Parameter b is the agents' sensitivity to variant value (s), and ranges from 0—not sensitive at all—to 1—fully sensitive—in steps of 0.1. Parameter d specifies whether the variant is in the agent's memory record, that is to say, whether the variant has been produced or seen at least once; d is 1 if the variant is in memory, and 0 otherwise—in other words, one cannot prefer to re-produce a variant due to its quality until one has been exposed to that particular representation. Parameter β is equal to $b \times d$. Thus, content bias (β) assigns a value from 0 to 1 to each variant. Note that, at each round, as explained in the model dynamics, each agent stores two variants in its memory (the one that was produced by the agent itself and the one that was produced by its partner). Thus, content bias does not affect the storage process in our model (agents store in memory all the variants they encounter), it only affects the agents' production. In simple words, the higher the value of β, the higher the probability of producing a cultural variant due to its quality. When content bias is 0, we have a neutral content model. Content bias values from 0 to 1 in steps of 0.1 are examined.

(i) Frequency bias (β'): It is a parameter identifying an agent's preference for variants that are more frequent in its history. It corresponds to the complement of β. This parameter is limited by the agents' memory size m, that is, the maximum amount of history (in rounds) that can influence the variant choice. Each variant found in the history is a variant that was either produced or observed by the agent. A memory size of m means that the model remembers the last M variants $h \mid M,m$ from the history h. Given an agent n and a variant x_v, then $f(x_{nv} \mid h_{nr|M,m})$

is the relative frequency of variant x_v in the memory of an agent n by round r. Thus, $f(x_{nv} \mid h_{nr|M,3})$ corresponds to the relative frequency of variant x_v in an agent memory for the last 3 rounds. Memory size between 2 and 4 rounds was associated with better model fit in Tamariz et al. (2014).

(j) Innovation rate (μ): Agents can generate novel variants. We fix the innovation probability at 0.02 by using a flat distribution $\phi(X)$ weighted by innovation rate, which means that 98% of variant choices would reflect the combined distribution (probability distribution yielded by all the parameters described above), while 2% would be a random choice among all V initial variants in X. The probability level selected reflects the innovation rate found in Fay et al. (2010) experimental data by Tamariz et al. (2014).

Model parameters				
Entity	Parameter	Symbol	Number of levels	Value(s)
Agent	Content bias	β	11	0.0 to 1.0 in steps of 0.1
	Confirmation bias	γ	11	0.0 to 1.0 in steps of 0.1
	Compliance bias	κ	11	0.0 to 1.0 in steps of 0.1
	Memory	m	1	3
	Innovation	μ	1	0.02
	Value system of an agent n	$S = \{s_{n1r,\ldots,s_{nVr}}\}$		
	Variant value (assigned by an agent)	s		Eq.5.3
	Agents' sensitivity to variant value s	b	2	[0,1]
	Variant in agent's memory record	d	2	[0,1]
	Agent' history	h		
	Flat distribution of variants	$\phi(X)$		
	Typical variant	x_v		
	Typical agent	n		
	Round	r		
Global	Initial set of agents	$A = \{a_i, \ldots, a_N\}$		
	Initial number of agents per micro-society	N	3	10,50,100
	Initial vector of cultural variants	$X = \{x_1, \ldots, x_V\}$		
	Initial number of variants per population	V	3	10,50,100
	Number of rounds	R	1	100
	Number of games per round	$N/2$	3	5,25,50
	Institution	$G = \{g_{1r}, \ldots, g_{Vr}\}$		
	Institutional value (assigned by G)	g		Eq.5.1
	Institutional power	ε	11	0.0 to 1.0 in steps of 0.1

Table 5.1: Parameters, state variables and scales.

For each round in the simulation, for each agent, the model yielded a probability distribution of variants (x) for a given history (h) of previous rounds, according to the following equation. We use the apostrophe (′) to denote the probabilistic complement: $a' = 1 - a$.

$$\Pr(x_{nvr} \mid h_{nr}) = \mu'\beta' f(x_{nv} \mid h_{nr|M,3}) + \mu'\beta s_{nvr} + \mu\phi(X) \tag{5.2}$$

where $\Pr(x_{nvr} \mid h_{nr})$ corresponds to the probability that an agent n produces variant v at round r given the specific history of agent n by round r.

We run each simulation with 14641 different parameter value combinations. For each parameter combination, we ran the simulation 200 times. For each model run we assume that all agents have the same connectivity dynamic, biases and memory sizes. The results below show the average and standard deviations of the number of runs of each parameter combination examined. All parameters and state variables can be found in Table 5.1.

The co-evolution of variant-choice, value system and institutions

The model is initialised with each agent having a cultural variant and a value system (Section 5.2.1). At each round, the value assigned by each agent n to variant-choice x_v is updated according to the following equation, where, as already noted, the apostrophe ($'$) denotes the probabilistic complement: $a' = 1 - a$:

$$s_{nvr+1} = g_{vr}\varepsilon\kappa + (x_{vr}\gamma' + s_{nvr}\gamma)\kappa' \tag{5.3}$$

where s_{nvr+1} is the value assigned by an agent n to a variant x_v in the next round $r+1$, g_{vr} is the institutional value assigned by an institution to a variant x_{vr}, ε stands for *institutional power*, κ for *compliance bias*, γ is a parameter that captures the relative weight of the current value system s_{nvr} in its future value system s_{nvr+1}, in a way that affirms the agent's prior value system or hypothesis (or *confirmation bias*), and x_{vr} represents the target variant, which takes value 1 if it has been produced in the current round, and 0 otherwise. Since one agent might eventually produce cultural variants which do not match its current value system but which inform its future value system, γ' captures the relative weight of current choices in the future variant value s_{nvr+1} at round $r+1$, regardless of the current variant value s_{nvr}. For illustration purposes, Figure 5.3 shows some characteristic parameter combinations that satisfy three different solutions for Equation 5.3.

As explained above, by implementing this algorithm our model is able to update agent's value systems from round to round. In turn, the form of the institutional value system is also updated at the beginning of each round. The co-evolution of institutional value systems and individual value systems model the value s assigned by each agent to each possible variant over time.

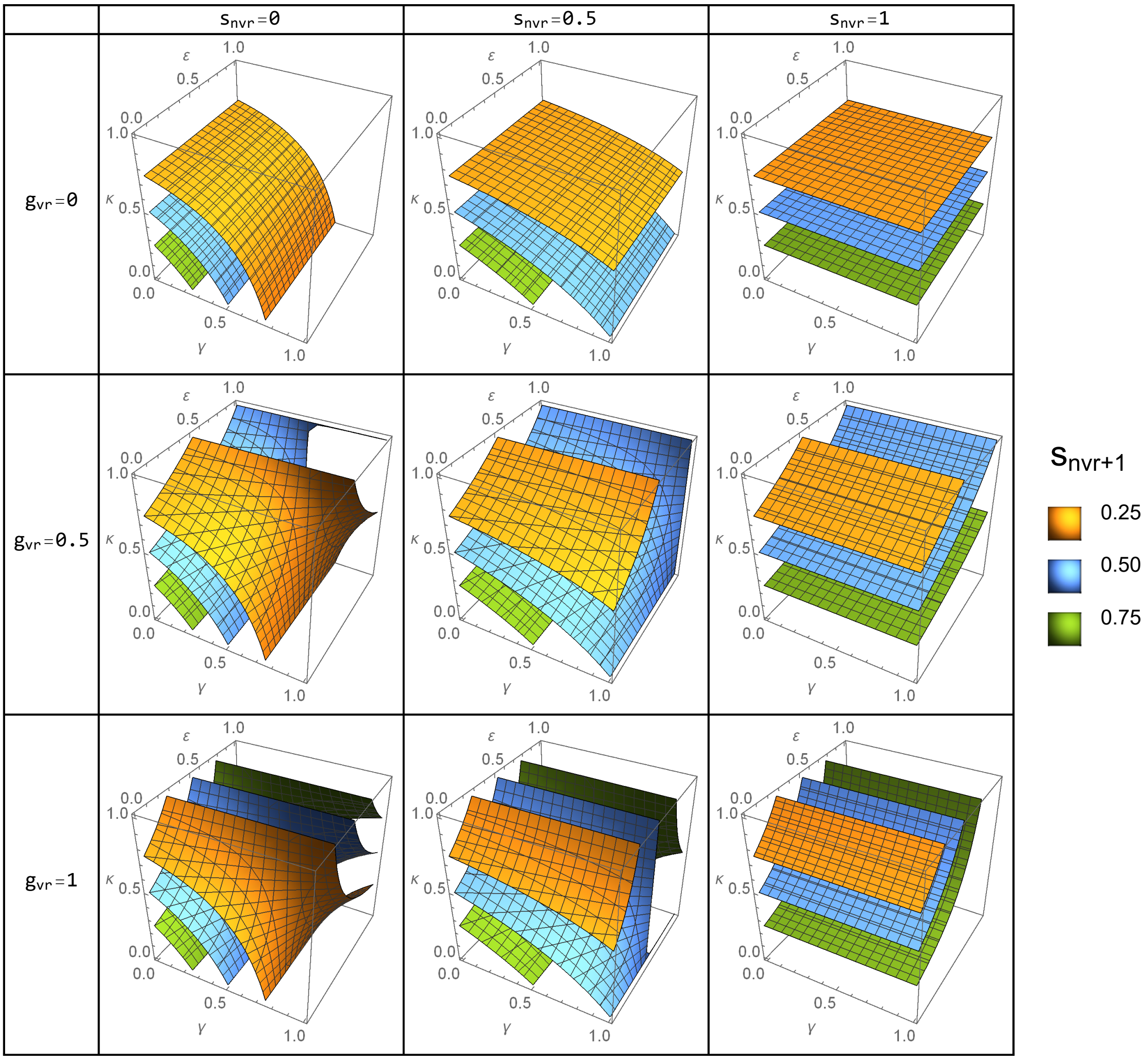


Figure 5.3: Illustration of some parameter combinations in the 3D space. Given that an agent n produces variant x_v at round r, plots show all possible combinations of values of institutional power (ε), compliance bias (κ) and confirmation bias (γ) that satisfy $s_{nvr+1} =$ 0.25, 0.5 and 0.75, according to Equation 5.3, for three different levels of g_{vr} and s_{nvr}. For example, if at round r a variant x_v is assigned value $g_{vr} = 0$ by an institution and value $s_{nvr} = 1$ by an agent n with compliance bias $\kappa = 0.5$, that means that the value assigned by the agent to variant x_v at round $r+1$ will be $s_{nvr+1} = 0.50$. Lower compliance bias (κ) tends to contribute more with variant value when institutional values (g) are weak (see top right). When institutional values are stronger and agent' values (s) are weak, confirmation bias and institutional power tend to invert the pattern of contribution to variant value (see bottom left).

Quantifying cultural diversity

In the present chapter, we are interested in the diversity of agents' produced variants and in the diversity of institutional values at each time step. As in previous chapters, we use Shannon's entropy to quantify diversity. In order to facilitate the comparison with other metrics and to eliminate the effect of different population sizes and time series, we normalise entropy in the same way as in Chapter 3 by $\log_2 n$ to obtain $H_n(V) \in [0,1]$:

$$H_n(V) = -\sum_{v_i \varepsilon V} \frac{p(v_i)\log_2 p(v_i)}{\log_2 n} \tag{5.4}$$

where V corresponds to the set of variants, and $p(v_i)$ is the probability of i^{th} variant in that set, and n is the number of variants. High entropy corresponds to high diversity and also to low convergence on shared conventions.

We also use a Simpson's index. However, unlike in Chapter 4, instead of taken the Simpson's reciprocal index, we use the Simpson's diversity index in order to facilitate the comparison with normalise entropy when testing robustness. Simpson's diversity index (D) is commonly used to measure the degree of dominance of variants in the community (Simpson, 1949). It is a simple mathematical measure that characterizes diversity in a data set. The proportion of variants relative to the total number of variants (p_i) is calculated and squared. The squared proportions for all the variants are summed, and the complement is taken:

$$D = 1 - \sum_{i=1}^{R} p_i^2 \tag{5.5}$$

where R corresponds to the total number of existing variants in the system (richness). D ranges from 0 (one variant dominates the cultural system completely) to 1 (all variants are equally present).

5.3 Results

5.3.1 Standard model: Initial random assignment of variants with replacement (wR)

We first considered a model in which agents were initially assigned a cultural variant selected from a pool of variants with replacement. Simulation outcomes show that the co-evolutionary processes of value systems, cognitive biases and institutions implemented in the model tend to stabilize the diversity of the produced variants over time.

Figures 5.4 and 5.5 show the evolution of the cultural diversity of the set of produced variants at each round for a selection of representative parameter combinations of institutional power (ε), compliance bias (κ), confirmation bias (γ) and content bias (β). Figures 5.6 and 5.7 show the evolution of the diversity of institutional values over time.

In general, ε facilitates the emergence of shared cultural conventions when $\kappa > 0$. This effect is amplified or minimised depending on how diverse institutions are at each time step. When $\varepsilon = 1$, institutions have an immense capacity to convey their values effectively to the agents. In this scenario, when institutions are less diverse, convergence on shared cultural conventions is faster and diversity reduction stronger. This can be observed when we compare results of simulations using PR, where each agent was initialised with a randomised value system, against simulations using OTA, where agents where initialised with an homogeneous and hegemonic value system. As figures 5.4 and 5.5 show, diversity of produced variants is lower under OTA, and this is due to the reinforcing role of institutions, which push diversity up (PR) or down (OTA) according to how diverse their own value systems (the ones that they convey to the society) are. In general, content bias (β) amplifies the effect of any parameter combination that allows the emergence of asymmetries in the dispersion of values among the set of existing cultural variants.

Interestingly, under *PR* and $\varepsilon = 1$, fully conformist populations ($\kappa = 1$) yield lower levels of convergence than populations with intermediate levels of κ. This result seems counterintuitive, but it is explained by the intrinsic properties of the institution: the emergent institution in this scenario is highly diverse and also has capacity to influence the agents' choices, resulting in a relatively high diversity of produced variants for all levels of content bias β. In other words, a compliance bias pushes diversity up when institutions are diverse, and pushes diversity down when institutions convey value systems with low diversity and strong dominance of one or few cultural variants.

Intermediate values of ε drastically reduce the formation of cultural conventions when compared to high levels of ε, and this effect increases with κ. Interestingly, a decrease in ε increases diversity in both scenarios, PR and OTA, regardless of the diversity of institutional values promoted by the institution. Indeed, we can expect diverse institutions, like those that emerge in PR, to boost diversity. But, how is it possible that institutions in OTA, which convey extremely unambiguous and hegemonic value systems, also increase diversity. The answer lies in the strength with which these values, whether diverse or not, are transmitted to the population by the institutions. When $\varepsilon = 0.5$ and $\kappa = 1$, institutions are just partially successful conveying their values. This leads to a weaker transmission of institutional values over time (i.e. on average, the institutional value assigned to each existing cultural variant loses value at each time step). This combined with a population of agents willing to nevertheless keep adopting institutional values, ends up weakening agents' value systems too. The result is a society where values fade over time, leading to a limit in which the change in the frequency of an existing variant in the population is due to random sampling of variants. However, when $\varepsilon = 0.5$ and agents do not fully conform, $\kappa < 1$, alternative options for the emergence of cultural conventions arise. This is the case when $\kappa = 0.5$ and $\gamma < 1$: In this scenario, cultural conventions can emerge from the mutual reinforcement of local interactions and institutional values. This is because as agents cease to fully conform with the social norms promoted by the institution, they begin to assign value to variants according to what is produced in their local interactions, which produces a stable convergence equilibria according to β. In a relatively wide range of these intermediate situations the model can be thought of as the most realistic example of most human communities, where neither are institutions fully powerful to direct the agents' choices, nor are agents fully conformist or fully non-conformist.

When $\varepsilon = 0$, institutions have no capacity to convey their values to the agents, so agents end up converging due to local interactions or not converging at all. For example, agents can attain moderately to high levels of convergence on cultural conventions without any institutional intervention when ε is 0 and γ is very high. In this scenario, agents converge due to their capacity to coordinate in their local interactions. On the other hand, a reduction in γ implies that the agents' value systems become more volatile and less dependent on their individual experiences, resulting in a decrease in convergence.

An unusual simulated case is when agents are fully conformist, $\kappa = 1$, and institutions have no power at all. In this scenario, the system tends to high diversity of produced variants because variant selection ends up being similar to a random choice among variants with null value. A model that can be thought of as a society of believers drifting around and looking for something to believe in. On the other extreme, when $\kappa = 0$ and $\gamma = 1$, we have a society where agents do not conform at all but are fully biased towards their prior beliefs. In this case, convergence on shared conventions is dependent on the degree of initial similarity between the agents' value systems.

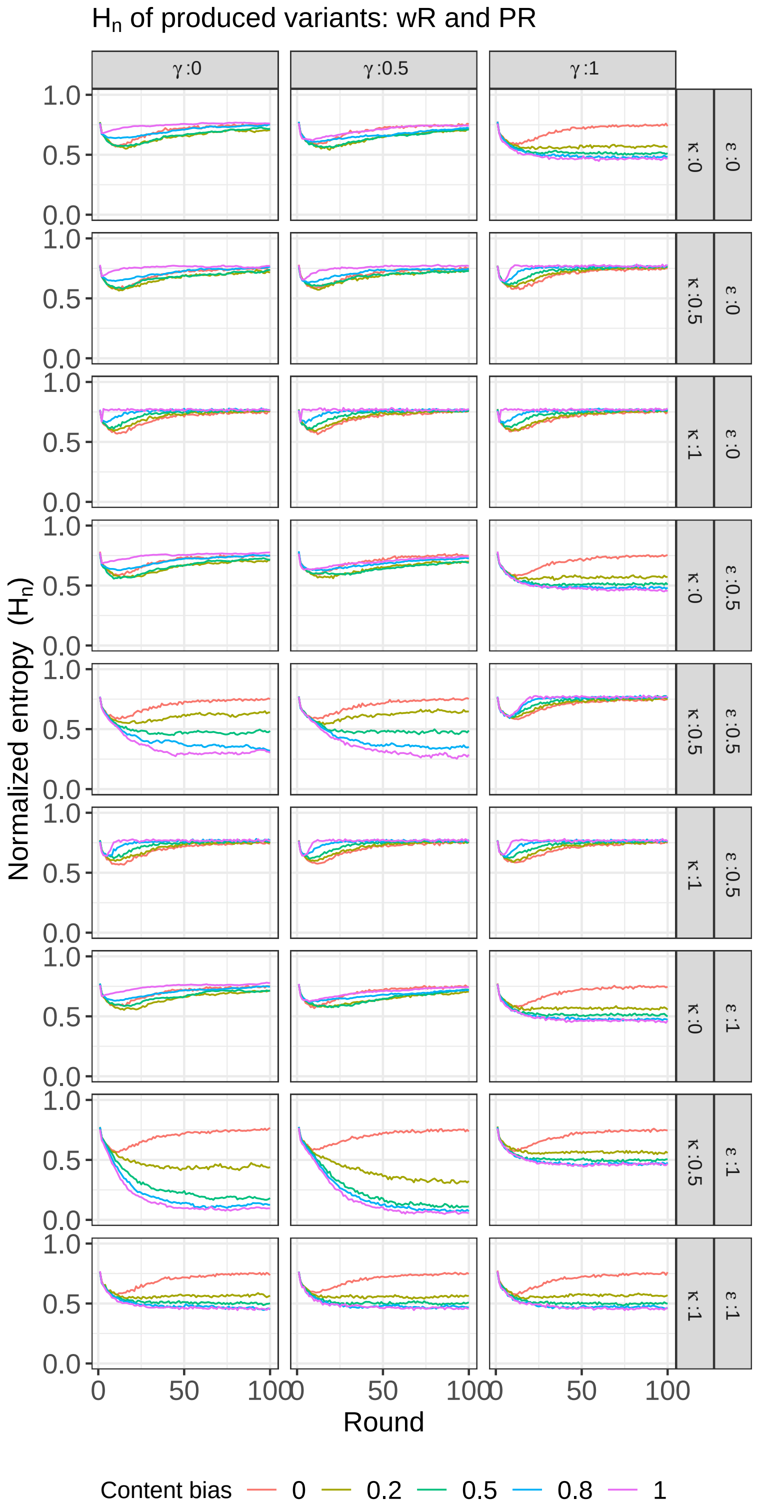


Figure 5.4: Cultural diversity (measured as Normalised Shannon Entropy H_n) of the set of produced variants over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X with replacement (wR) and initial randomised value system (PR).

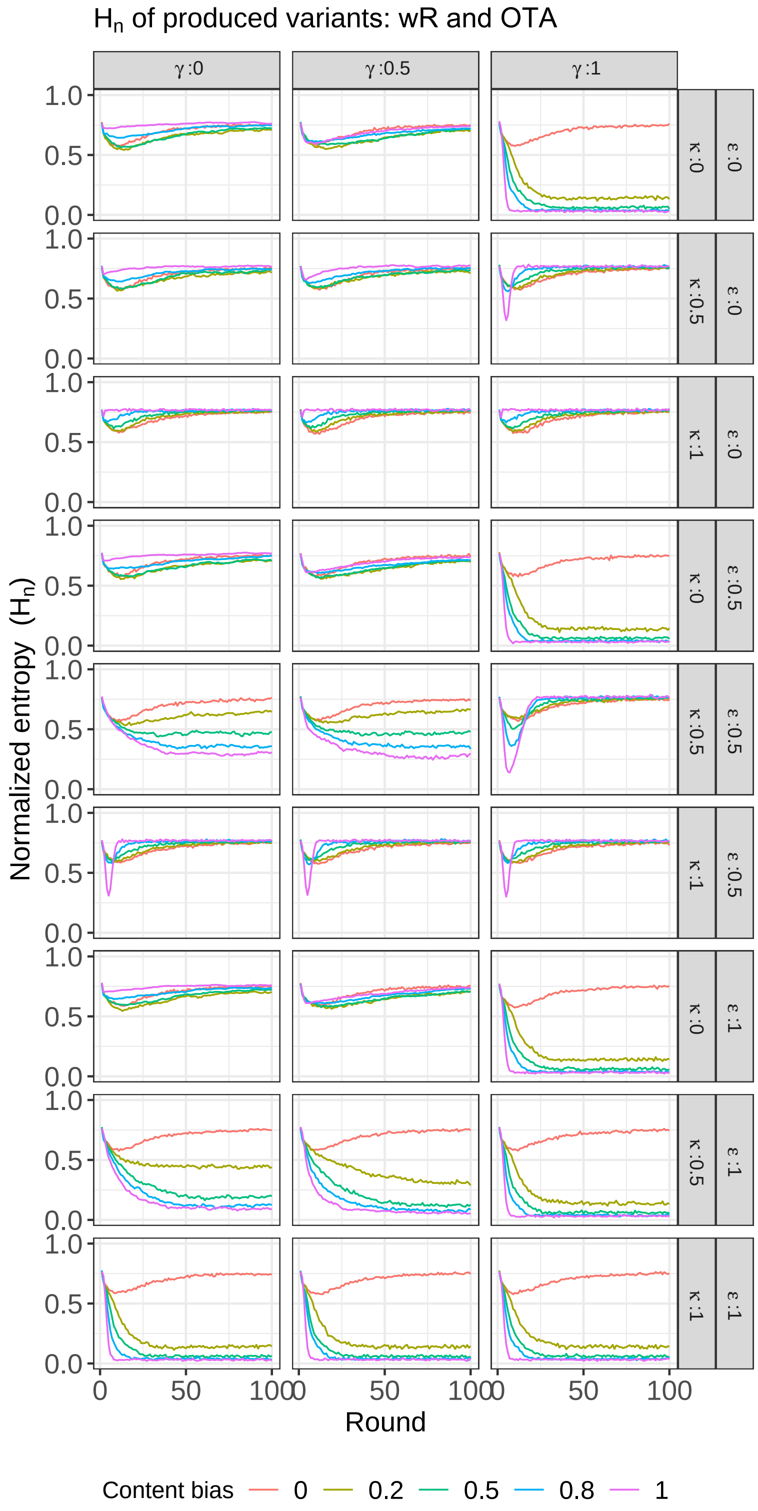


Figure 5.5: Cultural diversity (measured as Normalised Shannon Entropy H_n) of the set of produced variants over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X with replacement (wR) and initial randomised value system (OTA).

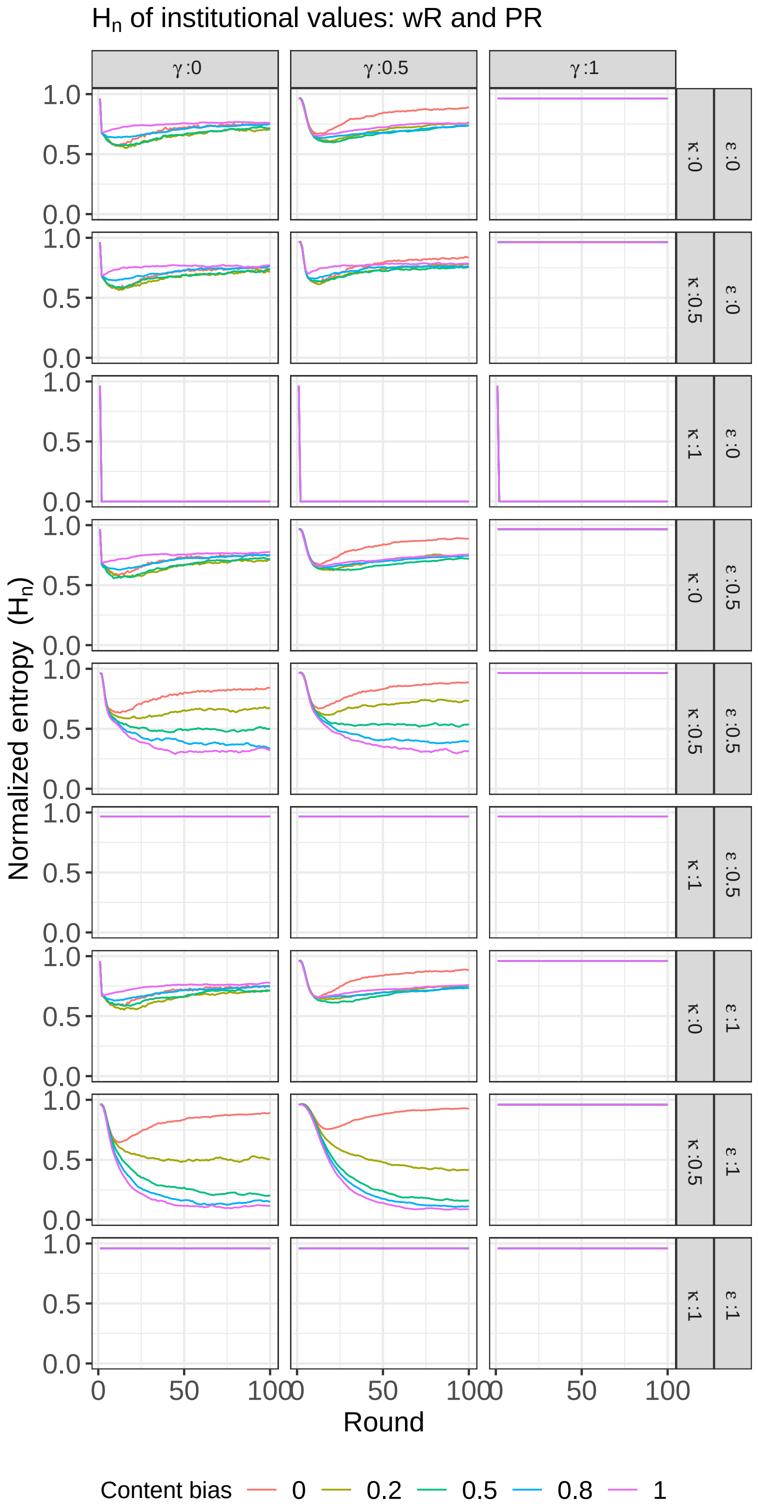


Figure 5.6: Diversity (measured as Normalised Shannon Entropy H_n) of institutional values over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X with replacement (wR) and initial randomised value system (PR).

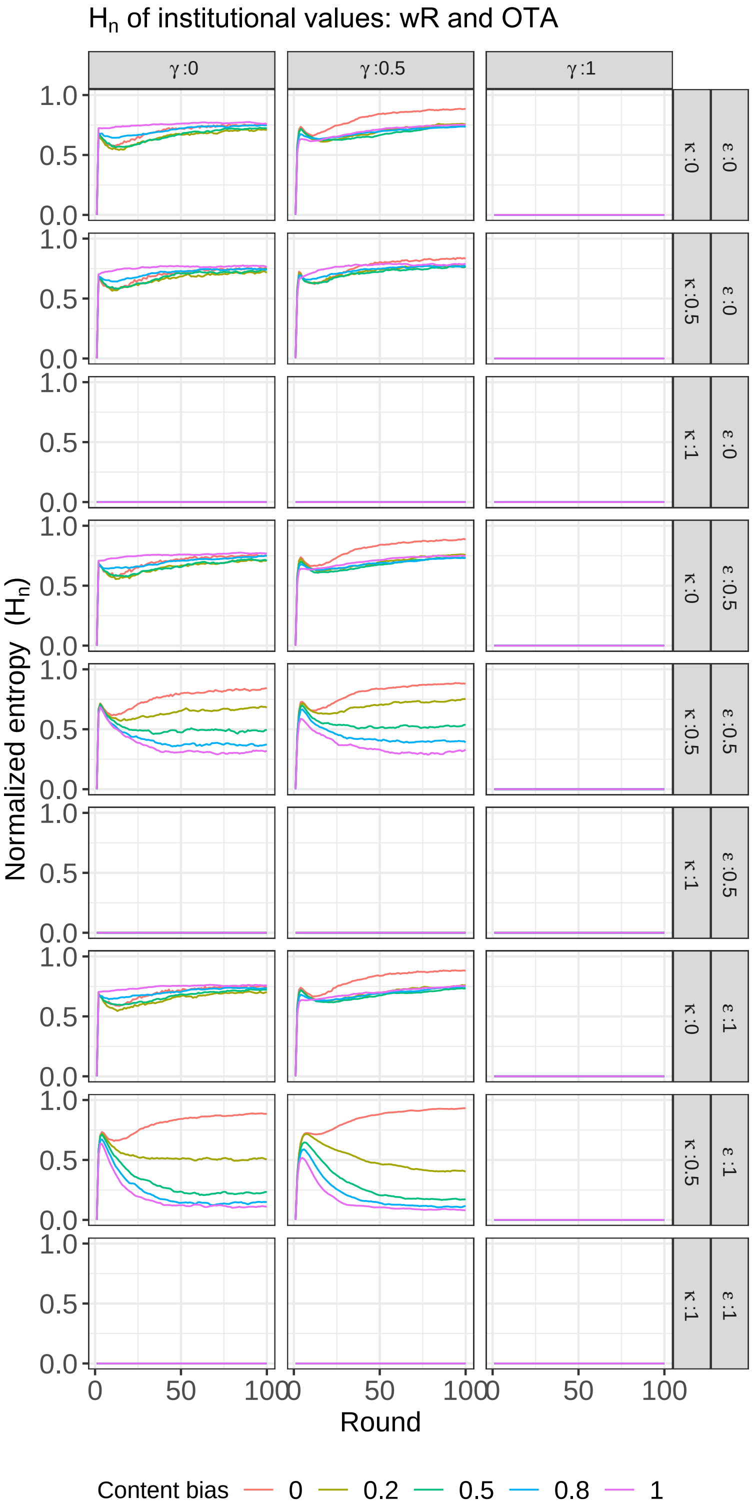


Figure 5.7: Diversity (measured as Normalised Shannon Entropy H_n) of institutional values over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X with replacement (wR) and initial randomised value system (OTA).

5.3.2 Assumptions concerning initial variant assignment: Assignment without replacement

Our model makes an important assumption about the initial state of diversity of cultural variants in the society, which depends on the method used to assign variants to agents at time 0. In the simulations examined so far, each agent is initialised with a cultural variant randomly selected from a pool of V distinct variants with replacement. This means that agents might share some variants in the initial state, resulting in a reduction of entropy. On average, this method yielded an initial diversity of around 0.75 (measured as Normalised Shannon Entropy, H_n). While we have shown that the co-evolution of value systems and institutions affect cultural diversity in different types of populations according to particular combinations of cognitive biases, we aim to asses the robustness of these findings and to what extent the model remains unchanged under different initial conditions of diversity. When agents are initialised with a cultural variant selected form a pool of variants without replacement, they start the simulation with a unique cultural variant, which means that diversity is the highest at time 0 (H_n=1). However, in both conditions (with and without replacement), convergence on shared cultural conventions remains unchanged. Although diversity is initially lower when agents share variants from the beginning, mean diversity at equilibrium was the same in both conditions and remained unchanged in response to diversity perturbations, which implies that the system behaves consistently against these different assumptions; compare Figures 5.4 and 5.5 with Figures 5.8 and 5.9. The same consistency can be observed in regard to diversity of institutional values; compare Figures 5.6 and 5.7 with Figures 5.10 and 5.11).

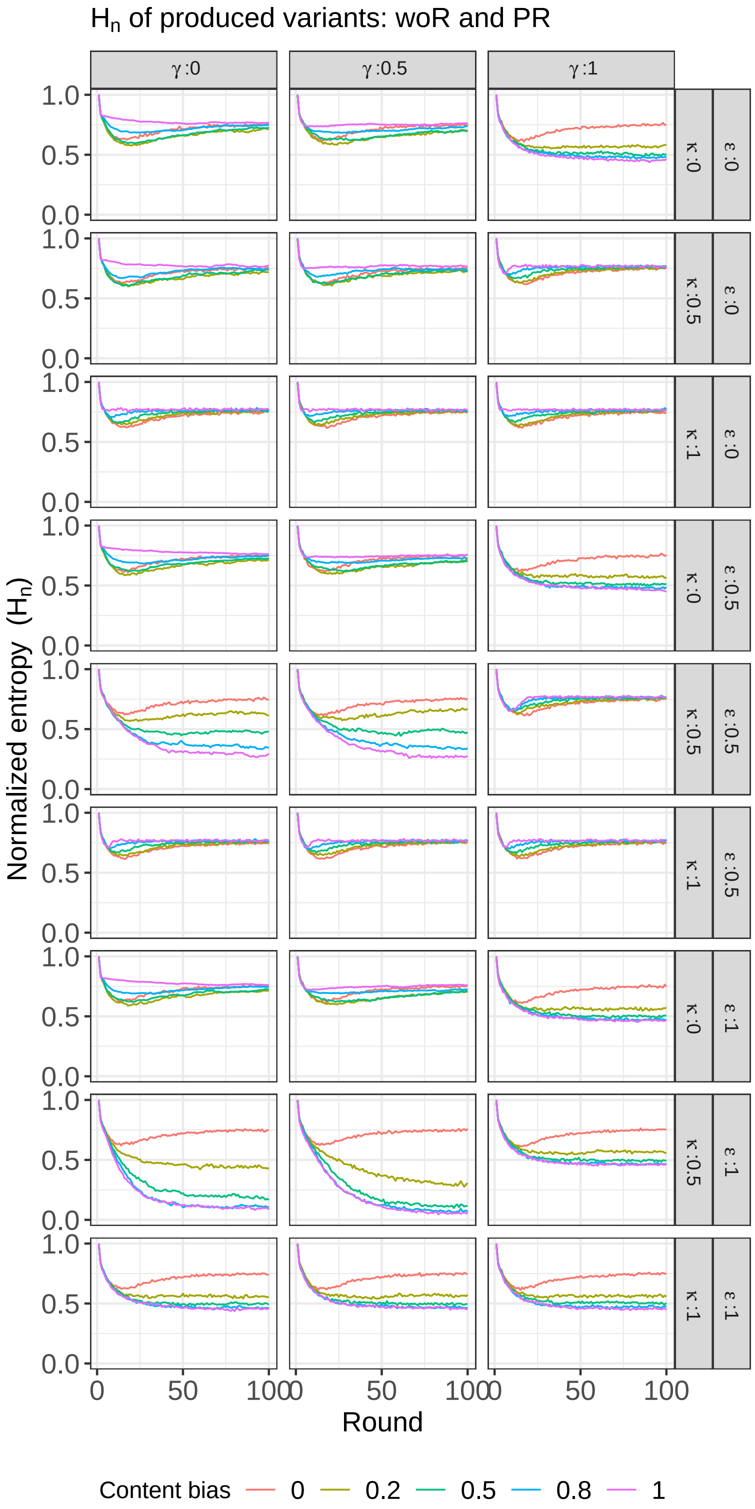


Figure 5.8: Cultural diversity (measured as Normalised Shannon Entropy H_n) of the set of produced variants over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X without replacement (woR) and initial randomised value system (PR).

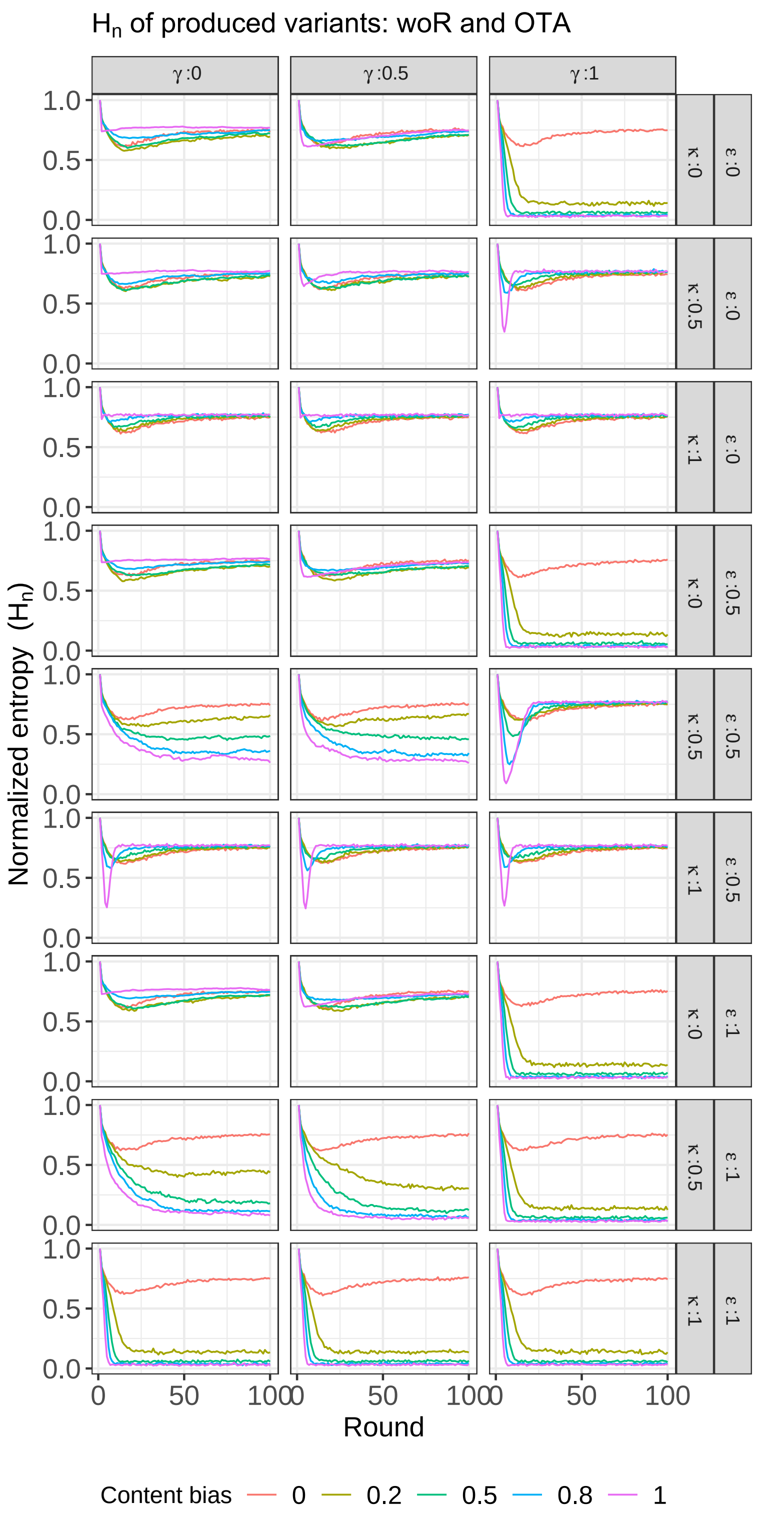


Figure 5.9: Cultural diversity (measured as Normalised Shannon Entropy H_n) of the set of produced variants over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X without replacement (woR) and initial randomised value system (OTA).

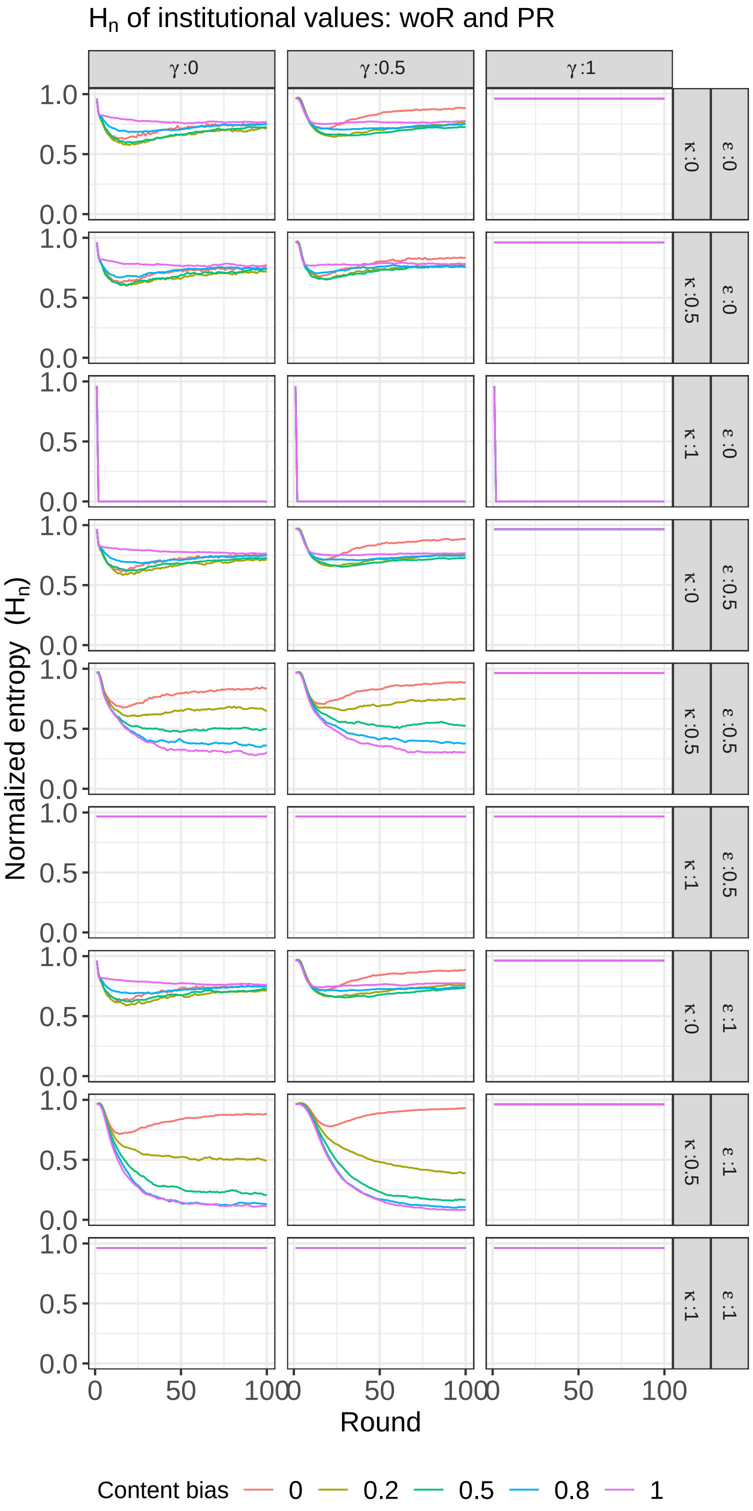


Figure 5.10: Diversity (measured as Normalised Shannon Entropy H_n) of institutional values over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X without replacement (woR) and initial randomised value system (PR).

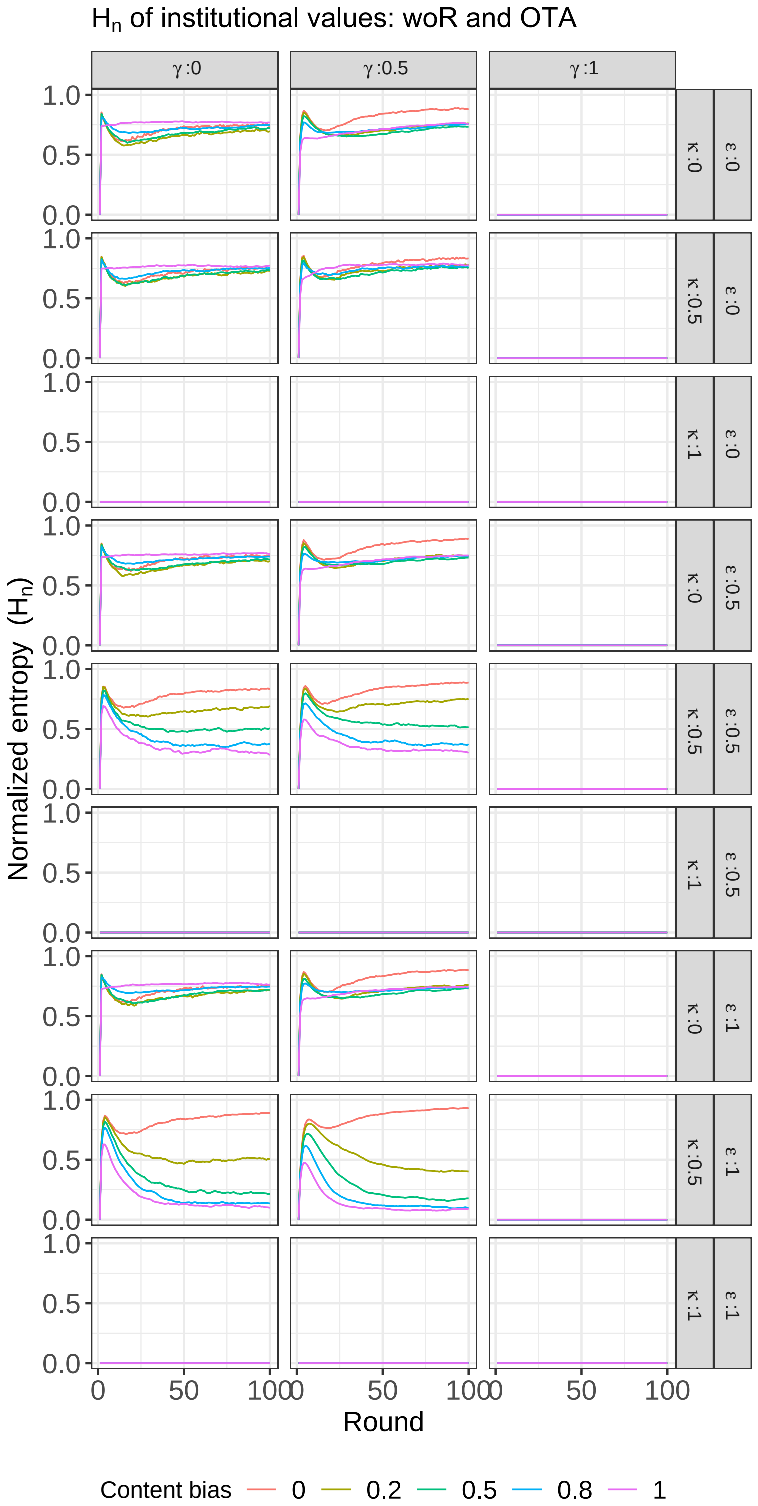


Figure 5.11: Diversity (measured as Normalised Shannon Entropy H_n) of institutional values over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X without replacement (woR) and initial randomised value system (OTA).

5.3.3 Population size

When population size increases, the emergence of shared cultural conventions is delayed, but the model yields qualitatively similar results for the population sizes examined. Figure 5.12 shows that the relative diversity of cultural variants remains higher in bigger populations until they eventually find their equilibrium. After 100 rounds, diversity outcomes are similar in both conditions. This is particularly true when content bias is greater than 0 ($\beta > 0$), because it allows agents to be sensitive towards variants value and, in turn, find similar equilibria under conditions of small and big population sizes. In contrast, for neutral content models ($\beta = 0$) differences between conditions are greater, with smaller populations reaching higher convergence. This is simply because the smaller the population, the higher the probability of sharing a variant by chance. Interestingly, in 100 agent micro-societies, when institutional power is limited (intermediate values of ε), time to equilibrium is longer. This is due to the combined effect of interacting forces that simultaneously push diversity up and down, which results in a complex system that delays the formation of conventions even longer. In the present example, institutions are diverse and weak (which pushes diversity up) and agents individual biases are moderate (which slightly pushes diversity down due to alignment in the local interactions). In other words, in bigger populations, intermediate levels of institutional power combined with intermediate levels of individual biases produce complex dynamics of antagonistic forces which translate into more time to reach consensus. This result implies that population size has an effect on model outcome. However, model results are still very similar qualitatively, which means that they hold for the range of populations sizes considered under the aforementioned assumptions.

5.3.4 Alternative alpha diversity indexes

A possible concern with the results of our study is that different diversity indexes might yield opposite trends. So far, we have used Shannon entropy (H_n) as a measure of diversity. Richness (i.e. simply a count of variant types) and Simpson's diversity index, are also widely used in ecology studies. While the Shannon index stresses the richness component and the unpredictability of variant types, the Simpson index gives more weight to the evenness component and the degree of dominance of variant types. These differential emphases on different aspects of diversity might give rise to the possibility of non-conformance of diversity indexes, as has been

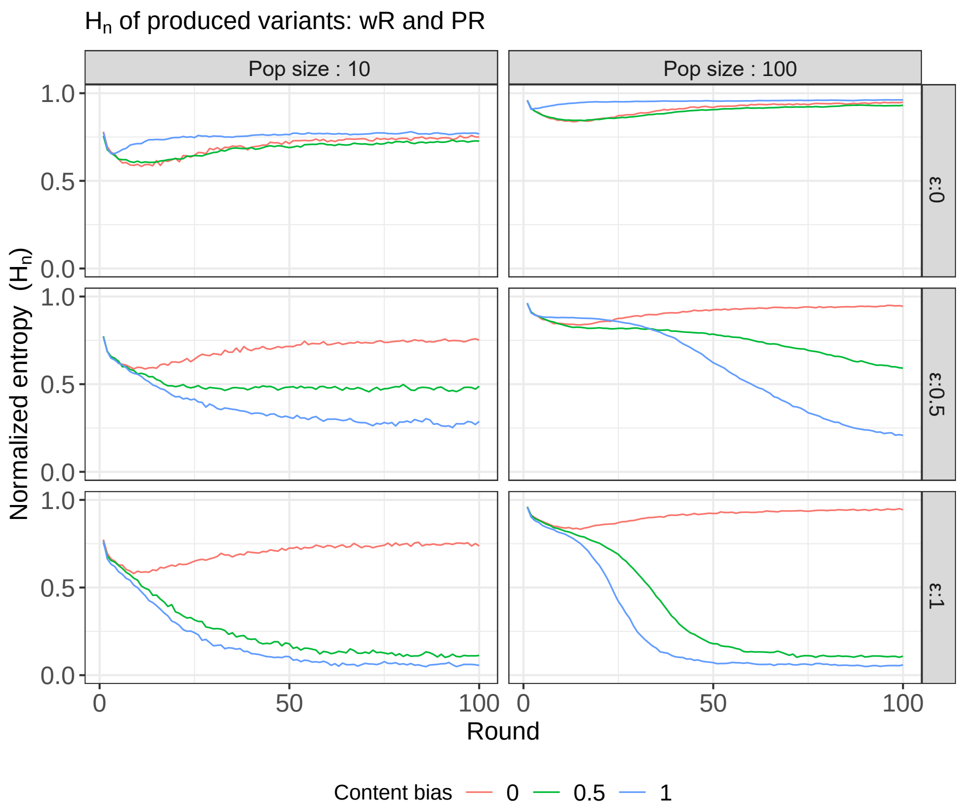


Figure 5.12: Cultural diversity (measured as Normalised Shannon Entropy, H_n) of the set of produced variants over time in 10 and 100 agents micro-societies. Results of three representative scenarios of *institutional power* (ε). Each scenario shows three different levels of *content bias*. Other cognitive biases were fixed at 0.5. Simulations were initialised with a random assignment of variants selected from X with replacement (wR) and a randomised value system (PR).

demonstrated in the past (Hurlbert, 1971; Nagendra, 2002).

In order to address this potential issue we also computed on Richness (R) and Simpson's diversity index (D) for a number of representative scenarios using the standard model (Figures 5.13 and 5.14). Results using these alternative alpha diversity indexes are fully consistent with our previous analyses. Model outcomes do not show opposite trends in response for the Shannon, Simpson and Richness indexes. Nevertheless, these additional analyses add relevant information about the evolution of the composition of the pool of cultural variants at each time step. Differences between highest and lowest $R(\simeq 0.5)$ are smaller than differences in $H_n(\simeq 0.68)$ (e.g. see R and H_n when $\varepsilon = 1$, $\kappa = 0.5$ and $\gamma = 0.5$), which suggests intense competition among the surviving cultural variants over time, along with a contained extinction probability in most of the scenarios examined. Simpson's D also confirms our previous analyses. Parameter combinations that facilitate the emergence of social conventions (measured as a decrease in H_n) coincide with those where there are few dominant variants at equilibrium (low D). Although with several important exceptions we have already explained, this happens mostly for moderate and high values of ε, κ and β.

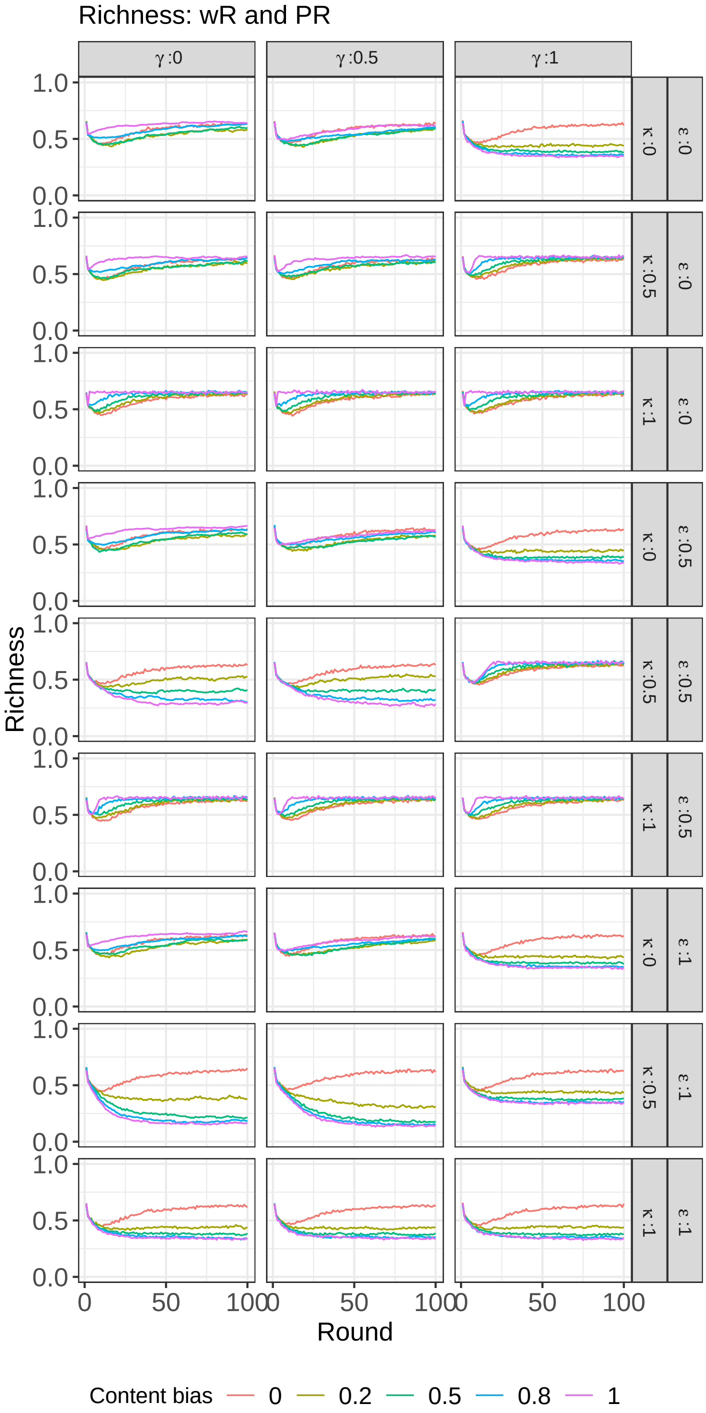


Figure 5.13: Richness (measured as the number of variants -operational taxonomic units- at each round) over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X without replacement (wR) and initial randomised value system (PR).

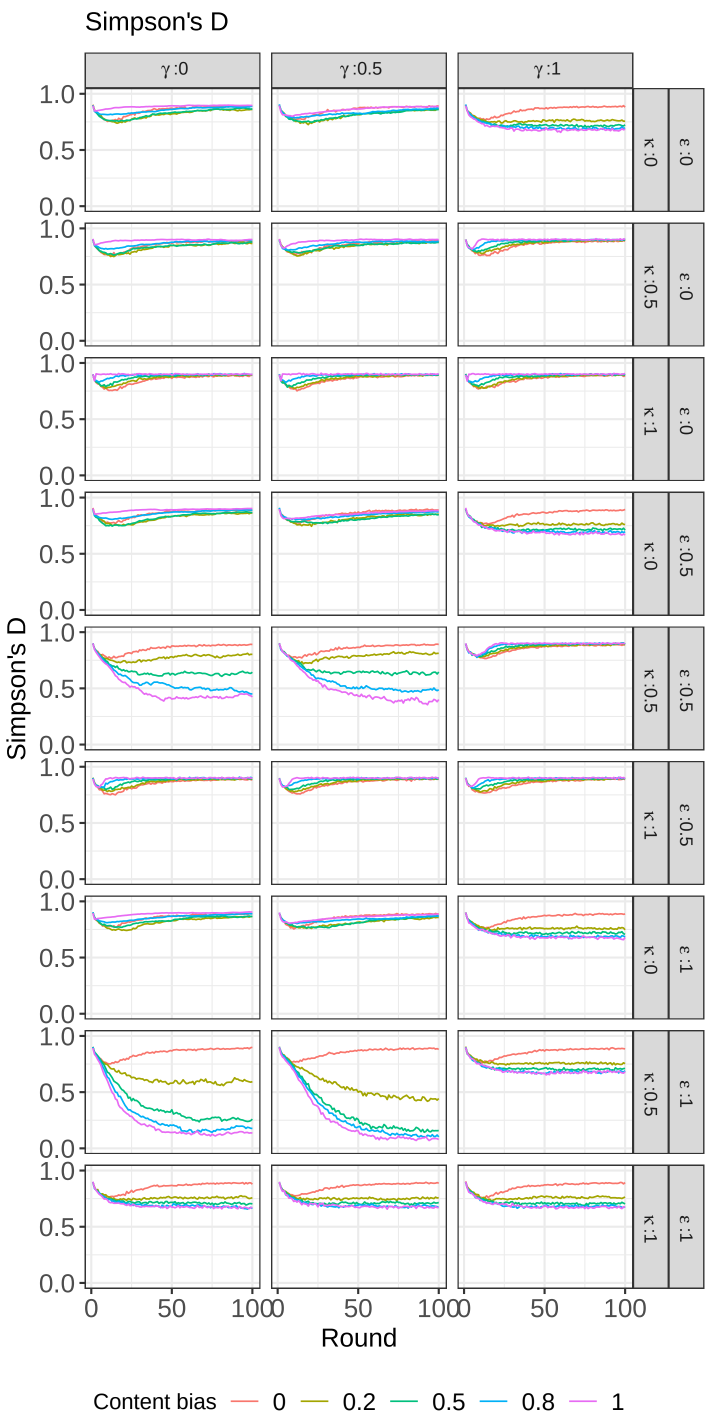


Figure 5.14: Diversity (measured as Simpson's diversity index D) of the set of produced variants over each level of *institutional power* (ε), *compliance bias* (κ), *confirmation bias* (γ) and *content bias* (β). Simulations with initial random assignment of variants selected from X without replacement (wR) and initial randomised value system (PR).

5.4 Discussion

We have provided the first detailed analysis of the co-evolution of diversity of cultural variants, institutional values and agents' value systems for varying conditions of compliance, confirmation, content and frequency biases, and for different levels of institutional power.

Institutional values and their capacity to convey those values to the population affects the agents' choices and in turn the diversity of cultural variants. In general, institutional power facilitates the emergence of shared cultural conventions. These results generally confirm previous studies suggesting that institutions facilitate the emergence of social conventions (David, 1994; Kearns et al., 2009; Voigt & Kiwit, 1998; Young, 2001), but they also show that powerful institutions can play an important role in the maintenance of cultural diversity. When institutions promote diversity and agents conform to the values promoted by institutions, the institutional capacity to transmit values efficiently becomes a prominent mechanism for the persistence of cultural diversity.

These results, however, are dependent on the institutional power and the cognitive biases of the population. In order to achieve high levels of convergence, societies with powerful institutions require agents to be at least moderately compliant, moderately content biased and moderately prone to consider new information to update their value systems. In the absence of institutional power and therefore of institutions that are in place to guide convergence, strong conventions can also emerge. This is consistent with previous studies showing that social conventions can emerge without the intervention of institutional mechanisms (Becker et al., 2017). However, in this scenario, high levels of convergence require a totally different combination of cognitive biases and value systems. Our model shows that, in the absence of institutional power, societies require more hegemonic and homogeneous values at the population level, higher levels of confirmation bias and lower levels of compliance to form strong cultural conventions. In this scenario, the agents' value systems are similar and they are additionally reinforced by a confirmation bias in the local interactions. Agents, therefore, end up having strong and similar values. As a result of this, population convergence on shared conventions increases quickly as content bias increases.

Institutional diversity arises from the diversity of the agent's value systems. The more diverse the agents' values are the more diverse institutions are. This is unsurprising, because it responds to the way in which the model is constructed. What is noteworthy is that institutions with unambiguous and hegemonic value systems can also increase diversity. However, this only happens when institutions are partially successful at conveying their values: When transmission weakens institutional values and agents have a moderate-high compliance bias, the result is that population value systems also fade over time and in turn cultural diversity skyrockets. Nevertheless, societies with moderately powerful institutions (either diverse or hegemonic) can stabilize the diversity of the produced variants at intermediate levels when agents are moderately compliant, especially when agents have a moderate-high tendency to consider new information to update their prior value systems. These results show that institutions, the agents' values and biases interact affecting the dynamics of social conventions: Different institutions can yield different or similar levels of diversity depending on how they interact with the agents' cognition.

Formal models show that conformity with institutional norms or beliefs can safeguard cultural diversity from erosion (Boyd & Richerson, 1985). In particular, strong conformity seems to affect human culture by facilitating stable in-group uniformity, which then stabilises between-group cultural diversity over time (J. Henrich & Boyd, 1998; N. Henrich & Henrich, 2007). In our model, compliance with institutional value systems works in a similar way in some scenarios: For example, when $\varepsilon = 1$ and the population behaves as an in-group (homogeneous and hegemonic value systems (OTA), an increase in diversity tends to increase convergence. In this context, when the population forms an out-group (PR), cultural diversity tends to increase when we move from intermediate levels to high levels of κ. Our model, therefore, suggests that compliance can be considered a prominent mechanism underlying the stabilization of cultural diversity.

However, our results also show that cultural diversity can be maintained without an individual compliance bias. This is particularly true when $\gamma < 1$, that is, when agents are more prone to use new information to update their value systems. Previous studies have suggested the possibility that the maintenance of the stability of cultural diversity could occur without a conformity bias (Acerbi et al., 2016). An interesting direction for future research would be to implement a *conformity bias* (disproportionate tendency to copy the majority) and test the results against the compliance bias implemented here. This would allow us to check whether a lack of confirmation bias, in combination with other evolutionary pressures, is a candidate for the preservation of cultural diversity in scenarios without an individual-level conformity bias.

Interestingly, in some scenarios (intermediate values of ε and κ), our study is consistent with recent studies showing that confirmation bias is determinant for the formation of communities of common interests, which in turn rises cultural variation (Del Vicario et al., 2017). In contrast to these results, in particular in PR scenarios, our simulations show that a strong confirmation bias without a compliance bias tends to increase convergence. It is still unclear, however, whether a confirmation bias alone can drive convergence on shared conventions without endowing individuals with other cognitive biases related to the intrinsic value of the cultural traits—in our case, a general individual-level content bias.

To ensure that our results were not dependent on the range of initial values of cultural diversity, we ran simulations using an initial assignment of variants without replacement. For all conditions the results were qualitatively identical to the runs with initial assignment of variants with replacement. We also ran simulations using a populations of 10 and 100 agents. Simulations results were again very similar between conditions. Additionally, we quantified diversity outputs using a number of alternative well-established alpha diversity indexes and we verified that all the different metrics yielded similar results. All these robustness tests mean that our model holds for a wide range of critical assumptions.

5.5 Conclusions

This chapter has introduced institutions, value systems and a set of cognitive biases into the dynamics of cultural evolution. This co-evolutionary model formalises the idea that cultural diversity changes with the processes of cognitive and institutional interaction. The interaction between cognitive biases, value systems and institutions reveal a complex relationship between individual-level compliance, confirmation bias and value systems and institutional-level values and power. In general, institutional power facilitates the formation of cultural conventions. However, institutional power can increase cultural diversity when agents are compliant and institutions promote diverse values.

In the absence of institutions that guide convergence strong conventions can also emerge. In our simulations, this occurs when populations of non-compliant agents have similar value systems and are endowed with strong individual-level confirmation bias. In more realistic scenarios, where institutional power is intermediate, a moderate compliance bias can be a potent mechanism underlying the stabilization of cultural diversity at intermediate levels.

The results presented in this chapter might be of interest to researchers investigating the impact of representative institutions on the evolution of linguistic, cultural, economic or social behavior (Becker et al., 2017; Bowles et al., 2003; Boyd & Richerson, 2008; Centola & Baronchelli, 2015; J. Henrich et al., 2005). In particular, an interesting research question for future work will be to explore in a formal model the relationships between conformity and compliance biases, how they interact and how they drive the emergence of cultural conventions in in-groups and out-groups. Future work using our co-evolutionary model will also be focused on mathematical modelling for a wider range of institutions.

The work hitherto developed in this thesis (Chapters 2 to 5) formalise into agent-based models (ABMs) a number of specific individual and social parameters to explore the evolution of cultural patterns. So far, we have modelled individual-level cognitive traits and society-level structures, run simulations and obtained averaged metrics of cultural products at the level of the population for both individual and collective features. Indeed, in the field of cultural evolution, agent-based models are recognised as essential tools to explore the effects of individual variation at the individual and population levels. The sources of individual variation are, however, more complex than what we can capture today with mathematical models. In spite of that, the utility of agent-based models is that they are versatile enough to include new sources of variation. That is probably why, in recent years, the use of ABMs have also expanded in the fields of ecology and evolutionary biology. This widespread use of ABMs has also led to the emergence of interdisciplinary collaborations between fields that have been traditionally separate, as well as to the use of complex systems approaches to unify conceptual frameworks. According to (DeAngelis & Mooij, 2005), five major types of individual variation can be identified in ABMs: spatial, ontogenetic, phenotypic, cognitive and genetic. In Chapters 2 to 5 we have used a number of state variables and parameters that capture general individual cognitive biases and other types of cognitive traits in addition to higher level society features. In Chapter 6 we also consider the importance of ontogenetic and phenotypic variation. In doing so we attempt to conceptualise a compatible framework for iterated learning, niche construction and ecological evolutionary developmental biology without needing to rely on strong genetic constraints.

Chapter 6

Iterated learning and evo-devo

Abstract

In this chapter we argue that ecological evolutionary developmental biology (eco-evo-devo) accounts of cognitive modernity are compatible with cultural evolution theories of language built upon iterated learning models. Cultural evolution models show that the emergence of near universal properties of language do not require the preexistence of strong specific constraints. Instead, the development of general abilities, unrelated to informational specificity, like the copying of complex signals and sharing of communicative intentions is required for cultural evolution to yield specific properties, such as language structure. We argue that eco-evo-devo provides the appropriate conceptual background to ground an account for the many interconnected genetic, environmental and developmental factors that facilitated the emergence of an organic system able to develop language through the iterated transmission of information. We use the concept of niche construction to connect evolutionary developmental accounts for sensory guided motor capacities and cultural evolution guided by iterated learning models. This integrated theoretical model aims to build bridges between biological and cultural approaches.

Keywords: language evolution; cultural evolution; extended evolutionary synthesis; niche construction; eco-devo

6.1 Introduction

Cultural evolution and biological evolution share a number of similarities that have long been recognised (e.g. Boyd & Richerson, 1985; Darwin, 1871/1981; Durham, 1991; Lumsden & Wilson, 1981). But since cultural inheritance and biological inheritance also encompass necessarily different features (e.g. Boyd & Richerson, 2005; Mace & Holden, 2005; Tëmkin & Eldredge, 2007), the study of both disciplines has followed relatively independent paths in terms of the methodologies and approaches used by each. Notwithstanding a number of scholars have argued that studying cultural phenomena within a unifying framework that takes insights from evolutionary biology is potentially useful to integrate separate disciplines (Charbonneau, 2016; Mesoudi et al., 2006). Cross-disciplinary approaches have also been defended for the field of language evolution (e.g. Bickerton, 2003; Christiansen, Dale, Ellefson, & Conway, 2002), although the uniqueness of human languages has undoubtedly delayed the construction of theoretical integrated frameworks incorporating both the findings in computational modelling and state-of-the-art empirical knowledge in evolutionary developmental biology.

Human languages are different from other animal communication systems. For example, they exhibit a semantically compositional structure that enables humans to manipulate long and complex chains of signals. This feature is known as *Fredge's Principle of Compositionality*, which essentially boils down to the fact that the meaning of a complex expression is a function of the meaning of its parts (for an overview: Krifka, 1999; Szabó, 2017, and references therein). Compositionality of meaning is generally assumed to be intimately connected to two other distinguishing properties of natural languages, namely their productivity and their systematicity (Fodor & Lepore, 2002), which in turn are related to a property of the syntactic principles responsible for the construction of complex expressions. This distinctive feature of language is often referred to as recursion (e.g. Hauser, Chomsky, & Fitch, 2002). Other properties of languages are duality of pattern, convexity, linearity or displacement (Hockett, 1960, 1966). The simultaneous presence of these distinctive features makes human languages open-ended communication systems (Kirby, 2017).

A common way to explain the origins of these features and, more generally, language structure, is natural selection (e.g Pinker & Bloom, 1990). But, as Eric Lenneberg pointed out, this approach can be problematic if it intends to explain evolution as a simplistic and unidirectional mapping of genotypes onto phenotypes, and it only pays attention to 'the biological usefulness of certain features of animal communication' (Lenneberg, 1967, p. 253). Firstly, because evolutionary biological dynamics can be radically altered by other external pressures such as the environment (Gilbert & Epel, 2009; Sultan, 2015) or, in the case of linguistic phenotypes, culture (Kirby, 2017); secondly, because dependencies between genes and phenotypes can not be drawn unidirectionally or attending to a single locus (Fisher & Vernes, 2015; Marcus & Fisher, 2003) when it comes to explaining language; and finally, because it has been shown that a constellation of processes that bias selection and modify the frequency of heritable variation, such as developmental biases and niche construction, can alter the way in which natural selection proceeds (Deacon, 2010; Laland et al., 2000; Lewens, 2019; Robert, 2004).

This chapter is structured as follows. In the next section, we will focus on the definition of domain specificity and we will propose a revision of the concept in light of alternative models that eschew traditional versions of genetic determinism. Then, in Section 6.3 we will review some relevant models suggesting that language regularities can be successfully acquired and transmitted without the need for strong genetic encoding. In Section 6.4, we expound the minimum requirements for iterated learning to work in the light of recent controversies. Then, in Section 6.5 we provide a brief review of the history that led to the development of eco-evo-devo models and argue for the need to abandon traditional dichotomies in order to better account for the linguistic phenotype. Finally, in Section 6.6 we will revisit a variety of studies that might be adding evidence to support the main hypothesis of this study: The conceptual apparatus of eco-evo-devo models is compatible with the findings of iterated learning models and helps dissolving the boundaries between a traditional dichotomy that has been limiting our understanding of the evolution of language. Using the notion of niche construction, where individual organisms play a much central role than in standard approaches, we will propose an integrated theoretical framework that stresses the need to connect the development of sensory-guided motor capacities and the requirements for iterated learning. Our conceptual model intends to help fill gaps in our knowledge about how *variational explanations* (changes due to variation within the population) and *developmental explanations* (changes due to variation within the individual) relate, as well as to provide a framework for language and cultural evolution to advance in the construction of new hypotheses upon which triple-inheritance models can be developed.

6.2 Domain specificity

The modularity of mind is a hypothesis about the architecture of mind according to which a number of cognitive systems, typically associated to perception, operate in characteristic ways that makes them, among other things, domain-specific and mostly impermeable to the operation of other modules and cognitive systems (Fodor, 1983, 1985). There is no conceptual or logical connection between the notion of modularity and nativism; but it is often the case that their proponents, on the basis of such considerations as the poverty of stimulus argument, assume that such functionally defined modules are associated to the corresponding Chomskyan-modules—that is, innate repositories of domain-specific information that are supposed to underlie our cognitive abilities in various domains. Accordingly, a system is domain specific if the class of objects and properties that it computes information about is restricted within narrow limits (Fodor, 2000; Robbins, 2017). Under this definition, humans would be endowed with systems of knowledge which serve as specialised evolutionary devices for specific tasks. For example, knowledge of language would be a domain-specific system that gives humans the ability required for the acquisition and use of language (Chomsky, 1986; Spelke & Kinzler, 2007). A particularly radical version of this stance is exemplified by evolutionary psychology and its massive modularity thesis according to which all extant human cognitive abilities (not just the peripheral ones) are modular and, also, adaptations to the environment of the Stone Age (Barkow, 1992; Plotkin, 1997). Differences as to the extent of modularity notwithstanding and focusing our attention on language, it is certainly true that both stances appear to be committed to some form of nativism according to which neurally specific modules for language are shaped by specific genes (Berwick & Chomsky, 2016; Pinker & Jackendoff, 2005). Researchers have found support for domain specificity in many different ways, e.g. looking at the competencies of infants, comparing human capacities with other animals or using the poverty of the stimulus argument as evidence for universal grammar (Berwick, Pietroski, Yankama, & Chomsky, 2011; Chomsky, 1967; Pinker, 1991).

But attempts to address key questions such us ‘why only us?’ (see Berwick & Chomsky, 2016) or, how do children acquire language without sufficient evidence in the primary linguistic data (Chomsky, 1965), have not always ended up proposing models that verify the existence of a domain-specific module for language. To be sure, even the human- and language-specificity of the computational operation Merge, the only putative genetically determined residue of UG in Chomsky’s Minimalist Program (Chomsky, 1995, and later work), has been called into question on the grounds of a detailed analysis from the perspective of the notion of biological homology (Balari & Lorenzo, 2013, 2015). In the field of cultural evolution, these same questions have been addressed using a variety of experimental and computational methods that, without relying on strong genetic constraints or domain specificity, model the successful acquisition and emergence of universal properties of language (e.g. Chater et al., 2009; Culbertson & Kirby, 2016; Kovas & Plomin, 2006; J. L. Morgan, Meier, & Newport, 1989; Scerif & Karmiloff-Smith, 2005; Smith & Wonnacott, 2010; Zuidema, 2002).

The notion of domain specificity has traditionally been linked with innateness in different ways, causing significant confusion in the field. However, if proponents of innateness argue that language acquisition is determined by genetic factors, and proponents of domain specificity claim that language is processed in localised modules that deal exclusively with a single information type, then we can no more argue that these two issues are automatically interchangeable in the debates about the evolution of language (Bates, 1994; Elman et al., 1996).

It is relatively common ground in the field of cultural evolution studies that domain-specific constraints, when genetically wired, might have evolved to take the form of weak biases or general capacities that, amplified by culture, interact with the linguistic system in domain-specific ways (Culbertson & Kirby, 2016). Be that as it may, we suspect that the whole debate may acquire a totally different flavor as soon as one adopts a developmental view. Firstly, because of the fact, firmly established already by 19th century embryologists like Karl Ernst von Baer, that all development follows a pattern going from the less specific to the particular and is hardly a matter of master control genes (Minelli, 2003). Secondly, because the traditional interpretation of the innate-acquired distinction, where what is innate is typically assumed to be internal to the object in question and, consequently, genetic, is most probably misleading (Keller, 2010; Wimsatt, 1986). We see no reason why the case of language should be different.

In the next section we will review some models that show how language features can emerge in the absence of strong genetic constraints, and demonstrate how such abilities as copying and sharing might be sufficient, when combined with iterative learning, to yield outputs that appear domain-specific without the need for strong language-related biological predispositions.

6.3 Challenging domain-specificity

The argument from the poverty of stimulus (henceforth, POS) states that children are not exposed to sufficient data within their linguistic environment to induce their native language. In 1967, Mark Gold provided a formal proof that has usually been interpreted as evidence for this argument (Gold, 1967). Gold's proof showed that, given a context-free grammar, regardless of the number of samples from an infinite language are presented to a learning algorithm, the algorithm can not accurately determine whether the samples belong to an infinite language or to a finite subset thereof containing the samples in question.

To investigate how a grammar that would be unlearnable by Gold's method could be acquired successfully, Zuidema (2002) constructed a model that uses *cultural evolution*. The model implements linguistic abilities using context-free grammars and three operations called *incorporation*, *compression* and *generalization*. When the algorithm is initiated it produces random strings, simulating transmission from the parent to the child. In these randomly generated strings some regularities may appear, for example: *aab*, *bab*, *cab*. In this example, the child can compress the substring *ab* into the non-terminal X: $S \mapsto aX$, $S \mapsto bX$, $S \mapsto cX$, $X \mapsto ab$. Then, say the child obtains another rule from another set of strings: $Y \mapsto d$. Now, the generalization operation can equate the non-terminals X and Y. This means that the child can obtain the unobserved strings *ad*, *bd*, *cd* from the resulting grammar. Over generations, in a population of agents, language becomes more structured and unseen strings more learnable, increasing communicative success. With this elegant model, Zuidema (2002) showed that POS is not necessarily a hindrance for learners to successfully acquire grammars from a class that is unlearnable by Gold's criterion.

To explore the extent to which language genes in the form of a highly specialised module could have co-evolved with language properties, another well known computational model was constructed by Chater et al. (2009). They simulated a population of language learning agents where arbitrary linguistic principles could become genetically encoded via the *Baldwin effect* (Baldwin, 1896; Weber & Depew, 2003, for a contemporary perspective). In evolutionary biology, the *Baldwin effect* describes a process where individuals have the ability to acclimatise to new pressures during their lifespan by learning a new behaviour. This mechanism would affect the individuals' reproductive success and the new trait could become gradually encoded in the genome over generations. However, Chater et al. (2009) showed in their study that this genetic encoding gets significantly reduced when the rate of language change is high enough. Therefore, they concluded, since language changes much more rapidly than genes, genetic evolution of domain-specific constrains is unlikely. As pointed out by Culbertson and Kirby (2016), there is nonetheless room for a more nuanced thesis that supports the existence of weak biases—that is, soft constraints that can impose a continuum of weak preferences—affecting language acquisition. And it could be the case that these weak biases of the individuals were not reflected in the spoken language.

Kirby et al. (2007) investigated this by testing how innate biases are related to universal properties of language. Their model shows that cultural transmission can amplify weak biases and end up producing language properties which are near universal. If this is the case, cultural transmission would have produced 'apparent adaptations', that prevented the evolution by natural selection of strong constraints in the form of domain-specific genes, mainly because those genes would be highly prone to drift.

We know, however, that the relationship between learners' biases and language structure is not straightforward when it comes to explaining linguistic variation. In a recent study that uses both a Bayesian model of learning and transmission and collected data from an artificial language learning experiment that mirrors the model, Smith et al. (2017) showed that weak biases can have a wide range of effects on language structure, from strong to weak, or even no effects whatsoever. Therefore, transmission and use are essential for understanding the interactions between biases and statistical learning.

For the purpose of this chapter, the examined models constitute sufficient evidence to illustrate the discussion for the next section. For a more detailed review of this line of work see Kirby (2017).

6.4 What does iterated learning actually require?

Berwick and Chomsky (2017, p. 172*ff*) argue that cultural evolutionary approaches have generally mistaken the word ‘universal’ as a property of the faculty of language with Greenberg’s linguistic universals (Greenberg, 1966), or properties of externalised languages. They claim that what ultimately evolves in these models is a population of learning agents’ choices and agents that already had the ability to choose between two alternative concept representations. In their words, this would not solve the problem of where a universal comes from (e.g. compositionality) because the ability to build context free grammars, generate infinite languages and/or even something like Merge is presupposed. Thus, they conclude, iterated learning models do not satisfactorily attempt to delimit any pre-existing innate universal grammar (UG) related with the language faculty.

But this might not be the case. Iterated learning models of language evolution define stochastic processes that can be mathematically characterised using Markov chains. To analyse the requirements for iterated learning, it is necessary to understand the core concepts that define the properties of Markov chains. In Appendix C, we provide an accessible summary that includes a brief presentation of Markov chains, along with a numerical example applied to language transmission. For now, it is important to note that Markov chains are very useful to analyse iterated learning processes by computing a transition matrix (a square matrix that gives the probabilities of different languages going from one to another) and finding the stationary probability of each language. For more detailed explanations of Markov chains see Kemeny and Snell (1983), Brémaud (1999, Ch. 2), or Griffiths and Kalish (2007), for example.

However, in the real world, learners have individual biases that affect the results of the predictions of an iterated learning process that has been reduced to a Markov chain. In order to construct learning algorithms that incorporate a wide characterization of these biases for a wide set of cognitive features, a number of researchers have used bayesian agents applied to human cognition (e.g. J. R. Anderson, 1990; Oaksford & Chater, 1998) and the emergence of linguistic regularities (e.g. Brighton, 2002; Kirby, 2001; Smith, Kirby, & Brighton, 2003). Interestingly, the predictions of these computational approaches have been successfully reproduced and tested against data obtained from psychology experiments with human participants (Kirby et al., 2008; Tamariz & Kirby, 2015).

Here, we review Bayes' rule applied to language acquisition. For a detailed analysis of iterated learning using learning algorithms based on Bayesian inference, see Griffiths and Kalish (2007).

The Bayesian framework used in iterated learning models computes the posterior probability of an event according to Bayes' theorem:

$$P(h \mid d) = \frac{P(d \mid h)P(h)}{P(d)} \tag{6.1}$$

where $P(h)$, named *prior* probability distribution, is the estimate of the probability of the hypothesis $h \in \mathcal{H}$ before d is observed—it encodes learner's biases. $P(h \mid d)$ is the posterior probability, the probability of h after d is observed. $P(d \mid h)$, named the *likelihood*, is the probability of observing d given h, and $P(d)$, named the *marginal likelihood*, is the probability of d averaged over all hypothesis,

$$P(d) = \sum_{h \in \mathcal{H}} P(d \mid h)P(h) \tag{6.2}$$

Applied to language acquisition, h is a language, and d the set of utterances sampled from the target language. Additionally, each learner has a *learning algorithm* (LA) that specifies the procedure for choosing h after observing d, and a *production algorithm* (PA) that specifies how they choose d given h.

Now, if we assemble this rule from generation to generation by forming an iterated learning process based on the principles of Bayesian inference, we have a Markov chain where each learner produces a set of data (a posterior distribution over languages) by combining a prior (representing their inductive biases) with the data produced by the previous generation. Then, this data is supplied to the next generation, and so on, as illustrated in Figure 6.1.

As stressed by Griffiths and Kalish (2007), a prior should not be interpreted as reflecting innate predispositions to language acquisition, but

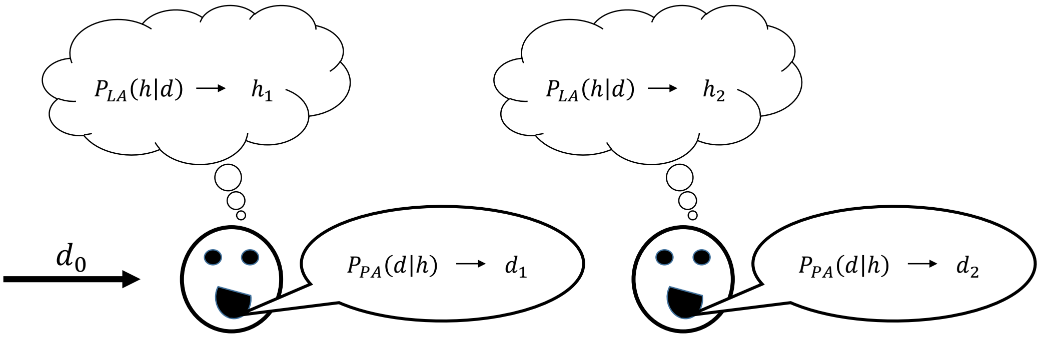


Figure 6.1: Iterated learning has been proposed as an explanation for the emergence of linguistic regularities and the existence of linguistic universals. Each learner sees a set of utterances (d) produced by the previous generation, forms a hypothesis (h) about the language from which those utterances were produced, and uses this hypothesis to produce the data that will be supplied to the next generation. Figure adapted from Griffiths and Kalish (2007)

.

as a collection of factors, not necessarily domain-specific constraints, that affect the agents' own hypothesis. So, although there might be a sense in which we could correctly say that there are basic functionalities built into the model, none of them are language related. In fact, these models require only two skills: The ability to learn data and the ability to produce data (for transmission). The mechanisms underlying these abilities may be quite elaborate, but, to make the point clearer, not innately determined (in the traditional sense) to deal with specifically linguistic data.

We can observe, then, that the concept of a pre-existing biological condition leading our species alone to possess language is not in fact discussed in the above computational models. Instead, what is challenged is the specificity and language-related origin of that genetic basis. In this line of thought, Kirby (2017) pointed out that if we want to look for human adaptations related to a precondition for language, then we might better look at the biological origin of these two traits: The ability to copy vast sets of behaviours, and our predisposition to share. These two non-language-specific predispositions seem to be the basic requirements for iterated learning models to work. It is important to stress here that the outcome of iterated learning is not simply built in into the properties of the learning algorithms (Griffiths & Kalish, 2007; Kirby et al., 2007; Smith et al., 2003). The Bayesian approach does not imply that there are not biological factors that play a role in the emergence of language regularities, but puts the emphasis on the interaction between agents' cognition and transmission.

If this is true, the relevant question to ask now if we are looking for human adaptations that biologically configured the so called language-ready brain in our species, is which current biological approach accounts best for the emergence of the key necessary biological changes that brought about the mentioned abilities (to copy and share). In the next section, we will suggest that evo-devo can be seen as the best general perspective to be taken when approaching this question, and we will discuss which implementation of evo-devo best fits the requirements of iterated learning to operate.

Throughout this investigaton, we will aim at showing that such traditional distinctions as general vs. specific or Faculty of Language (Narrow sense; FLN) vs. Faculty of Language (Broad sense; FLB) of Hauser et al. (2002), should be abandoned in order to construct a less simplistic developmental approach to the complex cognitive capacities that serve as the basis for iterative learning processes to give rise to language universals.

6.5 Which approach should we take to account for the preconditions for iterated learning?

The Modern Synthesis, a term popularised by Huxley (1942), gave rise to modern biology by gathering a number of postulates from natural selection, population genetics and Mendelian inheritance into an articulated corpus of empirical evidence and mathematical laws. Ernst Mayr, a key evolutionary biologist of the past century, was one of the main figures of this conceptual revolution in the field. Among his contributions, articulating the biological species concept and studying different forms of allopatric speciation stand out. Mayr's open skepticism towards what he called 'beanbag genetics' notwithstanding (Mayr, 1963), he nonetheless contributed to the consolidation of a biological thought centered on the notion of a 'genetic program' (Mayr, 1982) and on a neat separation of 'proximate' vs. 'ultimate' causes (Mayr, 1961). As a direct consequence of this stance, the mainstream orientation of the Modern Synthesis tended to ignore developmental processes and their role in evolutionary dynamics (Amundson, 2005; Maynard Smith, 1982; Robert, 2004), while, at the same time, organisms disappeared from the explanatory apparatus of evolutionary biology (Walsh, 2015). However, such a view on causality has often been considered as highly problematic (e.g. Laland et al., 2013;

R. C. Lewontin, 1974; Oyama, 2000b; Walsh, 2019), and Mayr's genetic program has also been shown to be unable to reflect the environmental context-dependency of phenotypic outcomes (e.g. Gilbert & Epel, 2009, 2015; R. C. Lewontin, 1983; Sultan, 2015).

In light of the discovery of the toolkit genes—highly conserved genes whose products regulate gene expression and control the organism's embryonic development—developmental geneticists and evolutionary biologists have been forced to confront each other's ideas in a more interconnected way. This filled the gap between both levels of analysis and gave rise to evo-devo, a new discipline that, since its origins, has been expanding upon the evolutionary synthesis (Carroll, 2008; Pigliucci & Müller, 2010, for two slightly different perspectives of this new synthesis).

Evo-devo, however, is not a unified theory (Benítez-Burraco & Longa, 2010; Hall, 2003), but a theoretical trend or general perspective where different evo-devo models fall here or there. Balari and Lorenzo (2013, chapter 6) describe three main categories of evo-devo approaches:

(i) those that encompass genome deterministic models;

(ii) those that include developmental factors beyond the genes but maintain a gene-centered approach; and

(iii) those that hold that disparate factors interact to bring about ontogenetic outcomes.

The last category approximates what may be categorised as the 'eco-evo-devo' approach, which shares a fair number of assumptions with the framework of the extended evolutionary synthesis (henceforth EES; Müller, 2020; Sultan, 2017). According to the proponents of the EES, developmental processes, including cellular products, intermediate phenotypic states, environmental inputs and behavioural practices, share with inclusive inheritance and niche construction, the potential to drive individual variation and, ultimately, evolution (Laland et al., 2015; Müller, 2017). Given this classification, it is not difficult to tell a priori which evo-devo category fits better with the concept of a strong domain-specific faculty of language and which one explains human motor capacities as the result of a complex architecture of interconnected developmental levels.

For example, as pointed out by Benítez-Burraco and Longa (2010), Chomsky has recently suggested non-trivial analogies between the biolinguistic approach (BA) and evo-devo (Chomsky, 2007, 2010). According to Benítez-Burraco and Longa, however, Chomsky's analogies mostly refer to that version of evo-devo that fully assumes a gene-centered perspective (e.g. Carroll, 2005), a stance that might have been appropriate as regards the Principles-and-Parameters Theory, but that, as these authors extensively argue, does not even fit well with a minimalist BA. Be that as it may, Chomsky has also advocated for a tripartite causal model according to which different aspects of the linguistic phenotype may be neatly attributed to well-delimited factors, namely genetic endowment, experience (i.e. the environment), and general principles not specific to the faculty of language like principles of data analysis, computational efficiency, or developmental constraints, among others (Chomsky, 2005). This is precisely the kind of analysis of causes that R. C. Lewontin (1974) showed to be impossible and that today still survives in the nature vs. nurture debate under its different guises (Keller, 2010).

Chomsky's views are direct heirs of a tradition where such dichotomies as internal vs. external, inherited vs. acquired, and genes vs. environment (or culture) have played an important explanatory role. But this stance radically comes into conflict with the idea, widely shared by most supporters of view (iii) above, that nature is not genetic but phenotypic; that nature is not a self-contained internal program but rather the open-ended product of a dynamic developmental interaction between internal factors including genes with external, environmental ones (Oyama, 2000a). Development thus arises from a complex network of causal interactions in which organism and environment co-construct each other (Laland et al., 2013, 2014) through reciprocal influences that effectively break the supposed barrier between the internal and the external (Sultan, 2015, 2019). An immediate consequence of this is that the genetic regulatory systems of developmental genetics loose their causal primacy in favour of the causal complex made up by the organism and its environment. Similarly, the concept of inheritance also changes to embrace an extended form of inheritance where the developmental 'resources' range from DNA sequence, to environmentally-induced epigenetic marks, to the location and ecological niche the organism inhabits (Bonduriansky & Day, 2018; Laland et al., 2015).

We have focused on Chomsky's attitudes towards a number of crucial issues that eco-evo-devo invites us to look at through a different prism, but it should be clear that some of these criticisms also apply to other frameworks, not necessarily friendly to Chomskyan thought. To be sure, since—and paraphrasing R. C. Lewontin (1974, p. 401)—the relevant questions are not whether the phenotype of an individual is the result of *either* environment *or* genotype, or of *either* biology *or* culture, or of *either* nature *or* nurture, because the phenotype, to the extent that all these dichotomies make real sense, is the result of *both*. Accordingly, the framework we are advocating for here definitely shares a number of central points with Cecilia Heyes's 'Cogntive Gadgets' (Heyes, 2018). Indeed, we agree that when the cognitive equipment of newborn humans is exposed to 'culture-soaked' human environments, it changes dramatically. Similarly, in this study we propose that we are born endowed with complex cognitive mechanisms that emerge as part of our development and that these mechanisms continue to develop throughout life within our deeply social and anthropised environments. These social environments constitute diverse human niches, which are in turn affected by horizontal and vertical transmission processes in which the ability to copy and share information at a fast rate play a crucial role (Laland, 2017a; Tomasello & Carpenter, 2007; Tomasello, Carpenter, Call, Behne, & Moll, 2005). But our approach here also differs from Heyes (2018), at least, in denying the necessity, if not the possibility, of retaining the nature vs. nurture dichotomy and of neatly identifying the causal contribution of each. The cognitive equipment of newborn humans may not differ substantially from the minds of closely related species. Yet, human cognitive diversity and evolved predispositions might be the product of observable causal processes whose causes can not be depicted as totally disentangled, because that would lead to epistemological contradictions when it comes to clearly delimiting objects, causes and effects. For example, DNA is both inherited and environmentally responsive, and we know this in enough detail to move beyond the nature-nurture debate (Robinson, 2004). Thus, in the model we propose here, the contingencies of those traditional categories are reduced to mere instrumental categories.

To the extent that this new paradigm constitutes the recognition of the need to adopt a pluralistic attitude toward the complex nature of the language faculty, the emergence of which can not be clearly quantified in terms of internal versus external structures, nor characterised as a unique object, we think that eco-evo-devo and EES approaches are also demanding the abandonment of such traditional distinctions as FLN/FLB and others we already referred to above. Many of these arguments that we will not expand here have been amply analysed using a variety of biolinguistic approaches (Balari & Lorenzo, 2018; Boeckx, 2014).

Having droppend such distinctions, we think, the search for the faculty of language is also freed, to some extent, from the metaphor of specificity/generality, facilitating the construction of a non-reductionist, less simplistic, general theory of language that encompasses a complex multifactorial cognitive human capacity that does not yield specific linguistic outputs by itself, but is required subsequently to give rise to the phenotype through learning and transmission.

In the next section we will revisit some studies that add evidence to support such an interconnection of factors underlying the so-called 'language-ready brain,' without the need for strong straight dependencies between specific genes and specific language properties—even if key genes have obvious subsequent dramatic effects on the development of language. We will focus on the relationship between genetic factors, such as *FOXP2*, and cognitive abilities. Then, in the light of niche construction, we will argue in favour of a general theory of language evolution that integrates the developmental architecture of cognitive abilities and iterated learning models.

6.6 Towards an integrated theory: insights from comparative genomics and niche construction

As soon as we depart from a simplistic gene-centered approach that relies on an incredibly lucky mutation or behaviourally assimilated trait to explain language complexity and our capacity to acquire it, we face the need to expand our approach to incorporate developmental processes that explain how complex functional phenomena evolve. During the last decades, neuroscientists have gathered evidence that some cognitive domains can operate as overlapping functional architectures. For example, language processing has traditionally been associated to Broca's area, but fMRI studies have also identified activation patterns in Broca's area associated with recognition, imitation or movement preparation (e.g. M. L. Anderson, 2010). These neural reuse theories seem to be incompatible with strong conceptions of structural or functional modularity and offer an interesting perspective for the conception of more comprehensive evolutionary-developmental models.

Recent evo-devo approaches have focused on the molecular analysis of behavioural traits such as learning and memory applied to the evolution of language. For example, genes relevant for language, including the transcription factor FOXP2, have been identified. *FOXP2* was initially identified as a genetic factor of a speech disorder in a family known as KE, and was thus the first gene to be associated with speech and language (Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998; Lai et al., 2000).

Despite the strong correlation between a *FOXP2* variant and developmental verbal dyspraxia (DVD) (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001), it must be noted that *FOXP2* belongs to a complex molecular network of genes that build proteins that in turn regulate the expression of other genes. In particular, *FOXP2* is controlled by a set of upstream regulators, and it in turn regulates a vast set of target downstream genes, by repressing or activating them (Shu, Yang, Zhang, Lu, & Morrisey, 2001; Vernes et al., 2007).

Although *FOXP2* is the best known gene in the field of language evolution, it doesn't work alone. A huge variety of gene products regulate neuronal development and function, including 'proliferation, migration, neurite outgrowth, and axon guidance, as well as development, maintenance, and plasticity of synapses' (Fisher & Vernes, 2015). From an evo-devo point of view, human speech can be described as a form of auditory-guided, learned vocal motor behaviour, and *FOXP2* and its regulatory molecular network might be key factors to 'shape neural plasticity in cortico-basal ganglia circuits underlying the sensory-guided motor learning in animal models' (Scharff & Petri, 2011). Indeed, the connection between vocal learning abilities in several species including humans and a number of homologous gene networks and brain structures is today incontestable (Jarvis, 2019, for a review). Since language is culturally transmitted, a cognitive impediment within these molecular networks would affect the emergence of language properties in a community through iterative learning processes. Actually, the emergence of language properties through cultural transmission requires both ingredients: The adequate development of neural circuits and the adequate social structure. Note that neither of these components by themselves would be able to result in linguistic specific outcomes.

So, to what extent have these two processes—the emergence of a regulatory neural network and the social requirements for iterated information transmission—developed independently one from the other? In Section 6.4, we showed that iterated learning does not require innately determined abilities to account for the emergence of linguistic regularities and in Section 6.5 we showed that an eco-evo-devo approach is a valid framework to account for the preconditions for iterated learning without the need of relying on strong genetic constraints. Both, iterated learning processes by generating novel linguistic phenotypes, and eco-evo-devo processes—such as developmental plasticity, genetic accommodation and extragenic inheritance—by facilitating evolutionary transitions and the alteration of environments and niche construction (Gilbert, Bosch, & Ledón-Rettig, 2015), shape evolution by constructing 'extended phenotypes '(Simon & Hessen, 2019), which in turn promote niche construction, that is, the ability to produce better nests, houses, institutions or environments (including linguistic ones).

Niche construction, therefore, can be thought of as an emergent property of triple-inheritance systems that take into account all the three transmission pathways of genes, culture and environment (Kobayashi et al., 2019). Just as adaptive behavioural phenomena results from iterative processes at different scales, niche construction selects the behaviour of the organisms in an iterative process during ontogeny (Simon & Hessen, 2019). Organisms' traits develop by interacting with the environment, and in turn increasing the expression of synergistic relationships between different levels of development (R. C. Lewontin, 1983). A similar argument was originally offered by Deacon (1997) and Bickerton (2009), where it is suggested that the repetitive use of symbolic communication can create socially artificial niches that in turn enforce new pressures on human cognition.

Using ideas originally developed by R. C. Lewontin (1983), Laland et al. (2000) constructed a version of this conceptual model by mapping the causal relationship between biological evolution and cultural change. This model proposes that biological evolution depends not only on natural selection and genetic inheritance but also on *niche construction*. According to this framework, phenotypes have a more active role in development and culture amplifies the human capacity to alter sources of natural selection. Cultural traits affect the environment and may have additional effects on how evolution proceeds. These changes, in turn, may persist throughout generations, beyond the lifespan of an individual organism. Crucially, cultural change can occur at a much faster rate than biological change. Culture, therefore, can relax or intensify selection and create new demands by changing ecology, which favours new adaptations (Whitehead, Laland, Rendell, Thorogood, & Whiten, 2019). Interestingly, a relaxation of selection at the organism level may have given rise to new complex synergistic features of the human language capacity, which may explain why so much language information is 'inherited' socially (Deacon, 2010). At the level of the population, as our species constructed its niche for enhanced social relations, 'self-domestication' or 'self-control' might have driven the selection of anatomical and behavioural traits whose functionality is related with mild neural crest cell deficits during embryonic development (Shilton, Breski, Dor, & Jablonka, 2020; Thomas & Kirby, 2010; Wilkins, Wrangham, & Fitch, 2014).

Niche construction can result from different sources (genetic, ontogenetic, and cultural processes) and affect both biological and cultural evolution (Laland et al., 2000); for a number of examples, see Naiman, Johnston, and Kelley (1988), for beavers, Laskowski and Pruitt (2014), for social spiders, and Feldman and Cavalli-Sforza (1976) and Lotem, Halpern, Edelman, and Kolodny (2017), for some cases concerning humans. As regards communication, niche construction has also been invoked in hypotheses about language evolution (e.g., Bickerton, 2009, 2014; Deacon, 2010; Laland, 2017b, among others). A number of learning biases and sensory-guided motor capacities (e.g. vocal control) evolved in response to new environmental and social pressures. Since this new communicative feature became extremely important within human populations for successful integration in human societies and, in turn, reproduction, it also could have brought about selection favouring better acquisition and transmission. This would obviously include our capacity to copy and share large sets of communicative variants. However, unlike deterministic and Baldwinian models, niche construction does not assume genetic assimilation of linguistic features, nor innate or language-specific knowledge. Instead, niche construction favours selection of motor capacities, cognitive biases and environments (e.g. social structures) that in turn facilitate the maintenance of such a niche.

Using a version of the previous niche construction framework, a general causal graph can be constructed by putting together all the interactions between the relevant variables (Figure 6.2). A key feature that differentiates our model is the inclusion of insights from both iterated learning models and current eco-evo-devo theoretical approaches. The model is constructed in the light of comparative genomics and niche construction: In our integrated version, niche construction processes, which are in continuous interaction with both the individual motor capacities of individuals and the environment, favour transmission trough iterative learning processes, resulting in a particular language phenotype. Thus, niche construction is considered as a key feature of the model, since it has a prominent role altering two main sources of variation that are directly related with iterated learning processes. On the one hand, sensory guided motor capacities from neural development, and on the other, environmental structures such as social structures, rules or cultural conventions. In turn, a change in these two sources of variation can modify the agents' learning and production algorithms during cultural transmission. This consequently connects neural development and the agents' own hypotheses in a way that could be potentially implementable by using iterated learning models. To capture this idea in an integrated model, each learner's *learning algorithm* (LA) and *production algorithm* (PA) should be constructed as a function of the variables altered by niche construction. Since cultural transmission of language occurs at a much faster rate than organic evolution, it can quickly create new pressures that, in fast iterative cycles, accelerate the emergence of new linguistic adaptations. This effect would relax selection at the level of the individual, due to a large redistribution of selective pressures and a diversification of social traits inheritance mechanisms.

We agree with Odling-Smee and Laland (2009) that niche construction will fail to account for the evolution/development of human language until we take into account the exceptionally powerful role of human cultural processes and the mutual scaffolding effects between them, cognitive abilities, and individual biases in our species (Wimsatt, 2014; Wimsatt & Griesemer, 2007). For example, since language change occurs at a faster rate than genetic changes, neural development selected for language might reflect 'the most persistent and invariant demands of the highly variable linguistic niche' (Chater et al., 2009; Deacon, 2010). This idea is coherent with a model that incorporates niche construction, organic and cultural evolution, where phenotypes (such as the language phenotype) have a much more active role in evolution (Gilbert & Epel, 2009, 2015; Laland et al., 2000; Sultan, 2015).

Moreover, since the construction of human cultural niches is able to favour effective cultural responses beyond the lifetime of individuals, with impact, or absence thereof, on human genetics, it adds more uncertainty into the evolutionary process. Depending on the time-frame used, the social and environmental structure, and the specific communicative feature examined, researchers might find different niche construction effects. Individual cases will require individual explanations, and they are showing little by little that overcoming the limitations of models based on general observations is more necessary than ever before. Here, computer modelling of language evolution that simulates population dynamics using iterative learning may be helpful to expand upon theoretical frameworks for language development like the one that we present here.

There are several reasons to think that cultural niche construction can

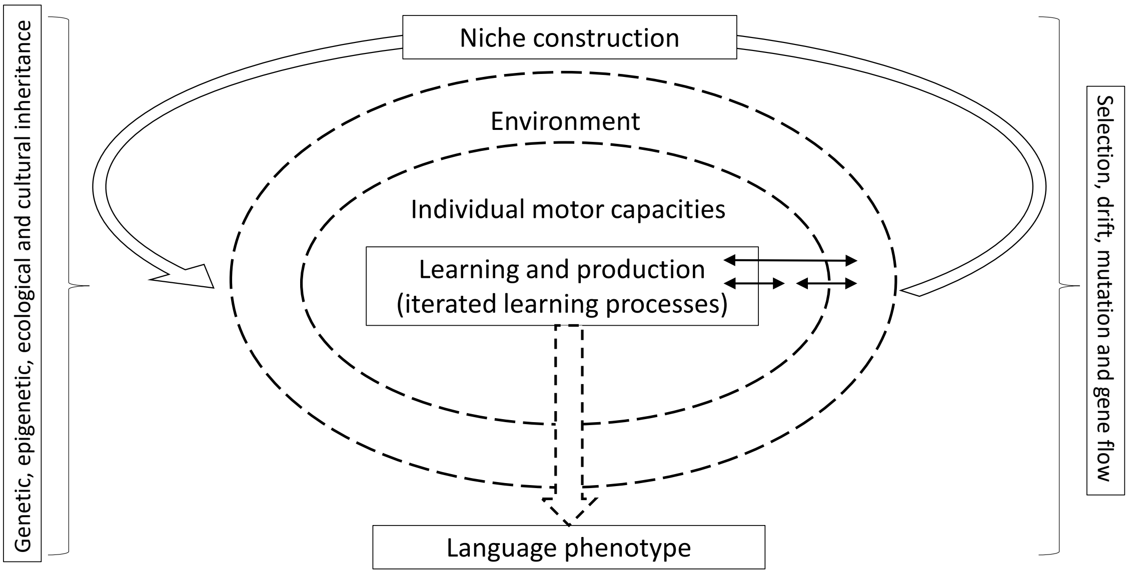


Figure 6.2: Integrated causal graph for the emergence of a language phenotype through developmental interactions. A change in the source variable causes a change in the destination variable. The area of the square in the middle represents learning and production through iterated learning processes. Learning and production algorithms interact with the environment and the neural plasticity underlying the individual sensory-guided motor capacities. Niche construction from all ontogenetic processes modifies human selective environments. In turn, individual motor capacities within a population with social structures favouring transmission through iterative processes would result in a particular language phenotype. This integrated theory does not assume that acquired language regularities became innate or specific. Instead, language phenotypes would have evolved due to selection affecting multiple levels of all these mechanisms (this is represented in the graph with the right hand bracket). From generation to generation, language change occurs faster than other biological processes (this is represented in the causal graph with arrow thickness). Dashed lines represent permeability between developmental categories. Some relations of this causal graph have been designed following an EES framework (Lewens, 2019).

offer an alternative framework to understand language evolution and bring about new hypotheses to test the compatibility of biological and cultural explanations of language. First, cultural niche construction itself is a useful eco-evo-devo approach to fill the gap between different approaches to language evolution (e.g. biological/cultural). Second, it does not assume that acquired language regularities themselves ever become innate (Deacon, 2003, 2010) nor cultural responses automatically genetically encoded (Odling-Smee & Laland, 2009). And third, it gives room to integrate under the same umbrella the developmental molecular processes leading to the language capacity and the iterated learning processes of language transmission leading to the emergence of language universals.

In recent decades, advances in molecular biology and computational modelling have incredibly narrowed down the processes of ecological inheritance related with language structures. Language is endowed with complex regularities that can not be explained by learning nor transmission alone, nor by genetic encoding of human behaviours. And such complexity will probably remain unsolved for several more decades, or more. In the meantime, one observation seems clear, whatever approach we take to explain language development and evolution, it will necessarily have to consider the vast interconnectedness of genetic, ontogenetic and cultural factors that shape language.

6.7 Conclusions

In this chapter we have reviewed a number of studies that show that iterated learning does not require strong genetic constraints in the form of a domain-specific module to give rise to near language universals. Instead, general abilities unrelated with informational specificity, such as the ability to copy and to share, are required to develop language through cultural evolution. These general abilities can nevertheless yield specific properties, and might have emerged from a complex multifactorial cognitive human capacity that includes genes, cellular products, phenotypic states, environmental inputs and behavioural practices. Here we have argued that the developmental explanation of human abilities and iterated learning through cultural transmission are mutually dependent processes and therefore compatible, insofar as both are common processes and interact stabilising selection at different levels. We have used the notion of niche construction to sketch an integrated framework that builds bridges between evolutionary developmental accounts for sensory-guided motor capacities and cultural evolution guided by iterated learning models. This integrated model aims to overcome traditional boundaries between biological and cultural approaches in the debates of language evolution.

Chapter 7

Conclusion

In this thesis I used a variety of approaches to explore some specific interactions betwe individual's cognition and their environment. In particular, I developed computational models to investigate how the relationship between individual cognitive biases, value systems and memory (at the individual-level) and network connectivity dynamics and institutions (at the society-level) affect cultural diversity and the emergence of cultural conventions in a population. The aim of this thesis was to formalise some of the relationships between the social and cognitive features mentioned above in agent-based models, as well as to test some preliminary predictions against experimental data. Additionally, using a conceptual framework, I aimed at exploring to what extent recent findings from iterated learning, ecological evolutionary developmental biology and niche construction are consistent with each other and can be used to formulate more integrated frameworks for the explanation of general language regularities.

For this purpose, I first presented in Chapter 2 an agent-based model to explore how the interactions between population network connectivity dynamics, content bias, coordination bias and memory affect the spread of cultural variants in pairwise interactive micro-societies. Simulation results extend previous studies in social learning showing that content biases are important drivers of convergence. I also showed that connectivity dynamics affect the time-course of the spread of variants in moderate-high content biased populations: When populations take longer to reach full connectivity, convergence onto a single cultural variant is slowed. This effect is amplified by content bias. In addition, results show that larger memory size and coordination biases, especially egocentric bias, slow down convergence, in particular for moderate-low content biased populations. The results presented here are consistent with theoretical and experimental studies showing that convergence is driven by content biases Gong et al. (2012); Tamariz et al. (2014), and also agree with studies on rational learning in social networks showing that convergence is partially determined by the degree of connectivity in the social network (Barkoczi & Galesic, 2016; Centola & Baronchelli, 2015; Centola et al., 2007; Mueller-Frank, 2013; Olfati-Saber & Murray, 2004). A practical application of the model is that it can be used to fit real data obtained from turn-based cultural processes and might be helpful to improve the organisation of the turn taking by mitigating undesirable effects linked with one particular connectivity. Importantly, the specific manipulation of the network connectivity dynamic that we have studied here has not been taken into account in previous experimental work, and it is especially relevant for those researchers that use pairwise interactive micro-societies of agents switching partners over time (e.g. Baum et al., 2004; Byun et al., 2018; Caldwell & Smith, 2012; Fay et al., 2008, 2010; Mesoudi & Whiten, 2008; Raviv et al., 2019b; Tamariz et al., 2014).

In Chapter 3, I reported on an experiment in the lab in which participants engaged in a Pictionary-like graphical communication task as members of a 4-participant micro-society. The experiment had two main goals: First, to evaluate the effect of two network connectivity dynamics (early and late) on the evolution of the convergence of the micro-societies on shared communicative conventions. Second, to compare the predictions of the agent-based model described in Chapter 2 against experimental data, and calibrate the model to find the best-fitting parameter setting. Experimental results support the hypothesis that the convergence of a small-scale society of human participants on shared communicative conventions can be affected by the order in which connections between individuals unfold over time. In particular, late connectivity dynamic can slow down convergence and early connectivity dynamic can accelerate convergence. Although the study presented here differs from previous work on network connectivity dynamics in several respects (e.g. population size and experimental procedure), and it might be limited by the number of independent observations collected, results are consistent with previous findings regarding some important points:

(i) subtle changes in a population network structure can affect the formation of conventions (Centola & Baronchelli, 2015); and

(ii) social influence seems to reduce diversity in networks where every agent has equal connectivity, when compared to networks where agent's access to information is not equal (Becker et al., 2017).

Experimental results also show, contrary to model predictions, that differences between network connectivity dynamics did not tend to reduce during postalignment, suggesting that participants who are exposed to the same convention for longer develop greater resistance to change their variants. Finally, model evaluation against experimental data shows that the agent-based model developed in Chapter 2 is able to accurately predict convergence under the set of experimental observations using an early connectivity dynamic. However, model predictions were less accurate for late connectivity dynamics because they underestimated convergence during prealignment. This suggested that a more realistic implementation of values across cultural variants along with additional selection pressures at the population level would be useful to improve model performance.

Chapter 4 expands on the agent-based model developed in Chapter 2 by implementing additional features related to the agents' value systems and external reinforcement. I systematically manipulated the hegemony of value systems, the homogeneity of value systems in the population and institutional sensitivity to agents' choices. Simulation results show that diversity is lower in populations with homogeneous value systems and in highly hegemonic value systems. These effects are amplified by content bias. In scenarios with responsive institutions, in which the value systems emerge from the choices of the population, diversity is maintained to a higher degree than in institutions that reinforce an existing value system. Interestingly, diversity differences between institutional conditions are greater in high-hegemony value systems and homogeneous populations, particularly for intermediate levels of content bias. This latter point suggests that institutional intervention is more effective when the agents' behaviours are not extreme. However, one limitation of the model developed in Chapter 4 is that institutions are fixed and hence they are not part of a co-evolutionary model of cumulative culture.

In order to overcome some of the limitations of the model described in Chapter 4, in Chapter 5 I presented a co-evolutionary model that includes value systems, institutions and a comprehensive set of cognitive biases. This co-evolutionary model formalises the idea that cultural diversity changes with the processes of cognitive and institutional interaction. In general, institutional power facilitates the formation of cultural conventions, although it can increase cultural diversity when agents are compliant and institutions promote diverse values. In the absence of institutions, strong conventions can also emerge when agents have similar value systems and are endowed with strong individual-level confirmation biases. In some scenarios, where institutional power is intermediate, a moderate compliance bias was shown to be an important mechanism underlying the stabilization of cultural diversity at intermediate levels. These results might be of interest to researchers investigating the impact of representative institutions on the evolution of linguistic, cultural, economic or social behaviour (Becker et al., 2017; Bowles et al., 2003; Boyd & Richerson, 2008; Centola & Baronchelli, 2015; J. Henrich et al., 2005). In particular, an interesting research question for future work will be to explore in a formal model the relationships between conformity and compliance biases, how they interact and how they drive the emergence of cultural conventions in in-groups and out-groups. Future work will also be focused on mathematical modelling for a wider range of emergent institutions.

At the current historical moment, and in the midst of a health, economic and social crisis, one cannot help but think about the implications that the models presented here have for the transmission of behaviors and the strategies to manage the dissemination of information in the context of a globalised society. The spread of infections (e.g. simple contagion of a virus) has a clear connection with the particular behavior of each individual and with the dynamics of complex behavior (complex contagion) (Centola, 2018). The creation of models that have the ability to integrate particular aspects of the culture of each country, the social networks in which cultural conventions emerge and the actions of local and global institutions and governments will undoubtedly be fundamental tools to develop more effective public policies for information management in the near future.

I suggest that my models of co-evolution of value systems and institutions are very relevant for research projects concerning the preservation of cultural diversity. In recent years, the processes of cultural shift and the extinction of languages have accelerated (Krauss, 1992). The models that I have presented here are a useful tool to explore the interaction between individual biases and the behavior of institutions such as policy makers, crucial to understand the implications of globalization on the conservation of cultural forms, such as endangered languages. These models, in combination with experimental verification and validation are one way to look for possible ways to estimate the economic and social trade-offs involved in preserving cultural diversity. In other words, the present line of research is crucial to develop more integrated and coherent theoretical models of cultural evolution, and in turn help to better structure the bases upon which public policies for cultural preservation stand.

The work developed in this thesis (Chapters 2 to 5) build on a number of mathematical frameworks that couple individual and social parameters to construct models that yield population outcomes. These models can be used to explore the co-evolution of individual cognition and society-level structures, and yield simulation results that correspond to averaged metrics at the level of the population for both individual and collective patterns of behavior. Indeed, the use of population-level outcomes have dominated recent explanations of how cultural and biological phenomena evolve. In particular, evidence gathered using agent-based computer simulations have overwhelmingly shown that cultural evolution plays a more potent role than usually assumed, which implies less dependence on biases or innate structures (Steels, 2011). Agent-based models are also recognised as very useful tools that allow the explicit inclusion of individual variation in greater detail than do classical differential-equation models (DeAngelis & Mooij, 2005). However, despite the growing number of studies using agent-based models in recent years for simulating cultural processes, the inclusion of such variation in model parametrization remains a crucial issue for continued progress in ecological and evolutionary theory. According to DeAngelis and Mooij (2005), five major types of individual variation can be integrated in agent-based models: spatial, ontogenetic, phenotypic, cognitive, and genetic. While Chapters 2 to 5 placed the interaction between individuals' cognition and social structures at the centre of the picture to explaining the emergence of collective patterns of behavior at the population level, in Chapter 6 I also considered the importance of ontogenetic and phenotypic sources of variation. This approach implied an acknowledgement of what R. C. Lewontin (1983) defined as *variational explanations* (changes due to variation within the population) and *developmental explanations* (changes due to variation within the individual). Individuals have highly complex responses to the environment, and these responses include the incorporation of phenotypical changes that may affect individual learning and production algorithms over time. The reviewed literature from iterated learning, niche construction and ecological evolutionary developmental biology allowed an in-depth exploration of the compatibility between eco-evo-devo accounts for cognitive capacities and cultural evolution guided by iterated learning processes. By doing so, I showed how iterated learning models do not require strong genetic constraints in the form of a domain-specific module to give rise to language regularization. I also showed how this idea is coherent with a model that incorporates both niche construction (as a developmental process in which organisms and environment co-construct each other), and cultural evolution, where phenotypes (such as the language phenotype) have a more active role in evolution. The proposed conceptual model might be useful to act as a hypothesis-generating framework around which cognitive scientists can structure new triple-inheritance formal models.

In sum, this thesis explored the relationships between specific features of human cognition and specific features of the social structure. I formalised some of the relationships between a number of relevant cognitive biases in social influence, memory and value systems (at the individual level) and specific network connectivity dynamics and institutions (at the level of the social structure). The results summarised above suggest two potential positive correlations that could drive the co-evolution of human cognition and human niches. Firstly, cognitive biases, and in particular content bias, are important drivers of the convergence of the population on shared cultural conventions, and they can dramatically amplify the effects of the network connectivity dynamics on the time-course of the spread of variants. Secondly, the co-evolution of institutions and value systems reveals a complex relationship between cognitive biases and cultural diversity. Institutions can facilitate or inhibit the emergence of cultural conventions depending on the diversity of the value systems they promote and the combination of cognitive biases in the population. Confirmation bias in the absence of institutions and compliance bias in the presence of strong institutions, seem to act as potent mechanisms for the formation of cultural conventions.

Finally, as mentioned above, this thesis formalised complex systems into agent based models and explored the possibility of new approaches with an integrative spirit. Indeed, culture and language are shaped by interrelated features at the individual and the socio-ecological level that can not be easily unraveled. Much remains to be discovered in the coming future about how these different levels interact affecting the evolution of cultural conventions. And one thing is sure, we will need more models, experiments and theory to keep learning about ourselves and others. In the end, the evolution of cultural conventions necessarily depends on us and on the things we do with others again and again.

Appendices

Appendix A

Content bias, coordination bias and memory

A.1 Content bias

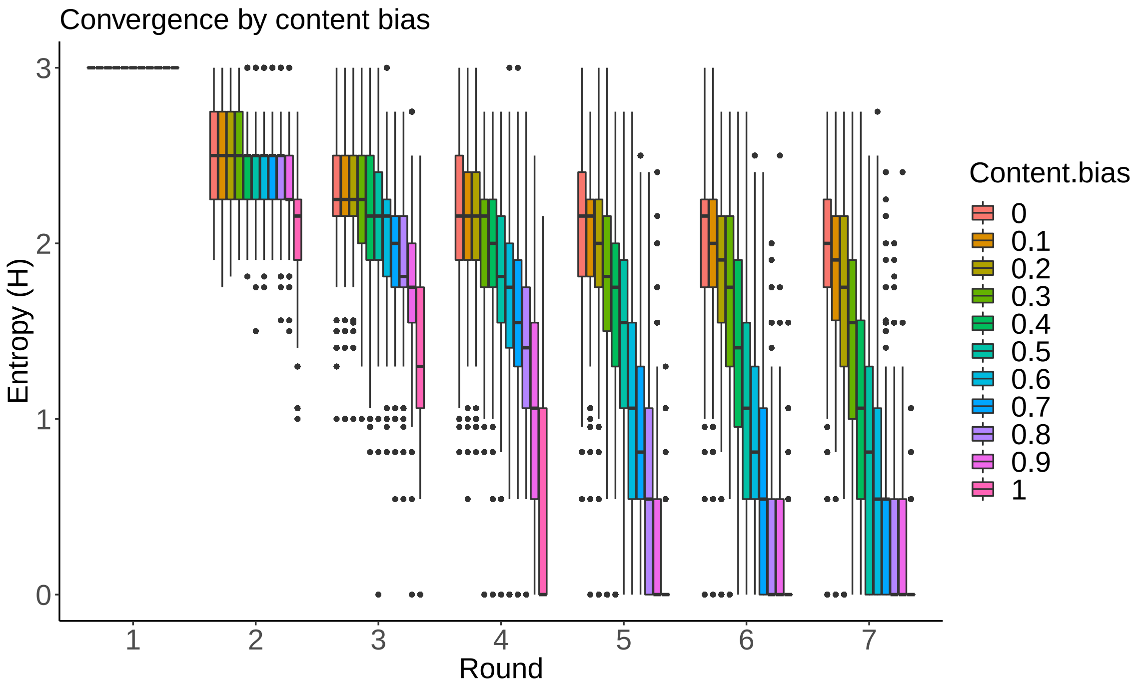


Figure A.1: Entropy (H) by round by content bias. Data subset of neutral coordination bias. In this plot, a drift model corresponds to content bias $= 0$. In this and subsequent boxplots: middle line is median, 50% quantile; lower hinge, 25$ quantile; upper hinge, 75% quantile; lower whisker is smallest observation greater than or equal to lower hinge $-$ $1.5 * IQR$; upper whisker is largest observation less than or equal to upper hinge $+ 1.5 * IQR$.

One of the advantages of our model is that the agents' responses are limited to a set of signals, so the number of permitted values of entropy is finite. It is therefore possible to accurately estimate conditional densities describing how the distribution of entropy changes over levels of content bias and over time. We compute conditional probability distributions of entropy by calculating the probability distribution of three entropy groups: high ($H \geq 2$ bits), medium (1 bit $\leq H < 2$ bits) and low ($H < 1$ bit). The probability of a non-convergent communication system (entropy higher than 2 bits) when content bias is 0 is approximately 99% by round 2:

$$P(H\,\text{high} \mid \text{Cont} = .0, \text{Gen} = 2) = 0.994,$$

and it drops to 50% by round 7:

$$P(H\,\text{high} \mid \text{Cont} = .0, \text{Round} = 7) = 0.504.$$

When the content bias is 1, the probability of a non-convergent communication system is 68% by round 2, and almost 0% by round 7. On the contrary, the probability of a conventionalised communication system (entropy lower that 1 bit) when content bias is 0 goes from 0% to approximately 0.04%, whereas when content bias is 1 goes from 0% to 99%. More details in Figure A.2.

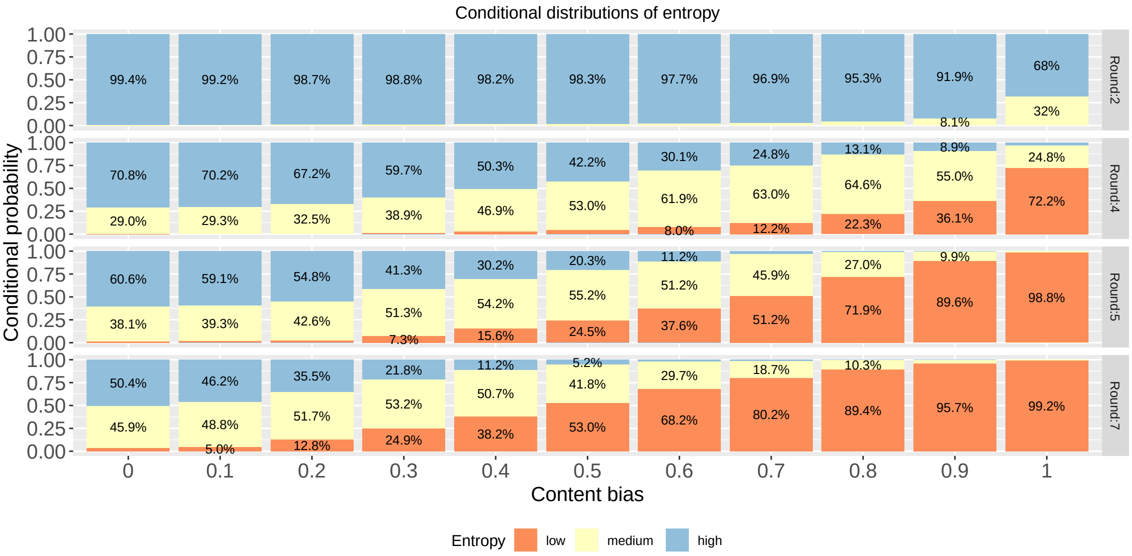


Figure A.2: Conditional distributions of entropy across levels of content bias: $P(y \mid x)$ against x. Bars represent conditional distributions of entropy (H): In blue high entropy ($H \geq 2$ bits), in yellow medium entropy (1 bit $\leq H < 2$ bits), in red low entropy ($H < 1$ bit). X-axis represents level of content bias, and y-axis represents conditional probability. For example, the probability of high entropy (low convergence) given that content bias is 0.5 ($P(H\,\text{high} \mid \text{Cont} = .5)$) goes from 0.9833 in round 2, to 0.0525 in round 7.

A.2 Coordination bias

Over rounds, the relationship between entropy and coordination bias remains characterised by a clear asymmetric distribution. When compared with neutral coordination, strong egocentric and allocentric behaviours reduce the entropy drop, and maintain variant diversity in the system. In the absence of content bias, entropy decreases only slightly towards its horizontal asymptote, which is highest when egocentric bias is strongest (Figure A.3 and Figure A.4 on the next page). Our results agree with Segovia-Martín et al. (2019) in recognizing a weak effect of coordination bias when content bias is strong, and a stronger effect of coordination bias when content bias is weak. Also, in both models, egocentric bias maintains diversity better than allocentric bias. However, when compared with a neutral model, allocentric bias reduced convergence in the current model, while increased it in Segovia-Martín et al. (2019). This is likely due to different implementations of variant quality distributions. We will explore the effect of variant quality on convergence in future research.

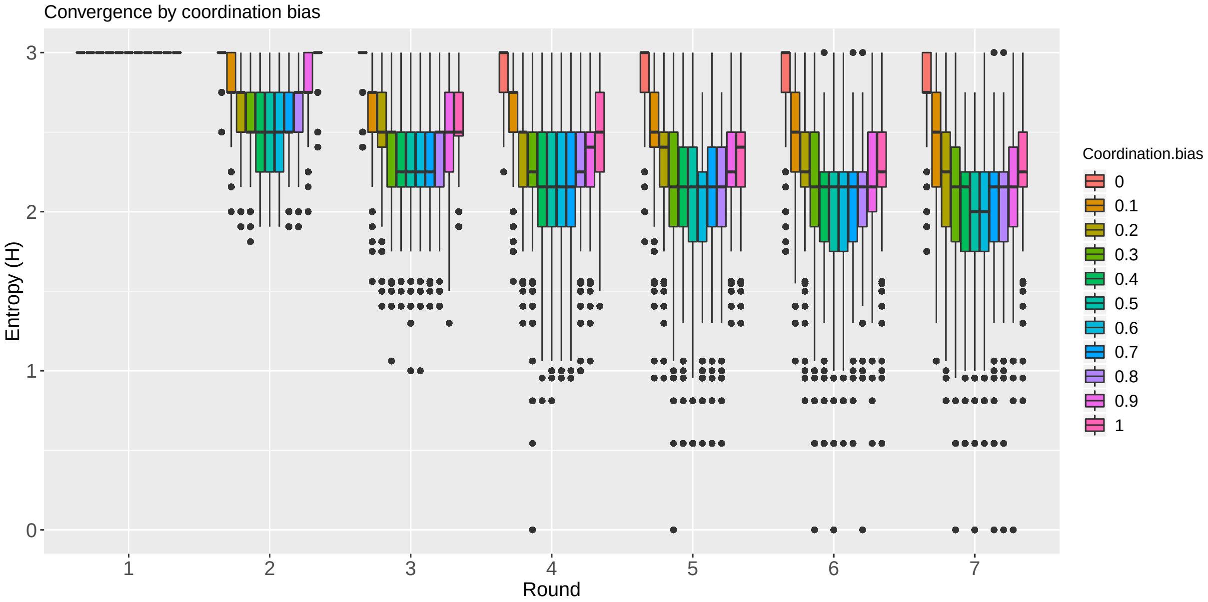


Figure A.3: Entropy (H) by round by coordination bias. Data subset of neutral content bias. In this plot, a drift model corresponds to coordination bias = 0.5.

The probability of a non-convergent communication system—entropy lower than 1 bit—when coordination bias is 0 (fully egocentric) remains close to 100% over rounds. When coordination bias is 1 (fully allocentric) the probability of a non-convergent communication system is lower but still quite high (79.33% by round 7). The probability of a non-convergent communication system is lowest when coordination bias is neutral (Figure A.4).

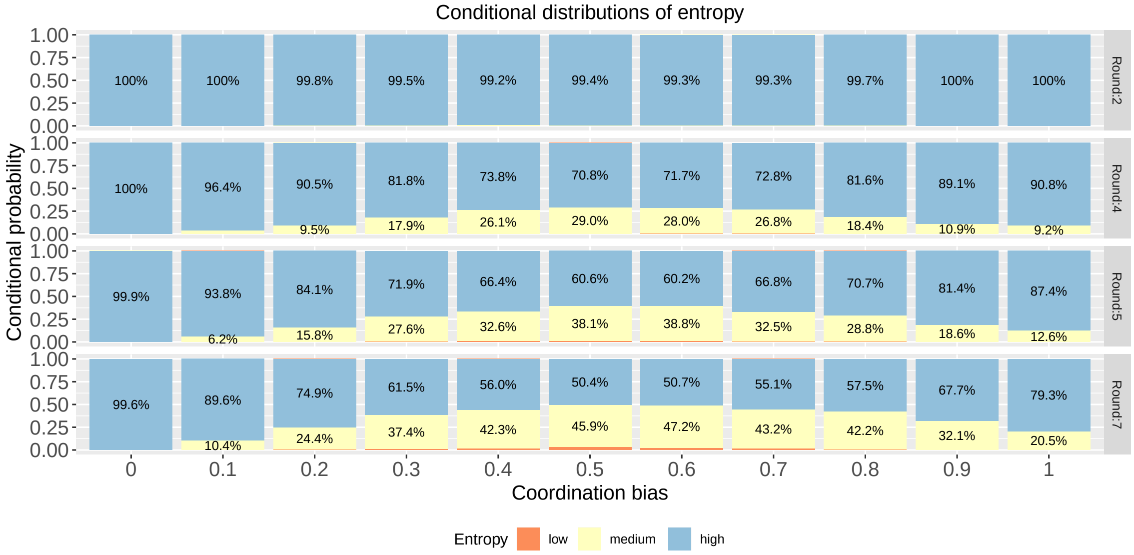


Figure A.4: Conditional distributions of entropy across levels of coordination bias: $P(y \mid x)$ against x. Bars represent conditional distributions of entropy: In blue high entropy ($H \geq 2$ bits), in yellow medium entropy (1 bit $\leq H < 2$ bits), in red low entropy ($H < 1$ bit). X-axis represents level of coordination bias, and right axis represents conditional probability.

A.3 Interactions between Content bias and Coordination bias

In all the rounds, the probability of a non-convergent communication system is higher when content bias is 0 (i.e. neutral) and coordination bias is 0 (i.e. strong egocentric bias). In contrast, the probability of a conventionalised communication system is higher when content bias is 1 (i.e., strong content bias) and coordination bias is 0.5 (i.e. neutral). Conditional distributions of entropy for all levels of coordination bias tend to equalise as content bias increases. A detailed multi-graph of conditional distributions can be found in Figure A.8 on page 174.

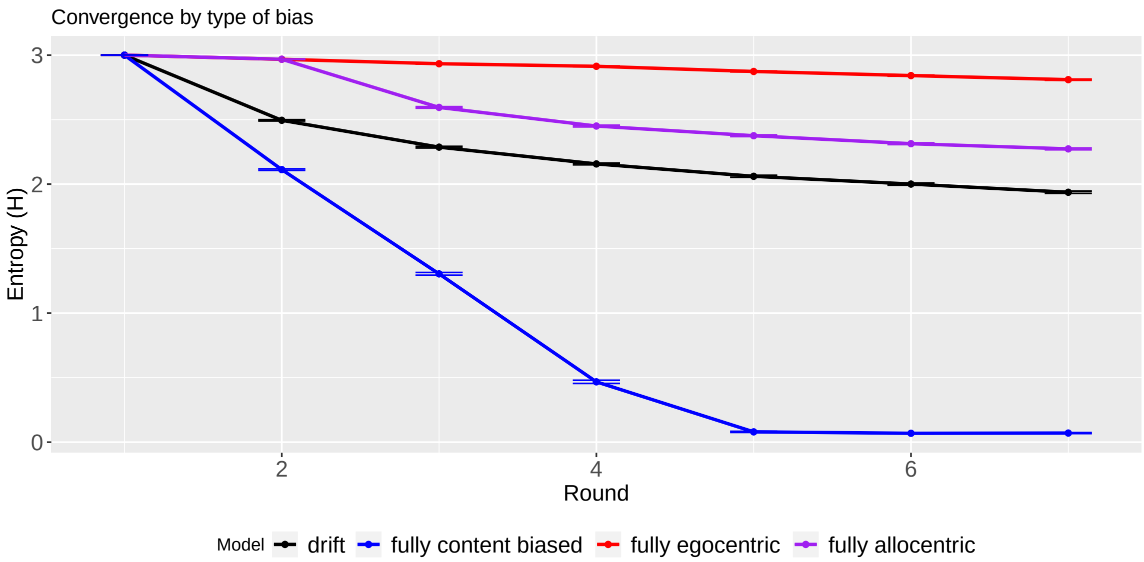


Figure A.5: Entropy by round by type of bias, averaged over the global data set, error bars indicate 95% CIs. Red: Strongest egocentric bias when content bias is neutral. Purple: Strongest allocentric bias when content bias is neutral. Blue: Strongest content bias when coordination bias is neutral. Black: Drift model (neutral content bias and neutral coordination bias). When compared to a drift model, content bias increases convergence while coordination bias, especially egocentric bias, decreases convergence.

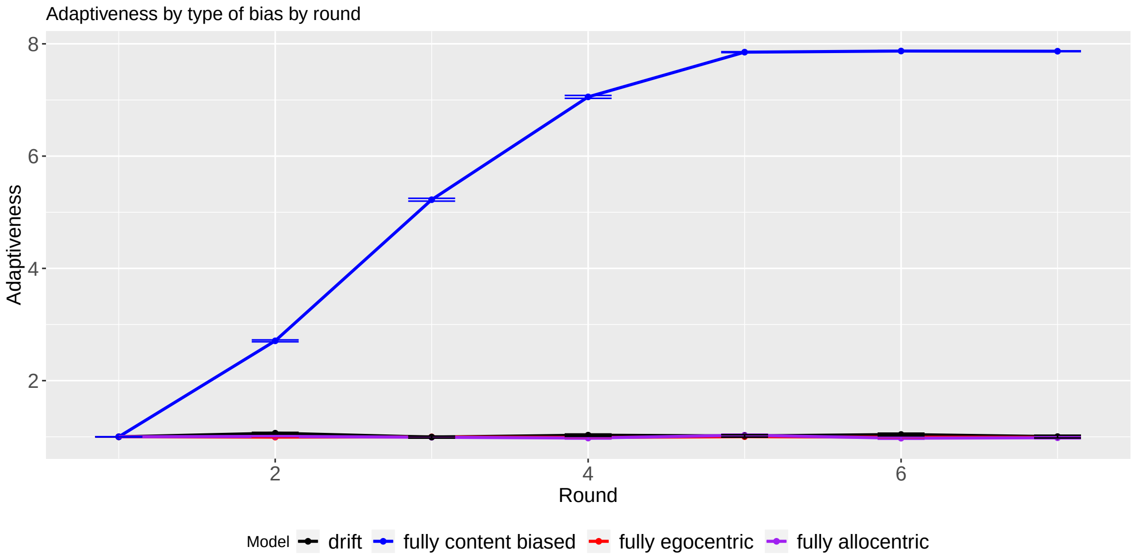


Figure A.6: Adaptiveness by round by type of bias, averaged over the global data set, error bars indicate 95% CIs. Red: Strongest egocentric bias when content bias is neutral. Purple: Strongest allocentric bias when content bias is neutral. Blue: Strongest content bias when coordination bias is neutral. Black: Drift model (neutral content bias and neutral coordination bias). When compared to a drift model, only content bias increases adaptiveness.

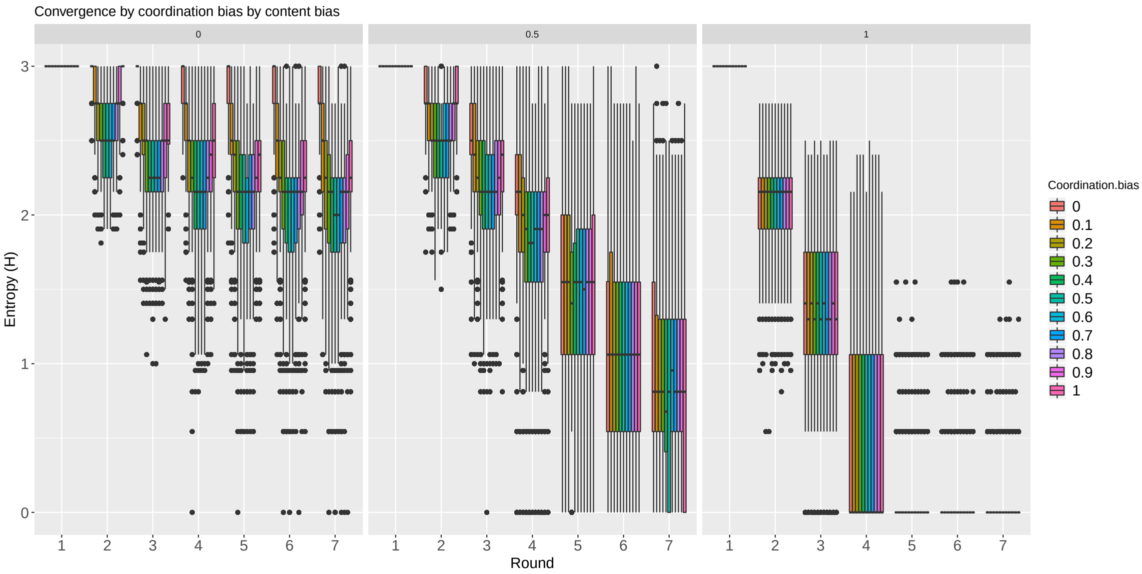


Figure A.7: Entropy (H) by round by each combination of biases. Examples for content bias 0, 0.5 and 1. A drift model has a content bias of 0 and a coordination bias of 0.5. X-axis represents rounds from 1 to 7, Y-axis represents entropy in bits.

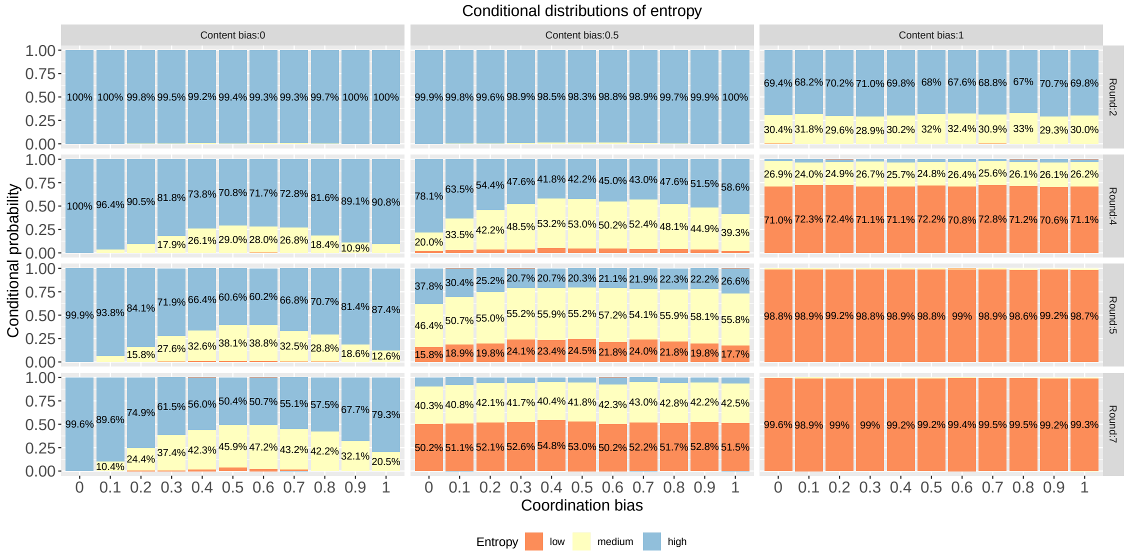


Figure A.8: Conditional distributions of entropy across levels of coordination bias: $P(y \mid x)$ against x. Bars represent conditional distributions of entropy: In blue high entropy ($H \geq 2$ bits), in yellow medium entropy (1 bit $\leq H < 2$ bits), in red low entropy ($H < 1$ bit). X-axis represents level of coordination bias, and right axis represents conditional probability.

A.4 Memory

The probability of a non-convergent communication system given a memory of 1 round drops from 96.04% in round 2 to 10.53% in round 7. Given memories of 3, 5 and 7 rounds, it drops from about 96% in round 2 to 16.46%, 24.40% and 28.62% in round 7, respectively. Differences between levels of memory are higher at round 7. The probability of a convergent communication system given a memory span of one round is 69.10%. Given a memory span of 3 rounds, it is 55.26%, 5 rounds 42.58% and 7 rounds 37.54% (see Figure A.11 on the following page).

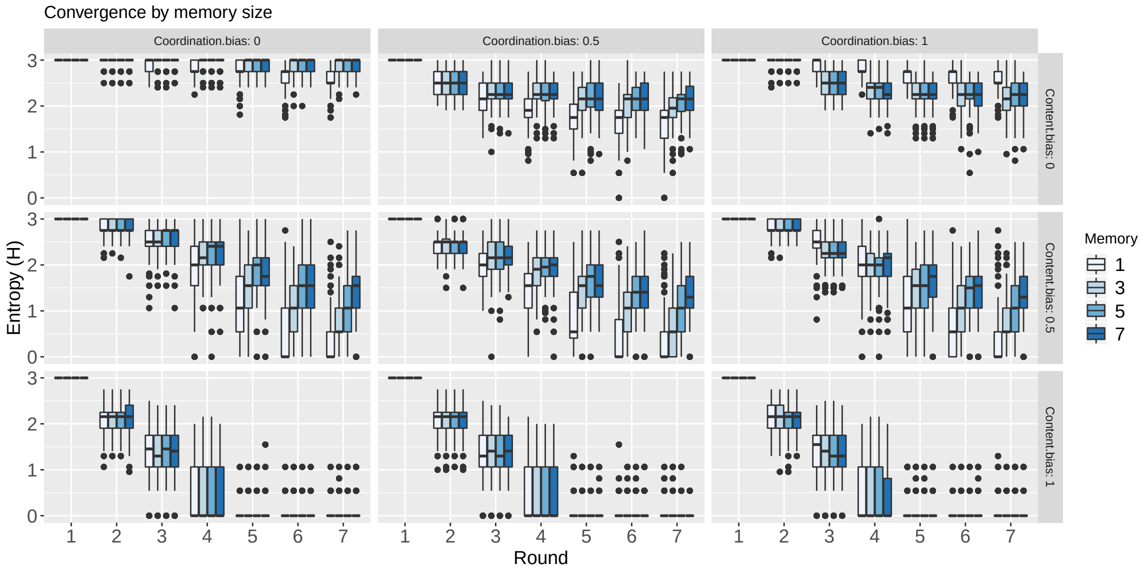


Figure A.9: Entropy (H) by round by each level of memory. X-axis represents rounds from 1 to 7, Y-axis represents entropy in bits. Coordination bias = 0 is fully egocentric and Coordination bias = 1 is fully allocentric. A drift model is represented by Coordination bias = .5 and Content bias = 0.

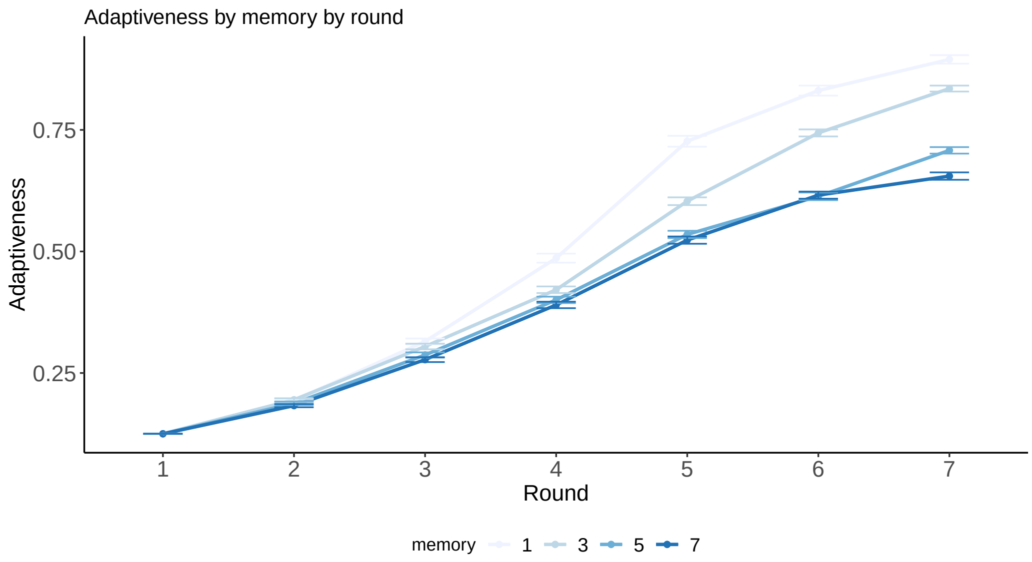


Figure A.10: Adaptiveness by round by each level of memory, averaged over the global data set, error bars indicate 95% Cis. Memory size decreases adaptiveness of the cultural system when compared with memory = 1.

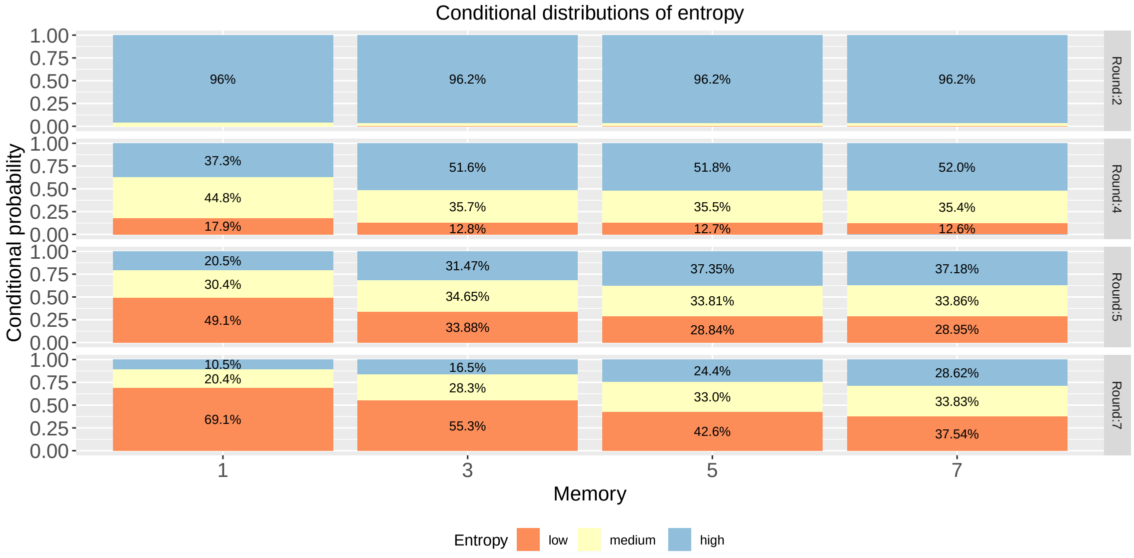


Figure A.11: Conditional distributions of entropy across levels of memory: $P(y \mid x)$ against x. Bars represent conditional distributions of entropy: In blue high entropy ($H \geq 2$ bits), in yellow medium entropy (1 bit $\leq H < 2$ bits), in red low entropy ($H < 1$ bit). X-axis represents level of coordination bias, and right axis represents conditional probability.

A.5 Time to convergence

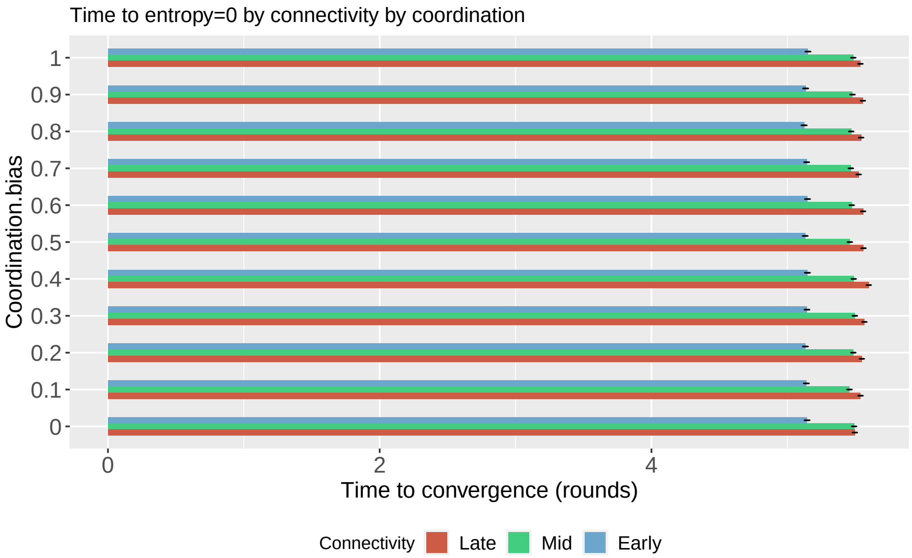


Figure A.12: Time to convergence (number of rounds until entropy is lower than 0) averaged over each level of connectivity and coordination bias after 7 rounds. Error bars indicate 95% CIs.

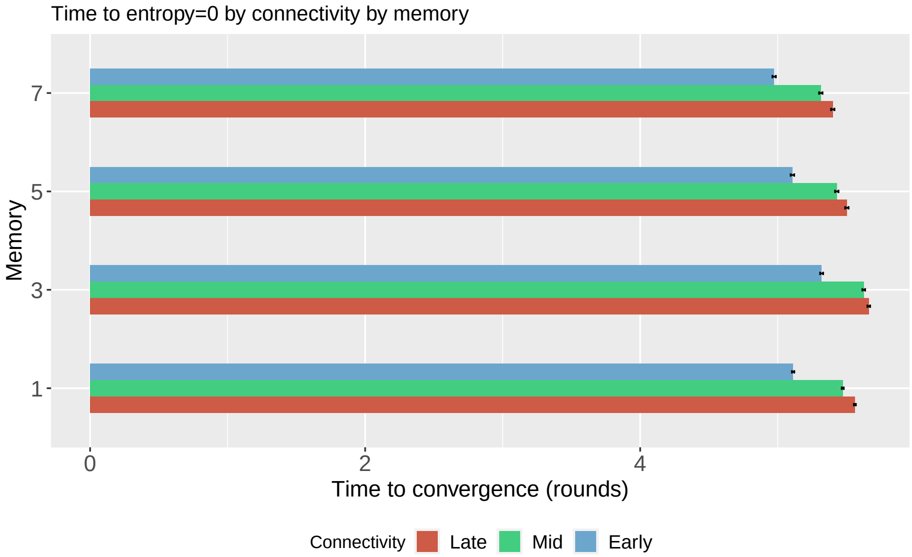


Figure A.13: Time to convergence (number of rounds until entropy is lower than 0) averaged over each level of connectivity and memory after 7 rounds. Error bars indicate 95% CIs.

Appendix B

Probabilistic model

Tamariz et al. (2014) constructed a parametrised model of participant variant choice. The model takes as input the history of the representational variants the participant had used or seen a partner use and returns a distribution over how they might next represent that concept.

$$\Pr(\sigma \mid h) = \mu'\beta'c'f(\sigma \mid h_{|E,3}) + \mu'\beta'cf(\sigma \mid h_{|A,3}) + \mu'\beta s\varepsilon + \mu\phi(\sigma) \quad \text{(B.1)}$$

The model takes four parameters as described in the main text, plus a coordination bias (c) which fixes the likelihood of being copied ascribed to variants produced by others and witnessed by the participant, and the variants produced by the participant themself. It takes values ranging from 0 (fully egocentric: Preferring self-produced variants over other-produced variants) to $+1$ (fully allocentric: Preferring other-produced variants over self-produced variants). Zero coordination bias treats variants in $h_{|E,m}$ and in $h_{|A,m}$ as equally worthy of reproduction, i.e. unbiased.

Together the parameters define the probability distribution shown in equation B.1, varying over potential representational variants x, for a given history h.

In this study, we control for coordination bias by using a drift model (coordination bias 0.5), which means that variants in $h_{|E,3}$ and in $h_{|A,3}$ were equally likely to be produced. This is equivalent - if we consider a neutral coordination bias - to:

$$\Pr(\sigma \mid h) = \frac{\mu'\beta'f(\sigma \mid h_{|E,3})}{2} + \frac{\mu'\beta'f(\sigma \mid h_{|A,3})}{2} + \mu'\beta s + \mu\phi(\sigma) \quad \text{(B.2)}$$

And, if we unify allocentric and egocentric memories, we have:

$$\Pr(\sigma \mid h) = \mu'\beta'f(\sigma \mid h_{|M,3}) + \mu'\beta s + \mu\phi(\sigma) \quad \text{(B.3)}$$

where s encompasses a variable that that takes one out of the 9 value systems examined, β corresponds to the level of content bias examined and $Pr(\sigma \mid h)$ yields a probability distribution of variants (σ) for a given history (h) of previous rounds. The overbar denotes the probabilistic complement (e.g. $\bar{a} = 1 - a$).

Appendix C

Markov chains and language transmission

A Markov chain is a sequence of random variables $X_0, X_1, \ldots, X_n$ such that,

$$P(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, ..., X_n = x_n) \tag{C.1}$$

and,

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n) \tag{C.2}$$

This means that the distribution of the forthcoming state X_{n+1} depends only on the current state X_n and is independent of all its predecessors.

The set of possible states of the Markov chain is $S = \{s_1, s_2, \ldots, s_r\}$, and the probability to move from state s_i to state s_j is named *transition probability* p_{ij}:

$$p_{ij} = P(X_1 = s_j \mid X_0 = s_i) \tag{C.3}$$

If the Markov chain is homogeneous, meaning that the underlying transition probabilities remain constant over time, then:

$$P(X_{n+1} = s_j \mid X_n = s_i) = P(X_n = s_j \mid X_{n-1} = s_i) \tag{C.4}$$

We can represent the distribution of transition probabilities with a *transition matrix* $T = (p_{ij})$, where each element in the matrix represents the transition probability p_{ij} from state s_i to s_j:

$$T = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{C.5}$$

To provide a numerical example of an iterated learning process, consider a Markov chain with 2 possible states (e.g. two possible languages l_1

and l_2):

$$T = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} \tag{C.6}$$

Here, $p_{11} = 0.8$ is the probability that $X_1 = s_1$ given that we observed $X_0 = s_1$, and so on. In the case of language transmission, p_{11} represents the probability that a learner acquires l_1 from data produced form l_1. p_{12} represents the probability that a learner acquires l_1 from data produced form l_2, and so on.

Assume that the current language of the chain is $X_0 = l_2$, e.g. $l^0 = \begin{pmatrix} 0 & 1 \end{pmatrix}$. Then, the probability distribution of languages after 1 step is:

$$l^1 = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{pmatrix} 0.1 & 0.9 \end{pmatrix} \tag{C.7}$$

After 2 steps, since $l^n = l^{n-1}T$, we have:

$$l^2 = \begin{pmatrix} 0.1 & 0.9 \end{pmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{pmatrix} 0.17 & 0.83 \end{pmatrix} \tag{C.8}$$

which means that the stationary probability of each of the two languages is determined by the fidelity of transmission in the Markov chain. Thus, the stationary probability of l_1 in the second step is 0.17, $P(X_2 = l_1 \mid X_0 = l_2) = 0.17$, while the probability of l_2 is $P(X_2 = l_2 \mid X_0 = l_2) = 0.83$.

References

Acerbi, A., Van Leeuwen, E. J., Haun, D. B., & Tennie, C. (2016). Conformity cannot be identified based on population-level signatures. *Scientific reports*, *6*, 36068.

Ager, P., & Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the us during the age of mass migration. *European Economic Review*, *64*, 76–97.

Amundson, R. (2005). *The changing role of the embryo in evolutionary thought: roots of evo-devo*. Cambridge: Cambridge University Press.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, *33*(4), 245–266.

Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological monographs: General and applied*, *70*(9), 1.

Ashraf, Q., & Galor, O. (2011). *Cultural diversity, geographical isolation, and the origin of the wealth of nations* (Tech. Rep.). National Bureau of Economic Research.

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., & Pagel, M. (2008). Languages evolve in punctuational bursts. *Science*, *319*(5863), 588–588.

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

Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., & Lemey, P. (2012). Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Molecular biology and evolution*, *30*(2), 239–243.

Balari, S., & Lorenzo, G. (2013). *Computational phenotypes: Towards an evolutionary developmental biolinguistics*. Oxford: Oxford University Press.

Balari, S., & Lorenzo, G. (2015). It is an organ, it is new, but it is not a new organ. conceptualizing language from a homological perspective. *Frontiers in Ecology and Evolution*, *3*, 58. doi: 10.3389/fevo.2015.00058

Balari, S., & Lorenzo, G. (2018). The internal, the external and the hybrid: The state of the art and a new characterization of language as a natural object. *Glossa. A Journal of General Linguistics*, *3*, 22. doi: https://doi.org/10.5334/gjgl.330

Baldwin, J. M. (1896). A new factor in evolution. *The American Naturalist*, *30*, 354, 422, 441–451, 536–553. Retrieved from `https://www.jstor.org/stable/2453130`

Barabási, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, *272*(1-2), 173–187.

Barabâsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, *311*(3-4), 590–614.

Barkoczi, D., & Galesic, M. (2016). Social learning strategies modify the effect of network structure on group performance. *Nature communications*, *7*, 13109.

Barkow, J. H. (1992). Beneath new culture is old psychology: Gossip and social stratification. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 627–637). Oxford: Oxford University Press.

Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in Neuroscience*, *10*(1–2), 136–148.

Baum, W. M., Richerson, P. J., Efferson, C. M., & Paciotti, B. M. (2004). Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evolution and Human Behavior*, *25*(5), 305–326.

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the national academy of sciences*, *114*(26), E5070–E5076.

Bellini, E., Ottaviano, G. I., Pinelli, D., & Prarolo, G. (2013). Cultural diversity and economic performance: evidence from european regions. In *Geography, institutions and regional economic performance* (pp. 121–141). Springer.

Benítez-Burraco, A., & Longa, V. M. (2010). Evo-Devo — Of course, but which one? Some comments on Chomsky's analogies between the biolinguistic approach and Evo-Devo. *Biolinguistics*, *4*(4), 308–323. Retrieved from `https://www.biolinguistics.eu/index.php/biolinguistics/article/view/175`

Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1547), 1443–1450.

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge, MA: The MIT press.

Berwick, R. C., & Chomsky, N. (2017). Why only us: Recent questions and answers. *Journal of Neurolinguistics*, *43*(Part B), 166–177. doi: https://doi.org/10.1016/j.jneuroling.2016.12.002

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science*, *35*(7), 1207–1242. doi: 10.1111/j.1551-6709.2011.01189.x

Bhavnani, R. (2003). Adaptive agents, political institutions and civic traditions in modern italy. *Journal of Artificial Societies and Social Simulation*, *6*(4).

Bickerton, D. (2003). Symbol and structure: a comprehensive framework for language evolution. *Studies in the Evolution of Language*, *3*, 77–93.

Bickerton, D. (2009). *Adam's tongue: How humans made language, how language made humans*. New York: Hill and Wang.

Bickerton, D. (2014). *More than nature needs: Language, mind, and evolution*. Cambridge, MA: Harvard University Press.

Bishop, J. H., & Wößmann, L. (2004). Institutional effects in a simple model of educational production. *Education Economics*, *12*(1), 17–38.

Björk, J., & Magnusson, M. (2009). Where do good innovation ideas come from? exploring the influence of network connectivity on innovation idea quality. *Journal of Product Innovation Management*, *26*(6), 662–670.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of child language*, *14*(2), 211–227.

Boeckx, C. (2014). What can an extended synthesis do for biolinguistics: On the needs and benefits of eco-evo-devo program. In M. Pina & N. Gontier (Eds.), *The evolution of social communication in primates* (pp. 313–326). Cham: Springer.

Bonduriansky, R., & Day, T. (2018). *Extended heredity. a new understanding of inheritance and evolution*. Cambridge: Princeton University Press.

Bormann, E. G. (1996). Symbolic convergence theory and communication in group decision making. *Communication and group decision making*, *2*, 81–113.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of economic literature*, *36*(1), 75–111.

Bowles, S. (2000). Economic institutions as ecological niches. *Behavioral and Brain Sciences*, *23*(1), 148–149.

Bowles, S., Choi, J.-K., & Hopfensitz, A. (2003). The co-evolution of individual behaviors and social institutions. *Journal of theoretical biology*, *223*(2), 135–147.

Boyd, R., & Richerson, P. (1985). Culture and the evolutionary process. chicago: Univ. *Press, Chicago*.

Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford University Press.

Boyd, R., & Richerson, P. J. (2008). Gene–culture coevolution and the evolution of social institutions. *Philpapers.org*.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*(2), 163–197.

Brémaud, P. (1999). *Markov chains. gibbs fields, monte carlo simulation, and queues*. New York: Springer.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial life*, *8*(1), 25–54. doi: 10.1162/106454602753694756

Byun, K.-S., De Vos, C., Roberts, S. G., & Levinson, S. C. (2018). Interactive sequences modulate the selection of expressive forms in cross-signing. In *the 12th international conference on the evolution of language:(evolang xii)* (pp. 67–69).

Caldwell, C. A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, *29*(3), 165–171.

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PloS one*, *7*(8), e43807.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, *22*(2), 249–254.

Carroll, S. B. (2005). *Endless forms most beautiful: The new science of evo devo and the making of the animal kingdom*. New York: WW Norton & Co.

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, *134*(1), 25–36. doi: https://doi.org/10.1016/j.cell.2008.06.030

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach* (No. 16). Princeton University Press.

Centola, D. (2010). The spread of behavior in an online social network experiment. *science*, *329*(5996), 1194–1197.

Centola, D. (2018). *How behavior spreads: The science of complex contagions* (Vol. 3). Princeton University Press.

Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of*

Sciences, *112*(7), 1989–1994.

Centola, D., Gonzalez-Avella, J. C., Eguiluz, V. M., & San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, *51*(6), 905–929.

Charbonneau, M. (2016). Evo-devo and culture. In L. Nuño de la Rosa & G. B. Müller (Eds.), *Evolutionary developmental biology: A reference guide.* Cham: Springer. doi: 10.1007/978-3-319-33038-9_47-1

Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences*, *106*(4), 1015–1020. doi: https://doi.org/10.1073/pnas.0807191106

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in cognitive sciences*, *7*(1), 19–22.

Chavalarias, D. (2017). What's wrong with science? *Scientometrics*, *110*(1), 481–503.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese*, *17*, 2–11.

Chomsky, N. (1986). *Knowledge of language. its nature, origin and use.* New York: Praeger.

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: The MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*(1), 1–22.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland & H.-M. Gärtner (Eds.), *Interfaces + recursion = language? chomsky's minimalism and the view from syntax-semantics* (pp. 1–29). Berlin: Mouton de Gruyter.

Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of language: Biolinguistic perspectives* (pp. 45–62). Cambridge: Cambridge University Press.

Christiansen, M. H., Dale, R. A., Ellefson, M. R., & Conway, C. M. (2002). The role of sequential learning in language evolution: Computational and experimental studies. In *Simulating the evolution of language* (pp. 165–187). Springer.

Christiansen, M. H., & Kirby, S. (2003). *Language evolution.* OUP Oxford.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591–621.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37–46.

Creanza, N., Kolodny, O., & Feldman, M. W. (2017). Greater than the sum of its parts? modelling population contact and interaction of cultural repertoires. *Journal of The Royal Society Interface*, *14*(130), 20170171.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, *6*, 1964. doi: 10.3389/fpsyg.2015.01964

Cuskley, C., Loreto, V., & Kirby, S. (2018). A social approach to rule dynamics using an agent-based model. *Topics in cognitive science*, *10*(4), 745–758.

Dacin, M. T. (1997). Isomorphism in context: The power and prescription of institutional norms. *Academy of management journal*, *40*(1), 46–81.

Darwin, C. (1871/1981). The descent of man, and selection in relation to sex. 1871. *Princeton: Princeton UP.*

David, P. A. (1994). Why are institutions the 'carriers of history'?: Path dependence and the evolution of conventions, organizations and institutions. *Structural change and* *economic dynamics*, *5*(2), 205–220.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: WW Norton & Co.

Deacon, T. W. (2003). Multilevel selection in a complex adaptive system: the problem of language origins. In B. H. Weber & D. J. Depew (Eds.), *Evolution and learning: The baldwin effect reconsidered* (pp. 81–106). Cambridge, MA: The MIT Press.

Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 9000–9006. doi: https://doi.org/10.1073/pnas.0914624107

DeAngelis, D. L., & Mooij, W. M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecol. Evol. Syst.*, *36*, 147–168.

DeGraff, M. (2001). *Language creation and language change: Creolization, diachrony, and development*. MIT Press.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific reports*, *7*, 40391.

Derex, M., Beugin, M.-P., Godelle, B., & Raymond, M. (2013). Experimental evidence for the influence of group size on cultural complexity. *Nature*, *503*(7476), 389.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987.

Derex, M., Perreault, C., & Boyd, R. (2018). Divide and conquer: intermediate levels of population fragmentation maximize cultural accumulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170062.

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is "huh?" a universal word? conversational infrastructure and the convergent evolution of linguistic items. *PloS one*, *8*(11), e78273.

Dixon, R., & Robert Malcolm Ward, D. (1997). *The rise and fall of languages*. Cambridge University Press.

Durham, W. H. (1991). *Coevolution: Genes, culture, and human diversity*. Stanford University Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: The MIT press.

Fay, N., & Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PloS one*, *8*(8), e71781.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological science*, *11*(6), 481–486.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3553–3561.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive science*, *34*(3), 351–386.

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158–180.

Feldman, M. W., & Cavalli-Sforza, L. L. (1976). Cultural and biological evolutionary processes, selection for a trait under complex transmission. *Theoretical Population Biology*, *9*(2), 238–259. doi: https://doi.org/10.1016/0040-5809(76)90047-2

Ferdinand, V., Thompson, B., Kirby, S., & Smith, K. (2013). Regularization behavior in a non-linguistic domain. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature genetics*, *18*(2), 168. doi: 10.1038/ng0298-168

Fisher, S. E., & Vernes, S. C. (2015). Genetics and the language sciences. *Annual Review of Linguistics*, *1*, 289–310. doi: https://doi.org/10.1146/annurev-linguist-030514-125024

Fitch, W. T. (2008). Glossogeny and phylogeny: cultural evolution meets genetic evolution. *Trends in genetics*, *24*(8), 373–374.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: The MIT press.

Fodor, J. A. (1985). Précis of *The Modularity of Mind. Behavioral and Brain Sciences*, *8*(1), 1–5. doi: https://doi.org/10.1017/S0140525X0001921X

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology.* Cambridge, MA: The MIT press.

Fodor, J. A., & Lepore, E. (2002). *The compositionality papers.* Oxford: Oxford University Press.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of economic perspectives*, *7*(2), 159–171.

Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics.* Cambridge university press.

French, J. R. P. (1956). A formal theory of social power. *Psychological review*, *63*(3), 181.

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, *15*(3-4), 193–206.

Fuentes, A., Wyczalkowski, M. A., & MacKinnon, K. C. (2010). Niche construction through cooperation: a nonlinear dynamics contribution to modeling facets of the evolutionary history in the genus homo. *Current Anthropology*, *51*(3), 435–444.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological science*, *23*(8), 931–939.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science*, *40*(1), 145–171.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Garcia, A. C., & Baker Jacobs, J. (1999). The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on language and social interaction*, *32*(4), 337–367.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3), 181–215.

Garrod, S., & Pickering, M. J. (2007). Alignment in dialogue. *The Oxford handbook of psycholinguistics*, 443–451.

Giddens, A. (2013). *New rules of sociological method: A positive critique of interpreta-*

tive sociologies. John Wiley & Sons.

Gilbert, S. F., Bosch, T. C., & Ledón-Rettig, C. (2015). Eco-evo-devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics*, *16*(10), 611–622.

Gilbert, S. F., & Epel, D. (2009). Ecological developmental biology: Integrating epigenetics, medicine, and evolution.

Gilbert, S. F., & Epel, D. (2015). *Ecological developmental biology: The environmental regulation of development, health, and evolution*. Sunderland, MA: Sinauer.

Godfrey-Smith, P., & Martínez, M. (2013). Communication and common interest. *PLoS computational biology*, *9*(11).

Gold, E. M. (1967). Language identification in the limit. *Information and control*, *10*(5), 447–474. doi: https://doi.org/10.1016/S0019-9958(67)91165-5

Gong, T., Minett, J. W., & Wang, W. S.-Y. (2008). Exploring social structure effect on language evolution based on a computational model. *Connection Science*, *20*(2-3), 135–153.

Gong, T., Shuai, L., Tamariz, M., & Jäger, G. (2012). Studying language change using price equation and pólya-urn dynamics. *PLoS One*, *7*(3), e33171.

González-Avella, J. C., Cosenza, M. G., Klemm, K., Eguíluz, V. M., & San Miguel, M. (2007). Information feedback and mass media effects in cultural dynamics. *Journal of Artificial Societies and Social Simulation*, *10*(3), 9. Retrieved from `http://jasss.soc.surrey.ac.uk/10/3/9.html`

Greenberg, J. H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA: The MIT Press.

Greig, J. M. (2002). The end of geography? globalization, communications, and culture in the international system. *Journal of Conflict Resolution*, *46*(2), 225–243.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*(3), 441–480. doi: 10.1080/15326900701326576

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical population biology*, *69*(3), 339–348.

Hahn, M. W., & Bentley, R. A. (2003). Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(suppl_1), S120–S123.

Hall, B. K. (2003). Evo-devo: evolutionary developmental mechanisms. *International Journal of Developmental Biology*, *47*(7-8), 491–495. Retrieved from `http://www.ijdb.ehu.es/web/paper/14756324/evo-devo-evolutionary-developmental-mechanisms`

Harary, F. (1959). A criterion for unanimity in french's theory of social power.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. doi: 10.1126/science.298.5598.1569

Henrich, J. (2004). Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses—the tasmanian case. *American Antiquity*, *69*(2), 197–214.

Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and human behavior*, *19*(4), 215–

241.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … others (2005). “economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, *28*(6), 795–815.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61–83.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *12*(3), 123–135.

Henrich, J., & McElreath, R. (2007). Dual-inheritance theory: the evolution of human cultural capacities and cultural evolution. In *Oxford handbook of evolutionary psychology.*

Henrich, N., & Henrich, J. P. (2007). *Why humans cooperate: A cultural and evolutionary explanation*. Oxford University Press.

Herzog, H. A., Bentley, R. A., & Hahn, M. W. (2004). Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_5), S353–S356.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Cambridge, MA: Harvard University Press.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88-111.

Hockett, C. F. (1966). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of language* (pp. 1–29). Cambridge, MA: The MIT Press.

Hodgson, G. M. (2015). On defining institutions: rules versus equilibria. *Journal of Institutional Economics*, *11*(3), 497–505.

Hovers, E., & Belfer-Cohen, A. (2006). “now you see it, now you don’t”—modern human behavior in the middle paleolithic. In *Transitions before the transition* (pp. 295–304). Springer.

Hudson Kam, C. L., & Chang, A. (2009). Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 815.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, *1*(2), 151–195.

Huntington, S. P., & Dunn, S. (2004). *Who are we?: The challenges to america’s national identity*. Simon and Schuster.

Hurlbert, S. H. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, *52*(4), 577–586.

Huxley, J. (1942). *Evolution: The modern synthesis*. London: George Allen & Unwin.

Janda, R. D., & Joseph, B. D. (2003). On language, change, and language change; or, of history, linguistics, and historical linguistics. *The handbook of historical linguistics*, 3–180.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? a practical guide to computing and reporting bayes factors. *The Journal of Problem Solving*, *7*(1), 2.

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. doi: 10.1126/science.aax0287

Jeffreys, H. (1961). *The theory of probability (3rd ed.).* OUP Oxford.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, *52*(1/2), 225–247.

Judd, S., Kearns, M., & Vorobeychik, Y. (2010). Behavioral dynamics and influence in networked coloring and consensus. *Proceedings of the National Academy of Sciences*, *107*(34), 14978–14982.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–294.

Kandler, A., & Laland, K. N. (2009). An investigation of the relationship between innovation and cultural diversity. *Theoretical population biology*, *76*(1), 59–67.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, *126*(3), 278.

Kearns, M., Judd, S., Tan, J., & Wortman, J. (2009). Behavioral experiments on biased voting in networks. *Proceedings of the National Academy of Sciences*, *106*(5), 1347–1352.

Keller, E. F. (2010). *The mirage of a space between nature and nurture*. Durham, NC: Duke University Press.

Kemeny, J. G., & Snell, J. L. (1983). *Finite markov chains*. New York: Springer-Verlag.

Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the effect of group size on cultural accumulation. *Evolution and Human Behavior*, *35*(4), 285–290.

Kendal, J., Giraldeau, L.-A., & Laland, K. (2009). The evolution of social learning rules: payoff-biased and frequency-dependent biased transmission. *Journal of theoretical biology*, *260*(2), 210–219.

Kendal, J., Tehrani, J. J., & Odling-Smee, J. (2011). *Human niche construction in interdisciplinary focus.* The Royal Society.

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

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110. doi: 10.1109/4235.918430

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, *24*(1), 118–137. doi: https://doi.org/10.3758/s1342

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. Retrieved from `https://www.pnas.org/content/105/31/10681` doi: 10.1073/pnas.0707835105

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245. doi: https://doi.org/10.1073/pnas.0608222104

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Klemm, K., Eguíluz, V. M., Toral, R., & San Miguel, M. (2005). Globalization, polarization and cultural drift. *Journal of Economic Dynamics and Control*, *29*(1-2), 321–334.

Kline, M. A., & Boyd, R. (2010). Population size predicts technological complexity in oceania. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1693), 2559–2564.

Kobayashi, Y., & Aoki, K. (2012). Innovativeness, population size and cumulative cultural evolution. *Theoretical population biology*, *82*(1), 38–47.

Kobayashi, Y., Wakano, J. Y., & Ohtsuki, H. (2019). Evolution of cumulative culture for niche construction. *Journal of theoretical biology*, *472*, 67–76.

Komarova, N. L., Niyogi, P., & Nowak, M. A. (2001). The evolutionary dynamics of grammar acquisition. *Journal of theoretical biology*, *209*(1), 43–59.

Komarova, N. L., & Nowak, M. A. (2003). Language dynamics in finite populations. *Journal of theoretical biology*, *221*(3), 445–457.

Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. *Trends in Cognitive Sciences*, *10*(5), 198–203. doi: 10.1016/j.tics.2006.03.001

Krauss, M. (1992). The world's languages in crisis. *Language*, *68*(1), 4–10.

Krifka, M. (1999). Compositionality. In R. A. Wilson & F. C. Kiel (Eds.), *The mit encyclopedia of the cognitive sciences* (pp. 152–153). Cambridge, MA: The MIT Press.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, *1*(3), 199–244.

Labov, W. (2011). *Principles of linguistic change, volume 3: Cognitive and cultural factors* (Vol. 36). John Wiley & Sons.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Levy, E. R., Hodgson, S., Fox, M., … Monaco, A. P. (2000). The *SPCH1* region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *The American Journal of Human Genetics*, *67*(2), 357–368. doi: https://doi.org/10.1086/303011

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519. doi: 10.1038/35097076

Laland, K. N. (2017a). *Darwin's unfinished symphony*. Princeton University Press Princeton, NJ.

Laland, K. N. (2017b). *Darwin's unfinished symphony: How culture made the human mind*. Princeton, NJ: Princeton University Press.

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, *23*(1), 131–146. doi: https://doi.org/10.1017/S0140525X00002417

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*(5), 719–745. doi: 10.1007/s10539-012-9335-1

Laland, K. N., Odling-Smee, J., & Turner, S. (2014). The role of internal and external constructive processes in evolution. *Journal of Physiology*, *592*(11), 2413–2422. doi: 10.1113/jphysiol.2014.272070

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., … Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1813), 20151019. doi: https://doi.org/10.1098/rspb.2015.1019

Laskowski, K. L., & Pruitt, J. N. (2014). Evidence of social niche construction: persistent and repeated social interactions generate stronger personalities in a social spider. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1783), 20133166. doi: https://doi.org/10.1098/rspb.2013.3166

Lass, R. (1997). *Historical linguistics and language change* (Vol. 81). Cambridge University Press.

Latané, B., Nowak, A., & Liu, J. H. (1994). Measuring emergent social phenomena: Dynamism, polarization, and clustering as order parameters of social systems. *Behavioral science*, *39*(1), 1–24.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667–694.

Lee, Y., Stabler, T. C. C. E. P., & Taylor, C. E. (2005). The role of population structure in language evolution. *language*, *22*(23), 24–25.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley & Sons.

Letki, N. (2008). Does diversity erode social cohesion? social capital and race in british neighbourhoods. *Political Studies*, *56*(1), 99–126.

Levinson, S. C., & Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity* (Vol. 5). Cambridge University Press.

Lewens, T. (2019). The extended evolutionary synthesis: what is the debate about, and what might success for the extenders look like? *Biological Journal of the Linnean Society*, *127*(4), 707–721. doi: https://doi.org/10.1093/biolinnean/blz064

Lewis, D. (2008). *Convention: A philosophical study*. John Wiley & Sons.

Lewontin, R. (1965). Selection in and of populations. ideas in modern biology. In *Proc. xvi int. cong. zool.* (Vol. 6, pp. 299–311).

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*(3), 400–411.

Lewontin, R. C. (1983). Gene, organism, and environment. In D. S. Bendall (Ed.), *Evolution from molecules to men* (pp. 273–285). Cambridge: Cambridge University Press.

Leydesdorff, L. (2001). Technology and culture: The dissemination and the potential'lock-in'of new technologies.

Lipo, C. P., Madsen, M. E., Dunnell, R. C., & Hunt, T. (1997). Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology*, *16*(4), 301–333.

Loh, J., & Harmon, D. (2005). A global index of biocultural diversity. *Ecological indicators*, *5*(3), 231–241.

Lonely Star Software, I. (2017). *Guess-a-Sketch*. `https://www.guessasketch.com/`. ([Retrieved from https://www.guessasketch.com/])

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, *108*(22), 9020–9025.

Lotem, A., Halpern, J. Y., Edelman, S., & Kolodny, O. (2017). The evolution of cognitive mechanisms in response to cultural innovations. *Proceedings of the National Academy of Sciences*, *114*(30), 7915–7922. doi: https://doi.org/10.1073/pnas.1620742114

Lumsden, C. J., & Wilson, E. O. (1981). *Genes, mind, and culture: The coevolutionary process*. Cambridge, MA: Harvard University Press.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.

Mace, R., & Holden, C. J. (2005). A phylogenetic approach to cultural evolution. *Trends in ecology & evolution*, *20*(3), 116–121.

Maffi, L. (2005). Linguistic, cultural, and biological diversity. *Annu. Rev. Anthropol.*, *34*, 599–617.

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, *20*(11), 1309–1312.

Manzo, J. F. (1996). Taking turns and taking sides: Opening scenes from two jury deliberations. *Social Psychology Quarterly*, *59*(2), 107.

March, J. G. (2005). Parochialism in the evolution of a research community: The case of organization studies. *Management and organization review*, *1*(1), 5–22.

Marcus, G. F., & Fisher, S. E. (2003). *FOXP2* in focus: what can genes tell us about speech and language? *Trends in Cognitive Sciences*, *7*(6), 257–262. doi: https://doi.org/10.1016/S1364-6613(03)00104-9

Marsden, P. V., & Friedkin, N. E. (1993). Network studies of social influence. *Sociological Methods & Research*, *22*(1), 127–151.

Martínez, M., & Godfrey-Smith, P. (2016). Common interest and signaling games: a dynamic analysis. *Philosophy of Science*, *83*(3), 371–392.

Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*(3), 764–769.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Mayr, E. (1961). Cause and effect in biology. *Science*, *134*, 1501–1506. doi: 10.1126/science.134.3489.1501

Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.

Mesoudi, A. (2011a). Cultural evolution. *eLS*, 1–8.

Mesoudi, A. (2011b). Variable cultural acquisition costs constrain cumulative cultural evolution. *PloS one*, *6*(3), e18239.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3489–3501.

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and brain sciences*, *29*(4), 329–347.

Minelli, A. (2003). *The development of animal form: Ontogeny, morphology, and evolution*. Cambridge: Cambridge University Press.

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic diversity and economic development. *Journal of Development economics*, *76*(2), 293–323.

Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). Package ‘bayesfactor’. *URLh http://cran/r-projectorg/web/packages/BayesFactor/BayesFactor pdf i (accessed 1006 15)*.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language*, *28*(3), 360–374. doi: https://doi.org/10.1016/0749-596X(89)90039-9

Morgan, T. J., Acerbi, A., & Van Leeuwen, E. J. (2019). Copy-the-majority of instances or individuals? two approaches to the majority and their consequences for conformist decision-making. *PloS one*, *14*(1).

Morgan, T. J. H., & Laland, K. N. (2012). The biological bases of conformity. *Frontiers in neuroscience*, *6*, 87.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks.

Theoretical Economics, *8*(1), 1–40.

Müller, G. B. (2017). Why an extended evolutionary synthesis is necessary. *Interface Focus*, *7*, 20170015. doi: 10.1098/rsfs.2017.0015

Müller, G. B. (2020). Eco-evo's contributions to the extended evolutionary synthesis. In L. N. de la Rosa & G. B. Müller (Eds.), *Evolutionary developmental biology: A reference guide.* Cham: Springer. doi: 10.1007/978-3-319-33038-9_39-1

Muthukrishna, M., & Schaller, M. (2019). Are collectivistic cultures more prone to rapid transformation? computational models of cross-cultural differences, social network structure, dynamic social influence, and cultural change. *Personality and Social Psychology Review*, 1088868319855783.

Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2014). Sociality influences cultural complexity. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1774), 20132511.

Nagendra, H. (2002). Opposite trends in response for the shannon and simpson indices of landscape diversity. *Applied geography*, *22*(2), 175–186.

Naiman, R. J., Johnston, C. A., & Kelley, J. C. (1988). Alteration of north american streams by beaver. *BioScience*, *38*(11), 753–762. doi: https://doi.org/10.2307/1310784

Neiman, F. D. (1995). Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in illinois woodland ceramic assemblages. *American Antiquity*, *60*(1), 7–36.

Newson, L., Richerson, P. J., & Boyd, R. (2007). Cultural evolution and the shaping of cultural diversity. *Handbook of cultural psychology*, 454–476.

Nia, H. T., Jain, A. D., Liu, Y., Alam, M.-R., Barnas, R., & Makris, N. C. (2015). The evolution of air resonance power efficiency in the violin and its ancestors. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *471*(2175), 20140905.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175–220.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological review*, *97*(3), 362.

Oaksford, M., & Chater, N. (1998). *Rational models of cognition.* Oxford: Oxford University Press.

Odling-Smee, J., & Laland, K. N. (2009). Cultural niche construction: evolution's cradle of language. In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 99–121). Oxford: Oxford University Press.

Olfati-Saber, R., & Murray, R. M. (2004). Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on automatic control*, *49*(9), 1520–1533.

Oyama, S. (2000a). *Evolution's eye: A systems veiw of the biology-culture divide.* Durham, NC: Duke University Press.

Oyama, S. (2000b). *The ontogeny of information: Developmental systems and evolution. second edition.* Durham, NC: Duke University Press.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition.* Princeton University Press.

Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (2000). *Bayesian optimization algorithm, population sizing, and time to convergence* (Tech. Rep.). Lawrence Livermore National Lab., CA (US).

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169–190.

Pigliucci, M., & Müller, G. B. (Eds.). (2010). *Evolution—the extended synthesis*. Cambridge, MA: MIT Press.

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530–535. doi: 10.1126/science.1857983

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–727. doi: https://doi.org/10.1017/S0140525X00081061

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, *95*(2), 201–236. doi: https://doi.org/10.1016/j.cognition.2004.08.004

Plotkin, H. (1997). *Evolution in mind*. London: Allen Lane.

Powell, A., Shennan, S., & Thomas, M. G. (2009). Late pleistocene demography and the appearance of modern human behavior. *Science*, *324*(5932), 1298–1301.

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on facebook. *Available at SSRN 2795110*.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019a). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151–164.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019b). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Reali, F., & Griffiths, T. L. (2009). Words as alleles: connecting language evolution with bayesian learners to models of genetic drift. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1680), 429–436.

Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution*. University of Chicago press.

Robbins, P. (2017). Modularity of mind. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2017 ed.). Retrieved from `https://plato.stanford.edu/archives/win2017/entries/modularity-mind/`

Robert, J. S. (2004). *Embryology, epigenesis and evolution: Taking development seriously*. Cambridge: Cambridge University Press.

Robinson, G. E. (2004). Beyond nature and nurture. *Science*, *304*(5669), 397–399.

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for anova designs. *Journal of Mathematical Psychology*, *56*(5), 356–374.

Rubinstein, A. (2006). Dilemmas of an economic theorist. *Revista de Economía Institucional*, *8*(14), 191–213.

Sabherwal, R., Hirschheim, R., & Goles, T. (2001). The dynamics of alignment: Insights from a punctuated equilibrium model. *Organization science*, *12*(2), 179–197.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7–55). Elsevier.

Scerif, G., & Karmiloff-Smith, A. (2005). The dawn of cognitive genetics? crucial developmental caveats. *Trends in Cognitive Sciences*, *9*(3), 126–135. doi: 10.1016/j.tics.2005.01.008

Scharff, C., & Petri, J. (2011). Evo-devo, deep homology and foxp2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1574), 2124–2140. doi: https://doi.org/10.1098/rstb.2011.0001

Schelling, T. (1978). 1978 micromotives and macrobehavior. new york, norton.

Seber, G. A. (1965). A note on the multiple-recapture census. *Biometrika*, *52*(1/2), 249–259.

Segovia-Martín, J., Walker, B., Fay, N., & Tamariz, M. (2019, Feb). Network connectivity dynamics affect the evolution of culturally transmitted variants. *arXiv e-prints*, arXiv:1902.06598.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, *104*(18), 7361–7366.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.

Shennan, S. J., & Wilkinson, J. R. (2001). Ceramic style change and neutral evolution: a case study from neolithic europe. *American Antiquity*, *66*(4), 577–593.

Shibanai, Y., Yasuno, S., & Ishiguro, I. (2001). Effects of global information feedback on diversity: extensions to axelrod's adaptive culture model. *Journal of Conflict Resolution*, *45*(1), 80–96.

Shilton, D., Breski, M., Dor, D., & Jablonka, E. (2020). Human social evolution: Self-domestication or self-control? *Frontiers in Psychology*, *11*, 134. doi: 10.3389/fpsyg.2020.00134

Shteingart, H., Neiman, T., & Loewenstein, Y. (2013). The role of first impression in operant learning. *Journal of Experimental Psychology: General*, *142*(2), 476.

Shu, W., Yang, H., Zhang, L., Lu, M. M., & Morrisey, E. E. (2001). Characterization of a new subfamily of winged-helix/forkhead (fox) genes that are expressed in the lung and act as transcriptional repressors. *Journal of Biological Chemistry*, *276*(29), 27488–27497. doi: 10.1074/jbc.M100636200

Simon, C., & Hessen, D. O. (2019). Selection as a domain-general evolutionary process. *Behavioural processes*, *161*, 3–16.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688.

Smaldino, P. E., & Schank, J. C. (2012). Movement patterns, social dynamics, and the evolution of cooperation. *Theoretical Population Biology*, *82*(1), 48–58.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*(4), 371–386. doi: 10.1162/106454603322694825

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160051. doi: https://doi.org/10.1098/rstb.2016.0051

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. doi: https://doi.org/10.1016/j.cognition.2010.06.004

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. doi: https://doi.org/10.1111/j.1467-7687.2007.00569.x

Stasser, G., & Taylor, L. A. (1991). Speaking turns in face-to-face discussions. *Journal of personality and social psychology*, *60*(5), 675.

Steels, L. (2006). Experiments on the emergence of human communication. *Trends in Cognitive Sciences*, *10*(8), 347–349.

Steels, L. (2011). Modeling the cultural evolution of language. *Physics of life reviews*, *8*(4), 339–356.

Stone, V. A. (1969). A primacy effect in decision-making by jurors. *Journal of Communication*, *19*(3), 239–247.

Stubbersfield, J. M., Tehrani, J. J., & Flynn, E. G. (2015). Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends. *British Journal of Psychology*, *106*(2), 288–307.

Suchard, M. A., Weiss, R. E., & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time markov chain evolutionary models. *Molecular biology and evolution*, *18*(6), 1001–1013.

Sultan, S. E. (2015). *Organism and environment: Ecological development, niche construction, and adaptation*. Oxford: Oxford University Press.

Sultan, S. E. (2017). Eco-evo-devo. In L. N. de la Rosa & G. B. Müller (Eds.), *Evolutionary developmental biology: A reference guide*. Cham: Springer. doi: 10.1007/978-3-319-33038-9_42-1

Sultan, S. E. (2019). Genotype-environment interaction and the unscripted reaction norm. In T. Uller & K. N. Laland (Eds.), *Evolutionary causation. biological and philosophical reflections* (pp. 109–126). Cambridge, MA: The MIT Press.

Szabó, Z. G. (2017). Compositionality. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (summer 2017 edition)*. Retrieved from `https://plato.stanford.edu/archives/sum2017/entries/compositionality/`

Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788), 20140488.

Tamariz, M., & Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cognitive science*, *39*(1), 171–183.

Tëmkin, I., & Eldredge, N. (2007). Phylogenetics and material cultural evolution. *Current anthropology*, *48*(1), 146–154.

Thomas, J., & Kirby, S. (2010). Self domestication and the evolution of language. *Biology & Philosophy*, *33*(1–2), 9. doi: https://doi.org/10.1007/s10539-018-9612-8

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental science*, *10*(1), 121–125.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–691.

Turner, M. A., & Smaldino, P. E. (2018). Paths to polarization: How extreme views, miscommunication, and random chance drive opinion dynamics. *Complexity*, *2018*.

Twigg, L., Taylor, J., & Mohan, J. (2010). Diversity or disadvantage? putnam, goodhart, ethnic heterogeneity, and collective efficacy. *Environment and planning A*, *42*(6), 1421–1438.

Vaesen, K. (2012). Cumulative cultural evolution and demography. *PloS one*, *7*(7), e40989.

Vale, G. L., Flynn, E. G., Kendal, J., Rawlings, B., Hopper, L. M., Schapiro, S. J., … Kendal, R. L. (2017). Testing differential use of payoff-biased social learning strategies in children and chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1868), 20171751.

Van Parijs, P. (2011). *Linguistic justice for europe and for the world*. Oxford University Press on Demand.

Verkuyten, M. (2006). Multicultural recognition and ethnic minority rights: A social identity perspective. *European review of social psychology*, *17*(1), 148–184.

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., … Fisher, S. E. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of *FOXP2*, a gene mutated in speech and language disorders. *The American Journal of Human Genetics*, *81*(6), 1232–1250. doi: https://doi.org/10.1086/522238

Voigt, S., & Kiwit, D. (1998). The role and evolution of beliefs, habits, moral norms, and institutions. In *Merits and limits of markets* (pp. 83–108). Springer.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic bulletin & review*, *14*(5), 779–804.

Walsh, D. M. (2015). *Organisms, agency, and evolution.* Princeton, NJ: Cambridge University Press.

Walsh, D. M. (2019). The paradox of population thinking: First order causes and higher order effects. In T. Uller & K. N. Laland (Eds.), *Evolutionary causation. biological and philosophical reflections* (pp. 227–246). Cambridge, MA: The MIT Press.

Weber, B. H., & Depew, D. J. (Eds.). (2003). *Evolution and learning: The baldwin effect reconsidered.* Cambridge, MA: The MIT Press.

White, J. W., Rassweiler, A., Samhouri, J. F., Stier, A. C., & White, C. (2014). Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos*, *123*(4), 385–388.

Whitehead, H., Laland, K. N., Rendell, L., Thorogood, R., & Whiten, A. (2019). The reach of gene-culture coevolution in animals. *Nature Communications*, *10*(1), 2405. doi: 10.1038/s41467-019-10293-y

Whittaker, R. H. (1960). Vegetation of the siskiyou mountains, oregon and california. *Ecological monographs*, *30*(3), 279–338.

Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, *197*(3), 795–808. doi: https://doi.org/10.1534/genetics.114.165423

Wimsatt, W. C. (1986). Developmental constraints, generative entrenchment, and the innate-acquired distinction. In W. Bechtel (Ed.), *Integrating scientific disciplines: Case studies from the life sciences* (pp. 185–208). Dordrecht: Martinus Nijhoff.

Wimsatt, W. C. (2014). Entrenchment and scaffolding: An architecture for a theory of cultural change. In L. R. Caporael, J. R. Griesemer, & W. C. Wimsatt (Eds.), *Developing scaffolds in evolution, culture, and cognition* (pp. 77–105). Cambridge, MA: The MIT Press.

Wimsatt, W. C., & Griesemer, J. R. (2007). Reproducing entrenchments to scaffold culture: the central role of development in cultural evolution. In R. Sansom & R. N. Brandon (Eds.), *Integrating evolution and development. from theory to practice* (pp. 227–323). Cambridge, MA: The MIT Press.

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M.-H. (2010). Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Systematic biology*, *60*(2), 150–160.

Young, H. P. (1993). The evolution of conventions. *Econometrica: Journal of the Econometric Society*, 57–84.

Young, H. P. (2001). *Individual strategy and social structure: An evolutionary theory of institutions.* Princeton University Press.

Ziman, J. (2000). *Technological innovation as an evolutionary process.* Cambridge University Press.

Zuidema, W. H. (2002). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Nips'02 proceedings of the 15th international conference on neural information processing systems* (pp. 51–58). Cambridge, MA: The MIT Press.