PhD Dissertation

Epidemiology of Representations: An Empirical Approach

—original title may change—

Sébastien Lerique[1]

Supervisor: Jean-Pierre Nadal[2]
Co-supervisor: Camille Roth[3]

[1]Centre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557, CNRS-EHESS, Paris). Email: sebastien.lerique@normalesup.org.

[2]CAMS, and Laboratoire de Physique Statistique (LPS, UMR 8550, CNRS-ENS-UPMC-Univ. Paris Diderot, Paris). Email: nadal@lps.ens.fr

[3]CAMS, Centre Marc Bloch (CMB, UMIFRE 14, CNRS-MAEE-HU, Berlin), and Sciences Po, médialab (Paris). Email: camille.roth@sciencespo.fr

Contents

Chapter 1

Introduction

CAM: This is an example comment from Camille

TODO: This might be too high-level, so make it more focused. Also add an early hint of what I'll do, which is not that broad. Remember to start easy (but rich) before going to more complex things.

Current scientific knowledge describes the complexity of life, and of human life in particular, through a wide array of theoretical and empirical approaches. Given the heterogeneity of the phenomena we aim to understand, we consider it no surprise that biology, psychology, cognitive science, linguistics, anthropology, sociology, or philosophy claim such varied problems and explanatory programmes. But diversity also begs for questions: how can we bring together programmes which, at times, seem to talk past one another in spite of taking humanity as the same core object? Do different programmes always correspond to different explanatory levels, or should we rather combine them as interlocking aspects of the same unique level? Which of those programmes build on incompatible ontologies or world views, making their perspectives on life irreconcilable? These questions have interested a great number of researchers throughout the 20th century and up to now. Indeed, if we aim to carve life at its joints, the way we combine the diverse theoretical and empirical programmes that describe it will constrain the resulting bones and muscles carved out. In other words, choosing an assembly of fields goes hand in hand with a view of what it means to be human, of what exactly nature and culture correspond to, and of how we can best approach the emergence, complexity, and evolution of human life through time.

LOCATE

Over the last 40 years the field of cognitive anthropology, along with the approach to cultural evolution it suggests, has emerged as the strongest view of how science could combine the findings of anthropology, cognitive science, and biology. This approach, now "a booming cottage industry on the borders of evolutionary biology, archaeology, and biological anthropology" (following Sterelny 2017, and the labels he suggests), has two central traditions: the Californian, initiated by parallel works of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985), and the Parisian, federated around the seminal work of Sperber (1996). The two traditions agree on the combination of fields they defend.

- it has a number of variants, of which SCE and CAT, which seem to agree on the philosophical framing of things, but have open debates and differing focuses
- it also has stark critiques

- at stake in those debates is the overarching vision/paradigm of how facets fit together, which defines the set of obvious and difficult things, guides the design of new questions, experiments, and implementations.
- recently new generations have developed in several fields: cognitive science is at a crossroads, evolutionary biology also. 4E / DST / Eco-evo-devo
- the particular question here is: observe actual phenomena, explain them. From there, what are the ways of accounting for all these phenomena in a unique framework, i.e. with a small number of guiding principles, that will give intelligibility at high level, and guide further questions. What are the differences between the various approaches, and which approaches are viable. Oyama 2001: "We believe that the heuristic value of the idea of developmental information in certain contexts is more than outweighed by its misleading connotations". (This problem, the cog-soc link, is also sometimes viewed as the question of what human nature is, because an answer to that question would define the unit of analysis for culture and everything else.)
- How could there be any problems, and how is this a real question, since it's all a matter of empiricism? How is this a scientific question? Well it's a question at the level of scientific programmes
- First, the breadth of things you can look at is such that a theory will let you define what things are worth looking at. It also makes some cases more natural to explain than others. So a first question is, how much of the phenomena does it capture easily.
- Inside those things, some phenomena are not explained and we have a good idea of what could be the explanation
- Some other phenomena are not explained, and we have no idea how to tackle them, like the hard problem of content
- And the approach you take on those unsolved problems can be somewhat foundational for the rest of your theory, and makes other phenomena more or less natural to account for
- I don't claim to help with anything in the HPC. But I do think that existing theories can be moved forward by improving current empirical methods, to try and see how they pan out in relation to these hard problems.

FOCUS

- I will focus on CAT
- CAT, or ER, was developed by Sperber in the 90's, in a series of innovative articles
- explicitly not a reductionism, and not a Grand Unified Theory
- he proposes a general ontology where we think of things as representations, public/private that circulate, and that global dynamical system has attractors (that are more or less contingent on a situation in space and time)
- that lets us give an ontology for a framework, where we can rephrase long-standing anthropological questions in terms of spread of representations
- the hypothesis of attractors is the main added value of the theory: it claims that because of the interaction of biology, cognition, environment, and current cultural state, the space of how things evolve is somewhat skewed

- as can be understood by the recent name change, it's not clear that representations are the focus (e.g. distance when talking, or even the Claidière & Sperber 2007 example of smoking). I believe the central intuition was that psychology, in the form of at-the-time-emerging cognitive science, has a rich part to play in any synthesis, and is a core ingredient in the appearance of attractors.
- indeed, that is what is emphasised in debates with SCE
- also, it is not an explanatory theory in the sense of quantum physics (as it's not a GUT): Sperber reminds us that finding an attractor is only a cue to looking for its underlying causes. Instead it's a way of thinking at a high level, like a philosophy maybe
- so in a way, it can't be proved or disproved. But it can be more or less fruitful in generating situations and ideas, and providing a global assembly of facts

ARGUE/EXPAND

- up to now, empirical approaches fall into three categories: experimental transmission chains on simple content, compilation of historical works or data, social network data analysis
- each has its problems: transmission chains are on excessively simple content, historical compilations miss the variety of situations (lose detail) and are hard-put to distinguish explanations, social network analysis doesn't look at cognitive factors
- my focus will be on short utterances, bringing case studies of two types, one of each
- in-vivo, where there is huge complexity, but it's ecological
- in-vitro, where one can control the complexity a little, but you have to set a task and anything you do is subject to that choice
- I choose this for two main reasons: first, the availability of data, or relative ease in collecting it (though we'll discuss that). Second, language is at the core of the criticisms on CAT, so it's a good way to go to tease approaches apart.
- discuss the usefulness of those cases for CAT (and vice versa): does it help capture and account for the phenomena? And above all, do we observe attraction/convergence in these case studies? If not why not? Do we transform meaning in a particular fashion given a particular situation, and if such effects are observed can they be somehow generalised?
- show how developing these experiments helps:
- (1) formulate existing questions in more detail, and possibly ask new questions: convergence becomes a concrete question with real measures in these experiments, obstacles to or inappropriateness of convergence can become evident also
- (2) show the pain-points and problems in CAT. Notably:
  - the question of what information is, i.e. the information dualism of representations
  - -> level of description for convergence/transformation (can't decide on a principled level)
  - -> interpretation and meaning
  - taking context & environment into account, which is tackled by RT and Niche Construction, although that may not be enough, as it can maintain the dualism of nature/culture (might follow from the information dualism), even if you believe that it started early in biological evolution

OUTLINE

- I start with a detailed review of the literature on current approaches to cultural evolution, neighbouring areas, and parallels, alternatives, or downright criticisms
- I go on to detail the open problems of the field, seen both from its advocates and from its critiques
- I then explain how I contribute to those problems
- The second part details the first case study, an in-vivo "experiment" led on quotes on the internet
- The third part details the second case study, an in-vitro transmission chain experiment of short utterances of various types
- The final part revisits the contributions such experiments can bring to highlight (1) what has been unexplored, (2) the main challenge further case studies in this area are faced with and how it relates to criticisms, (3) how it would be possible to move forward.

1.1 Relevant works

1.1.1 20th social science

Social science has concerned itself with the stability, temporal evolution, and spatial variations and regularities of cultures since the start of its discipline. Émile Durkheim already, in his seminal study ([1897] 2012), was looking at the regularities of suicide rates over the years, and the correlation of those rates with a partition of society into religion-related groups. The continuity of suicide over time, and the links between suicide rate and social group, he argued, suggest we should study suicide as a *social fact*: a phenomenon which, in spite of manifesting itself individually, has emerging properties in large groups, with a causal life at the level of other emergent social phenomena. In particular, this emergent level has effects on individual psychology (one of Durkheim's aims was to establish the autonomy of sociology as a natural science of society). By what mechanisms does such an effect operate, and what role does it play in shaping the stability and evolution of cultures? Contemporaries of Durkheim, as well as later researchers, made this matter one of their central preoccupations.

But as those works also acknowledged, such a question raises issues in need of prior clarification: what is the exact status of culture in relation to psychology, or even biology, and how separable are they? Correspondingly, the question of how culture and psychology are part of one another and, if at all separable, how are the two related, or how best to describe the intermingling of these possible levels, has generated much debate throughout the 20th century.

Mauss (1936), in his studies of the ways in which people of different societies and throughout history use their bodies differently, represents one approach to that question. He noticed and began documenting the resting postures, the attitudes, the ways of walking, of swimming, or of sleeping, that different communities adopt, pass on to their offspring, and evolve through time. His endeavour focused precisely on describing (parts of) culture as an embodied and physical property of life, incorporated through the everyday practices of a community, into which children grow by imitation, teaching, or other kinds of learning. In this sense, he argued, there is no normal way of walking, there are only *ways* of walking: by living with different bodily practices, different communities develop different bodies, none of which is standard. Indeed most of these *techniques of the body* documented by Mauss play a role in the physiological development of people growing into them, with noticeable effects: they will make one able to crouch for long periods of time or sleep while standing or horse-riding, and will also, Mauss argued for instance, change the silhouette of

an adult body by influencing the way bones develop. While such practices are undeniably linked to the physiology and to the history of communities, and in that sense are both biological and social, Mauss also asked what part psychology plays in their development, and what influence they have in turn on psychology: he considered these practices complete *physio-psycho-sociological* phenomena.

The question of the incorporation of culture, and behind it the inseparability of culture, physiology, and psychology, has remained central in social science. Works closer to us have contributed to the matter: for Bourdieu (1980), societal norms are also incorporated in perception, and shape our everyday unreflective interpretation of events, our way of navigating life (our *sens pratique*). According to him, norms become embedded in the fullest sense, through life, into our low-level perception and exploration of the environment. Together, incorporated norms form what he calls a person's *habitus*, a concept he puts at the centre of his theory of social reproduction, building on the idea that members of a society grow into a *habitus* leading them to perceive events in ways that reinforce existing power structures.

Social scientists often criticise Bourdieu's approach for not providing a satisfactory account of individual agency, as in this view it still seems opposed to structure acting as a constraint on action. Another prominent approach, the *Theory of Structuration* developed by Giddens (1984), offers a more balanced way out of the tension between agency and structure (or norms), and in doing so reflects yet a different view of the relation between psychology and culture. Acknowledging that approaches that conceive of structure as an external constraint on individual agency cannot resolve that opposition, Giddens (closer to the notion used in structuralist works) sees structure itself as a set of properties of social systems that bear an inherent duality, and cannot be meaningfully isolated from agency: on one side, action arises by using existing structured resources, and its reliance on structure, or referring to it, is what makes it action and not noise; on the other side, the production of action is a new reinforcement of the structure on which it relies. This notion of structure is akin to Saussure's notion of linguistic structure, which in turn inspired the structuralist tradition in social science: on one side, Giddens recalls, in producing an utterance we rely on incorporated syntactic rules; on the other side, the production of a meaningful utterance contributes to maintaining language as a structured totality.

ADD: steiner-autonomy-2009, who connects Giddens (and others) with the questions of social cognition

Anthropology and sociology have produced considerable amounts of work concerned with the nature of culture in relation to psychology, the most prominent parts of which are reviewed by Risjord (2012). The sample I exposed here represents the approaches which have most influenced the initial questions of the present thesis. Common behind the works of Mauss, Bourdieu, Giddens, and authors contemporary to each of them, lies a certain interest in eliminating biology-culture, nature-nurture, or substance-form dualisms which we routinely rely on in our conception of life. This concern has remained central in contemporary social anthropology, and is worth keeping in mind when studying the cultural evolution approach that I will focus on in what follows.

While current writings often argue that such a separation is not "sharp" (Acerbi 2016, 2; relying on Morin 2016), the dualism it comes from permeates most of our theories about non-human animals, human beings, and life in general. This thesis is no exception, as I am myself an apprentice of the traditions I present and discuss here. However I see no reason to believe it is the best conceptual dicing one could achieve, and am (for now) agnostic about whether or not it should be maintained; whenever such distinctions appear in what follows, for instance between cultural and biological evolution, I will thus be referring to the conception of the works discussed, and will attempt to clarify when otherwise. I will come back to the consequences of this dualism in Section 1.1.5, and

will propose a more detailed discussion in Chapter 4.

1.1.2 Evolution of culture

Today's discussion of the ties between culture and psychology is more influenced by proponents from cognitive science. Inspired by the generality of the populational approach underlying contemporary evolutionary theory, a new wave of analyses developed throughout the eighties and nineties by proposing a blend of (a) the gene-centred level of evolutionary theory, (b) insights that the application of generalised evolutionary principles may, or may not, bring to the study of cultural change, and (c) a construal of the two processes of change, genetic and cultural, as parallel evolutions that actively interact, notably through psychology, learning, and environmental engineering. The combination of (a) and (b) is not new, as a similar inspiration underlay the development of memetics. Tim Lewens clarifies the position of this new wave regarding previous evolutionary approaches to culture:

> "Cultural evolutionists frequently begin their theorizing from a starting point that does not invite use of the meme concept, and may therefore appear neutral regarding its propriety. Their project is to integrate various forms of learning into evolutionary theory, in a way that leaves open the degree to which learning has anything in common with genetic inheritance." (Lewens 2012, 466)

(c), which differentiates current cultural evolutionists from previous approaches, has been developed into two parallel approaches, which Sterelny (2017) terms the Californian and the Parisian traditions.

ADD: scott-phillips-simple-2017, mesoudi-cultural-2016

1.1.2.1 Californian

Cavalli-Sforza and Feldman (1981) first articulated the combination of these three ideas in detail, by building on the fact that an evolutionary process need not be mediated by genetic transmission to take place. Indeed, following the synthesis provided by Lewontin (1982): for any type of item, the combination of (a) a transmission process leading to nonrandom dependence of subsequent instances on preceding instances (e.g. offspring and parent phenotype), (b) at least some variation of the properties of items (against which to select), and (c) some level of differential survival and spread depending on those properties, will lead to evolution by natural selection. By looking at genetic models of evolution as a special case of those general principles of evolution, Cavalli-Sforza and Feldman (1981) developed the mathematical analysis of purely phenotypic transmission, which can take different paths in a given population: vertical (from parent to offspring), oblique (from a non-parent member of the previous generation, to member of the next generation), horizontal (inside one generation), or any combination thereof. Boyd and Richerson (1985; 2005) further developed this line, leading to the formulation of *Standard Cultural Evolution* (hereafter SCE) which offers a systematic analysis of part of the interactions between cultural and genetic evolution, an approach which is now vibrant with empirical work (see Acerbi and Mesoudi 2015 for a review of recent studies). A notable feature of this programme is that it does not constrain itself into a particular view of what culture is. While the main authors do define culture as "information that people acquire from others by teaching, imitation, and other forms of social learning" (Boyd and Richerson 2005, 3), a definition which at first sight might prove difficult to reconcile with a non-informational view of culture (such as the incorporated views from Section 1.1.1), the mathematical models do not constrain the

concept of culture as much. The approach also has deep links with the analysis of *Niche Construction* (e.g. Odling-Smee, Laland, and Feldman 2003), which offers promising steps towards a reconciliation with non-informational views of culture stressing the importance of the development process in evolution. I return to this subject in more detail in Section 1.1.4.

ADD: mesoudi-cultural-2011?

1.1.2.2 Parisian

In the mid-nineties Dan Sperber formalised a second influential approach to the question of the evolution of culture: in a series of innovative articles gathered in Sperber (1996), the author puts forward a research programme called *Epidemiology of Representations* (better known today as *Cultural Attraction Theory*, hereafter noted CAT), and seeks to provide the cognitive and social sciences with a common framework with which to address interdisciplinary questions. One of the guiding questions of Sperber's work is the following: how can we explain both the diversity of culture across regions, and its relative stability through time, knowing that all human beings are more or less made of the same ingredients? In developing an answer, Sperber commits himself to presenting a coherent ontology where the status of each object he refers to is well defined, while at the same time connecting with the many ontologies he identifies in anthropology.

The framework he suggests then starts from an ontology made of "mental representations", which correspond to those defined and studied by classical cognitive science, and "public representations", which are the expressions of mental representations in diverse cultural artefacts such as pieces of text, utterances, pictures, myths, built structures, etc.. New mental representations are constantly formed in people's minds whenever they perceive or interpret public representations. For instance, say I am thinking of a tune (mental representation), and I whistle it (public representation); someone else hears it, and forms their own mental representation. Most of the time, the new representation in that person's head is different from my original representation. This last point is a defining feature of the theory Sperber proposes, in contrast to not only memetics, but also to SCE as outlined above (Sperber 1996, 25–26, 31).

On this basis, Sperber proposes to model human societies as large dynamical systems of people continuously interpreting public representations into mental representations, and producing new public representations through their situated actions (in which mental representations play a role). To explain culture then, in this framework, is to analyse the processes by which representations circulate through a society, with different levels of change along the way. Those processes are many and heterogeneous, which corresponds to the diversity of cultural domains that exist in societies, but the basic ontology remains grounded in the same notion of mental representations from cognitive science.

By developing such an ontology to connect disciplines, Sperber proposes a credible bridge between the notions of representation in social science and that of mental representation in cognitive science, without reducing one area to the other or making simplistic assumptions about the phenomena encountered. In this sense, his proposal is that of a naturalistic ontology for the study of culture which builds on cognitive science principles. It is amenable to anthropology, and encourages the combination of the two bodies of knowledge in a well defined way. Sperber can then rephrase interdisciplinary questions in terms of spread and transformation of representations. For instance: what types of representations are less transformed than others as people integrate them or perceive them, and produce them anew, making them circulate in a society? Such representations, spreading wider than others, become *cultural representations* and will characterise a given society (for instance

the habit of clothing, eating customs or values, or technological knowledge).

Why are those representations so stable, and how do they evolve? Sperber introduces an additional concept to analyse this evolution: the dynamical system of representations which models a society's culture exhibits attractors, called *cultural attractors*, that depend on the complex interaction of psychological and ecological factors, and on the distribution of representations at a given moment in time (Sperber 1996, 106–18). Cultural attractors are one of the core concepts in CAT providing intelligibility to the evolution of culture and to the reciprocal influence of psychology, culture, and environment. As such, a central goal in the CAT research stream has been to identify existing attractors, and explain their emergence based on the interaction of psychological and ecological factors.

The most important intuition in Sperber's proposal, and what differentiates it from previous works, is the centrality of psychology for the evolution of culture (Sperber 1996, 31) and its role in the emergence of attractors. He substantiates this by relying heavily on contemporary cognitive science, and in particular by adopting and extending the view of the modularity of mind initially defended by Fodor (1983) (his view goes further than Fodor's, as he argues for a *massive modularity of mind* applying not only to perception but also to conceptual processes). Combining this view with epidemiology of representations, Sperber argues, results in a theory avoiding both the blank slate approach to psychology (Sperber 1996, 63–66) and a naive application of neo-darwinist formalism to the specific case of culture (Sperber 1996, 101), while still being able to account for the diversity of cultures (Sperber 1996, 120). This specificity of CAT is often highlighted as one of the main differences with SCE (Sterelny 2017, 47); CAT is nonetheless compatible with the framework of gene-culture co-evolution developed by SCE (Sperber 1996, 114), and most authors consider that both theories are compatible but focus on slightly different questions (Sterelny 2017; Acerbi and Mesoudi 2015).

The space opened by the development of SCE and CAT has generated much debate (see e.g. the peer commentary to Mesoudi, Whiten, and Laland 2006) and some heated criticisms (Ingold 2007; answered in Mesoudi, Whiten, and Laland 2007; Lewens 2012, 461–2 provides further discussion and context). Sections 1.1.4 and 1.1.5 come back to these debates and criticisms, which I group into two broad categories. The first concerns the specifics of the evolutionary approach adopted by SCE and CAT: what evolutionary mechanisms should they take into account (for instance niche construction, or epigenetic or extended inheritance) and, correspondingly, how long the restriction of inheritance to two isolated and parallel channels will remain the best approximation. The second critique concerns what Thompson (2007) has termed the "informational dualism" of the notion of representation in cognitive science: at the concrete level, it questions whether a representationalist approach can naturalise meaning, a core aspect in the analysis of culture and life; at the programmatic level it suggests that, compared to the focus on dynamic couplings that Thompson and his colleagues develop under the "enactive approach" banner, a focus on representations in cognitive science has less heuristic value in guiding the exploration of the interactions between culture and psychology. Regardless of whether proponents can reconcile SCE and CAT with the second critique, I believe both approaches can benefit greatly from these debates, and that empirical study has a crucial role to play in exploring the implications of each side.

The empirical exploration of CAT has proven a difficult task, for two main reasons. First, Sperber clarifies that CAT is not, and should not be, a "grand unitary theory":

> "An epidemiological approach … should not hope for one grand unitary theory. It should, rather, try to provide interesting questions and useful conceptual tools, and to develop the different models needed to explain the existence and fate of the various families of cultural representations." (Sperber 1996, 83)

Indeed, CAT accomplishes this:

> "What the epidemiological analogy suggests is a general approach, types of questions to ask, ways of constructing concepts, and a plurality of not too grand theoretical aims." (Sperber 1996, 61)

As a consequence, testing CAT means evaluating the fruitfulness of its paradigm. No single study, or collection of studies for that matter, will reach a yes or no answer to the validity of CAT. But we can reach a collective consensus about the usefulness of approaching the psychology-culture link with the tools and questions developed by that theory.

As the name CAT indicates, an important part of that toolbox is the notion of cultural attractor, which is amenable to empirical study: the presence or absence of a well-defined attractor in a given situation can be turned into a testable hypothesis. In addition at the meta-analytical level, one can readily evaluate the usefulness of a cultural attraction approach in concrete domains, by analysing whether or not the questions it encourages result in fruitful studies, on average, for that domain. So while testing CAT empirically is not a simple matter of yes or no, the theory provides a clear set of tools and lines of thought that are well worth evaluating in a wider, programmatic appraisal.

The second challenge to empirical exploration resides in the fact that quantitative data on out-of-laboratory cultural artefacts is not easy to collect. Although theoretical models to guide that exploration are gradually appearing (Claidière and Sperber 2007; Claidière, Scott-Phillips, and Sperber 2014), developing methods for the quantitative study of attractors is still an open problem. The literature on CAT shows at least two important methods that have been used up to now. First, the meta-analysis of large numbers of anthropological or historical works, which has uncovered several relevant effects. The way portraits are painted over the centuries, for instance, has been show to increasingly favour direct- versus oblique-gaze portraying (Morin 2013). Miton, Claidière, and Mercier (2015) also used this technique to propose an explanation for the historical stability of the bloodletting practice, in spite of its medical ineffectiveness. Exploiting similar records, Baumard et al. (2015) showed a link between the evolution of religious values and the affluence of societies in which they develop. Morin and Miton (EHBEA 2017? [Citation needed]), using the same approach, model the evolution of heraldic coats of arms to test which hypotheses account best for the distribution of patterns and colours through time.

A second approach is to reproduce the evolution of content with human or animal participants in the laboratory, where subjects repeatedly transmit or interact around pieces of content initially chosen by the experimenters. The resulting data is an approximation of the rapid evolution of artefacts in real life, yet in a controlled situation that can then be statistically analysed with access to all the parameters. This technique has been used in studies ranging from the evolution of visual patterns transmitted in a group of apes (Claidière et al. 2014), to the role of argumentation in the transmission of solutions to simple but counter-intuitive reasoning problems (Claidière, Trouche, and Mercier 2017). # ADD: baumard-mutualistic-2013. A number of other disciplines have developed methods that can be relevant to the study of CAT and SCE, and I review these in Section 1.1.3.

While authors have acknowledged that there is no core incompatibility between CAT and SCE (Acerbi and Mesoudi 2015; Sterelny 2017) and a growing number of social scientists is discussing both approaches positively (see the comments to Mesoudi, Whiten, and Laland 2006; and Slingerland 2008), the work I present in this thesis is mainly focused on CAT itself, and less so SCE. While the prevalence of the cultural attraction approach in the French academic landscape naturally had an impact on this choice, I see three other arguments for focusing on CAT. First, it puts cognitive science squarely in the middle of the problem because of the importance of psychology for the study culture. I admit to sharing that opinion (and given current evidence, it is not more than an opinion). Second, its initial goal is not so much to develop a mathematical theory of culture (which

will surely flourish in due time), but to interest all disciplines studying human life in a common framework by using *principled philosophical arguments*. Third, this approach in turn leads to a clear articulation of the philosophical principles defended, discussed, and tested by the theory at the cultural and cognitive levels.[1] # ADD: smaldino-let-2014. As such, it seems that CAT is in a good position to generate productive debate between alternative approaches (such as the propositions coming from social anthropology, and those discussed in Section 1.1.5) which, at the moment, are still competing on the principled level. In particular, the challenges that arise when attempting to study CAT experimentally, as will be the case here for linguistic utterances, should prove helpful in developing a path to the resolution of the critiques faced by both CAT and SCE.

1.1.3 Neighbouring empirical areas

A number of adjacent disciplines have developed interest in and contributed to CAT. Simulations and experimental studies of the evolution of morality (Baumard, André, and Sperber 2013), religion (Boyer 2001), or reasoning (Claidière, Trouche, and Mercier 2017; Mercier and Sperber 2011) have applied the theory; studies on the evolution of language, and of online content through digital media, are also part of this cluster. The latter two areas are particularly relevant for applying CAT to the evolution of linguistic material, as they have already explored factors in this process for a variety of domains, as well as conditions under which gradual changes can build up cumulatively, and the role that such accumulation plays in long term evolution. Essential to these studies are the serial reproduction paradigm and its derivatives, better known today as transmission chains and micro-societies, which have all been used extensively in recent works.

1.1.3.1 The versatile serial reproduction paradigm

The serial reproduction paradigm was first applied in a series of influential studies on memorisation and recall by Bartlett ([1932] 1995), who was experimenting with ways of reproducing the evolution of content through its iterated production by and transmission to different people, under somewhat more controlled conditions than what can be achieved in field work. The paradigm was only one among several other techniques introduced by Bartlett, but it came to have a lasting impact on later studies linking memory to culture. Similar to a game of Chinese Whispers, people participate in a chain along which content is transmitted; the experimenter gives a first participant initial material, typically a picture or a short piece of text, with instructions to read or memorise it; that participant is later asked to recall or reproduce the material, and the experimenter uses their output as input for the next participant, thus constructing a chain of successive memorisation (or perception) and recollection (or reproduction) of the initial material. Participants may or may not know that they are part of a chain. The setup approximates the transmission and change process that happens in everyday life, and while it is quite idealised and imperfect it allows experimenters to explore effects of different factors on the process, and examine the evolution of the artefacts produced as a trace or signal of the overall phenomenon at work.

Important independent variables that may vary include the chain structure (e.g. the number and sources of preceding recollections a participant is exposed to), the reading or memorising instructions and context, the interval between exposition and recall, the possible task given during that interval (or, conversely, a possible overlap of exposition and recall, turning the task into copying, or

[1] In the words of Tim Ingold, an outspoken critic of both CAT and SCE: "[Sperber's work] has the virtue of rendering unusually explicit the assumptions built into much contemporary theorising about culture and cognition, and of driving them through to their logical conclusions" (Ingold 2001, 113).

online interaction between participants), the recall instructions and context, the ordering and organisation of participants in the chain, or the ordering of the content itself in the chain (e.g. is a given participant exposed to material from a single generation all in one go, or to material successively sampled from random generations). Typical studies will then analyse the trends in the transformation of artefacts that were produced, possibly comparing the behaviour across different chains. When faced with the complexities of quantitatively analysing the content itself, studies often opt to contrast a simpler measure (such as rate of change) across two different populations, two conditions in the independent variables mentioned above, or two minimally different types of content. In CAT parlance, this method corresponds to the simplest instantiation of the causal chains of public representation to mental representation to public representation (and so on) described by Sperber (1996, 99), and has thus garnered much attention in the literature related to cultural attraction.

Early work contemporary to Bartlett's applied the technique in a variety of settings. A first stream of research explored ways in which verbal memory and social or cultural background can interact. Maxwell (1936), for instance, contrasted the evolution of a story containing deliberate inconsistencies in groups of different social status and age such as soldiers, priests, educated men or women, students, or boy scouts. He found indications that different groups shortened the story at different rates, and conserved or transformed different pieces of the initial story. Northway (1936) also contrasted the conservation of story parts across groups of children from schools with different social backgrounds, relating the proportion of good recollections of a given item to the everyday activities of the children, or the types of transformations (addition, recasting, modification) to the different age groups and to the diversity of backgrounds in a given group. A later stream of research focused more on pictorial content, such as Ward (1949) who related a record of European coin types from 4th to 1st century B.C. to the trends obtained by Bartlett in serial reproduction of pictures. He confirmed that what Bartlett called "representative detached details", that is parts of a picture that represent a distinctive pattern or object even when isolated from the whole, are well preserved both in artificial and historical serial reproductions. Ward thus suggested that comparing historical data to results of serial reproduction experiments can be a useful method to investigate the influence of universal-psychological versus local-cultural factors in theories of historical evolution of artefacts. Hall (1950) further explored hypotheses made by Bartlett on the effect of titles for pictorial and verbal serial reproduction: in those new experiments, it appeared that titles had a considerable influence on the reproduction of both pictures and texts, by acting as an interpretive frame that guides (or confuses) the participants in interpreting the material.[2] More recently the reliability of Bartlett's studies has been discussed, shifting the focus towards a better control of experimental conditions: Gauld and Stephenson (1967), while still praising Bartlett for the progress that his work represented compared to that of his predecessors, showed that changing the exposition instructions, for instance by giving strict memorisation instructions, or by adding a simple sentence asking participants to be "accurate", would considerably reduce the transformation rate in the resulting chains, a fact that should question the effects measured by a setup that lacks explicit memorisation instructions. The authors also examined the level of conscientiousness of participants in accomplishing their task, showing that the measure correlated negatively to the participants' transformation rate. Noting that the effect was not explained away by a measure of intelligence of participants, they finally suggested that "errors could, it seemed, be avoided, if the subject was so inclined" (1967, 45). Today, one overcomes such an issue by designing experiments that create an intrinsic motivation for participants to adopt the behaviour under study (see Claidière, Trouche, and Mercier 2017 for an example). In contemporary work, Kashima (2000b) offered a reappraisal of the social aspect of Bartlett's contri-

[2] Hall, noting that "the function of the headline or title to some story or article is that of giving a particular emphasis to certain aspects of the text, and is one of the main methods of distorting and biasing what is remembered" (Hall 1950, 120), later reflects on the fact that such results have crucial political implications given the development of mass media.

butions. Kashima argues that while Bartlett's legacy is mostly seen in psychological studies that adopt methodological individualism, Bartlett fiercely opposed that approach; indeed, his view of the interaction of culture and psychology had much in common with views from social science such as those evoked in Section 1.1.1, and with today's social psychology view of a deeper integration between culture and biology.

One recurrent problem, found across most of the above studies, is the difficulty of quantitatively analysing meaning in the material transmitted along chains. Northway (1936), through her focus on meaningfulness, Hall (1950, 120) and Gauld and Stephenson (1967, 42) all indicate that analysing content itself is a laudable but yet unreachable goal. In facing that difficulty, studies recur to analyses of form, survival, or other measurable aspects of the participants' productions. For linguistic material for instance: the length of a recollection, the number of words, concepts, or propositions accurately recalled, or a contrast of the concepts conserved at the end of the chains. Today the problem remains, and contemporary studies of meaningful material use the same techniques to index or approximate changes in the content of artefacts produced (indeed, the work presented in this thesis will be no exception). While much progress can be made by using these approaches to quantifying aspects of the material, or by instead focusing on artefacts that bear no content, it seems a consensual account of meaning, and a corresponding means of analysis, will be necessary for a full-fledged theory of cultural change to establish itself. I will return to this issue in Section 1.2.

1.1.3.2 Contemporary revival

Owing to the development of SCE and CAT, the last two decades have seen a regain of interest for the serial reproduction paradigm (now known as a transmission chain) and its derivatives (generally known as cultural transmission experiments), resulting in the development of new case studies and methodologies. A number of effects have been catalogued by recent works. Bangerter (2000), for instance, showed that in transmitting a scientifically styled account of human sexual reproduction, participants tended to personify ovum and sperm, and attribute stereotypical gender roles to them. Mesoudi and Whiten (2004) have argued that the loss of detail that is repeatedly observed in transmission chains, and in particular in the transmission of reports of everyday events, is due to a hierarchical encoding of memories that biases participants' recollections in favour of higher-level descriptions. Works studying the evolution of religion have focused on the effect of counter-intuitive information: by studying transmission chains of stories made of elements with varying degrees of counter-intuitiveness, Barrett and Nyhof (2001) and Norenzayan et al. (2006) observed a conservation advantage for minimally counter-intuitive elements which supports the eponymous Minimal Counter-Intuitiveness account of religion (Boyer 2001; and see Purzycki and Willard 2016 for a critical discussion). Other similar transmission advantages have been identified in relation to stereotypes: for instance, elements of a story that are consistent with gender stereotypes are less degraded than elements that are inconsistent with such stereotypes, but only when relevant to the story plot (Kashima 2000a); stories made of social information, that is featuring human interactions and plots, are also better transmitted than stories involving non-interacting people or than stories about physical nonhuman elements alone (Mesoudi, Whiten, and Dunbar 2006). The extensive reviews provided by Mesoudi and Whiten (2008) and Whiten, Caldwell, and Mesoudi (2016) give a broader idea of the effects studied and methods used in the literature.

ADD: bebbington-sky-2017, stubbersfield-serial-2015

A more recent stream of research explicitly focuses on the individual transmission step in a chain. For instance, setups with two parallel chains that cross-fertilise each other have been shown to improve transmission rates (Eriksson and Coultas 2012): if at each generation, the participants of two parallel

chains both receive two inputs, one from each chain at the previous generation, information loss is decreased compared to a single chain where participants read twice the same piece of content (of the previous generation). Acerbi and Tennie (2016) further modelled such error-correcting redundancy, simulating minimal scenarios that could favour its evolution given fixed cost-reward constraints. Eriksson and Coultas (2014) also decomposed real world transmission into three phases: choose-to-receive, encode-and-retrieve, and choose-to-transmit. Focusing on emotional selection studies showing that participants are more willing to pass on stimuli that elicit disgust (Heath, Bell, and Sternberg 2001), the authors showed that any of those phases can be the target of a selection pressure. Such pressures can additionally counteract each other, showing that the transmission process is more complex than was initially assumed. Studies of digital media have adopted the increased level of detail, as digital communication lets users copy-paste and bypass the encode-and-retrieve phase of transmission (Acerbi 2016).

Methodological exploration has also led to more free-form interaction setups: in a study of the transmission of risk perception, Moussaïd, Brighton, and Gaissmaier (2015) chose to make the interaction and underlying transmission of information an open-ended process. After initiating a chain by providing the first participant with a set of documents to read on their own, later participants were left to talk freely in successive dyads, and the whole session was recorded on film for subsequent analysis. An analogous change of setup was operated by C. A. Caldwell and Millen (2008b; 2008a) who investigated the cumulative aspect of the evolution of building techniques in a series of experiments asking participants to construct spaghetti towers or paper planes, later evaluated by their height and flight distance. A crucial point in those setups, Sterelny notes (2017, n. 12), is that participants could observe the preceding generation during their experimenting and building, meaning they had access to rich context for the learning phase of transmission. Taking another step to embrace interaction with a setup where participants had to repeatedly build Lego cars in pairs (though without transmission), McGraw et al. (2014) take the complete joint interaction to be their object of study and propose to use the constructions resulting from interacting dyads as a trace of the processes that took place in the interaction. By looking at the characteristics of the cars built by participants, the authors claim to present "methods for discerning, and quantifying, schema-like intersubjective understandings in material form" (McGraw et al. 2014, 4; see also Mitkidis et al. 2015; and Wallot et al. 2016), effectively integrating critiques that I discuss below in Section 1.1.5.

The works reviewed here show that there is much room for exploration, on one side, of the methodological choices in transmission chain experiments (some of which will turn out to be more important than others in the trends observed), and on the other, of the theoretical background that sustains a given study. Such experiments have therefore much to bring to underlying theoretical debates.

1.1.3.3 Language change: Experimental semiotics and Iterated learning

ADD: tamariz-experimental-2017

A related strand of research has developed similar methods to study the emergence and evolution of language. A central experimental paradigm is that of iterated learning, which resembles a transmission chain for artificial languages where participants learn a communication system at each step: a first group of participants must learn to use a simple artificial language (e.g. an artificial vocabulary for naming a range of objects), after which a second group of participants must learn that language through some interaction with or transmission from the first group. The process is then iterated over successive generations, leading to the evolution of the artificial language initially introduced. A related setup used in experimental semiotics surfaces the evolution of interaction without transmission across generations. It consists in pairing participants and assigning them a task that

they must cooperatively solve over repeated iterations, without changing partners. Most often, participants can use a communication channel (or must create it) to help in coordination, such that the setup exposes the way participants iteratively develop conventions over the channel. Studies using these two paradigms have shown that simple biases in participants' interactions and learning capabilities can lead the final evolved communication system to exhibit non-trivial structure. Several factors have been shown to influence the process, including the structure of the objects referred to by the artificial language, the transmission or interaction task (and its surrounding context) for iterated learning, and the reinforcement rules favouring expressivity of the language that will interact with learnability pressures.

A first major goal in this stream is to provide an non-nativist account of the emergence an structure of communication systems (Kirby, Dowman, and Griffiths 2007). Galantucci (2005), for instance, studied pairs of participants facing a cooperative task that required them to develop ad hoc communication conventions. The communication channel, a form of shared whiteboard, distorted participants' input by a constant drift such that they could not enter letters or pictorial drawings. The study highlighted the variety of strategies used by participants to develop conventional sign systems, and in particular the way those systems were adapted to the interaction history of dyads. The authors showed that sign systems were deeply meshed with the situated and time-dependent information that was available, allowing the participants to solve their task by dynamically coordinating (Galantucci 2005, 748–49). Further stripping down the set of assumptions built into the experimental setup, Scott-Phillips, Kirby, and Ritchie (2009) created a cooperative task without providing any communication channel parallel to the task itself. To solve the challenge, participants had to develop a communication system where their own behaviour in the task could become an embodied communicative signal, aside from accomplishing their action for the task either simultaneously or at other times. Here too, the authors underline the importance of developing a dialogue, based on common ground provided by the history of interactions, to bootstrap the creation of a general communication system that can solve the task in all situations. These studies show that communication systems are inherently embodied and situated, but still involve some degree of intentional design by the participants.

Iterated learning setups, on the other side, have focused on the unintentional emergence of structure. Kirby, Cornish, and Smith (2008), for instance, studied the evolution of an initially random artificial vocabulary set which participants had to learn, use, and extrapolate to name events (events were a combination of a shape, a colour, and a movement). By filtering ambiguities in the output vocabulary produced by one participant and given to the next, the authors were able to create a pressure for expressivity of the complete vocabulary; combined with the learnability pressure inherent to the task, and the fact that participants had to extrapolate to unknown events, the vocabulary gradually evolved to regularise variation with the emergence of compositionality corresponding to the three dimensions of the events to name. Crucially, the participants were not aware of the goals of the experiment, nor that they were part of a chain: the emergent structure in the vocabulary thus appeared without intentional design. Using a comparable setup where participants had to extrapolate a colour-naming vocabulary, a subset of which was then transmitted to the next generation, Xu, Dowman, and Griffiths (2013) observed that (probably culture-specific) biases in colour grouping played an important role in the convergence of vocabulary terms. Combinatorial structure and distinctiveness has also been shown to emerge in a set of acoustic signals devoid of meaning that participants had to learn and reproduce (Verhoef, Kirby, and de Boer 2014). Cornish, Smith, and Kirby (2013) encountered similar results for sequences of categorical items that had no inherent meaning: since participants must reproduce a whole set of sequences, or acoustic signals, at each generation, the set behaves as a interconnected system for which learning pressures gradually increase the combinatorial structure.

Experimental semiotics and iterated learning also cross-fertilise each other: after Garrod et al. (2007) used a Pictionary-like collaborative task, without transmission, to study the emergence of symbolism in a lexicon, a process they called "interactive grounding", Fay et al. (2010) extended the findings to a micro-society. Fay and colleagues showed that globally shared symbols can emerge through the gradual alignment of such interactive groundings, leading to an increasingly refined and streamlined symbol system. The authors thus proposed "symbolisation" as an additional mechanism in the emergence of communication systems, based on intra-generational collaborative coordination through interaction, and parallel to the inter-generational learning biases and bottlenecks studied by iterated learning studies. Taking another leaf from experimental semiotics, Winters, Kirby, and Smith (2015) exploited the observation that a large part of the meaning of an utterance comes from its situational context: they studied the influence of the situations in which participants use vocabulary items on the structure of the language that evolves from interaction and iterated learning (thus extending Silvey, Kirby, and Smith 2015). The authors show that, if the situations in which participants communicate do not contrast items on all the dimensions on which they differ, then the vocabulary set often evolves to not encode those dimensions. In other words, if the usage situations shield the participants from certain contrasts between items, the final vocabulary is often under-specified with respect to the full item space: it does not encode the dimensions that discriminate the unobserved contrasts, instead adapting to be useful only for the contrasts that users observed. This stream of research is active and gradually relaxing the constraining hypotheses made by initial studies. Carr et al. (2017), for instance, recently made the space of items more realistic by exploring the emergence of vocabulary sets where participants communicate about a continuous unbounded set of items.

A second, closely related question concerns how an already structured communication system evolves given a set of external, learning, or interaction pressures. Croft (2013) provides a general framework for this question, inspired by principles from biological evolution and by recent debates on the nature of the evolutionary process, and crucially focused on identifying an adequate unit of analysis for the evolutionary study of language change. Indeed the author combines two key insights. On one side, Hull's General Analysis of Selection (Croft 2013, 16) lets him abstract out the principles of evolutionary processes and distinguish between replicator, interactor, and selection, three core components that provide "a model for disentangling different cultural evolutionary processes and identifying their interconnections" (Croft 2013, 18). On the other side, he draws on the critique that Developmental Systems Theory opposes to a gene-centred view of biological evolution (Oyama, Griffiths, and Gray 2001), and insists that a theory of language change should consider utterances to be full life cycles made of pronunciation, meaning, and interpretation in context. He thus proposes the Theory of Utterance Selection, which takes linguemes (i.e. the linguistic structure of sounds, words, constructions and utterances) to be replicators, but always part of a larger cycle; language speakers are the interactors (2013, 16), and the theory defines language as the "population of utterances in a speech community" (2013, 35).

A number of existing studies fit well in this framework. Tamariz et al. (2014), for instance, used a common setup requiring participants to develop a pictorial vocabulary for a pre-given set of words, and modelled the trends participants exhibit in adopting new signs as they go through the interactions of the experiment. The authors found that participants do not select new signs neutrally; rather, they tend to favour signs they have used in the past, even if their partners use different ones, unless they encounter a sign they find obviously superior in representative power. Similarly, in a picture-description transmission chain using an artificial minimal language, Smith and Wonnacott (2010) showed that the accumulation of individual participant biases will regularise the marking of plurals in the evolved language. Kirby et al. (2015), while studying the emergence of structure as the result of combined pressures of expressivity and compressibility (the latter often attributed to learning, Tamariz and Kirby 2015), hint to the fact that the way structure emerges and evolves is

highly dependent on the combination of such pressures. They note, in particular, that "there is some suggestive evidence that structure in language can be modulated by the composition of populations" (Kirby et al. 2015, 99): different communication patterns at the population level, or a different fabric in the population responsible for the transmission and evolution of a language (e.g. more second-language learners, or more children learners), should lead to differences in the evolution of language structure. Regarding symbolism, Caldwell and Smith (2012) extended the micro-society Pictionary-like task studied by Fay et al. (2010) to one where participants were gradually replaced, inducing increased symbolism and successful transmission of the evolved symbols at the same time. Initial members of the micro-society constructed highly iconic representations of the meanings to convey, but as the experiment introduced newcomers to those signs through observation and consequent use, the drawings gradually lost the iconic link to their referent and became simpler.

Large portions of the iterated learning literature draws on and contributes to a parallel theoretical track which laid down the first analytical predictions for models of Bayesian agents learning and producing languages in chains. Griffiths and Kalish (2007) were the first to show that the analytical structure of iterated learning with uniform Bayesian agents can correspond, depending on the way the agents produce new iterations, to well-known statistical inference methods (Gibbs sampling and a flavour of the EM algorithm). In such a setup, iterated learning predictably converges towards distributions determined by the internal prior distributions of agents (i.e. their inference bias). As a consequence, in those analytically derived situations one can straightforwardly predict the final distributions that should evolve under iterated learning, a fact that Kalish, Griffiths, and Lewandowsky (2007) verified with humans in a function learning task. Griffiths, Christian, and Kalish (2008) further exploit this result by using it in the reverse direction: since the outcome of iterated learning, for specific setups, is predictable on the basis of participants' priors, one can use such experiments to investigate the inductive biases of participants. The authors confirmed this, showing that the method infers well-known participant biases in category learning tasks. Griffiths, Kalish, and Lewandowsky (2008) then explored the relevance of these findings for the study of cultural evolution, showing in particular that individual cognitive biases can have significant effects on long-term cultural evolution. Reali and Griffiths (2009) further related those results to the evolution of vocabulary, showing that they are consistent with experimental cases of word-meaning mapping regularisation. Finally Perfors and Navarro (2014), through analytical derivation and experimental confirmation, reintroduced the impact of the external world in those results; the authors showed that the structure of referents (i.e., the external world) will also play a role in the final evolved language, provided it has an effect on the choice of items people actually talk about (versus, the choice of items talked about only depends on the language itself).

Scott-Phillips and Kirby (2010) and Tamariz and Kirby (2016) provide further reviews of the iterated learning literature, and Galantucci, Garrod, and Roberts (2012) and Roberts and Galantucci (2017) offer reviews of experimental semiotics. An interesting and important development in recent works is the reintroduction of pragmatics into theoretical questions. Scott-Phillips (2017), in particular, reaffirms the central role of pragmatics in the creation and understanding of meaning in context, and argues for a much stronger focus on the evolution of pragmatics itself, that is, as he envisions it, on the evolution of *ostensive communication*. Let me close this review by noting that there is an increasing convergence both in the literature and in empirical questions, of the cultural evolution and language evolution fields; the intersection of questions from the two fields is likely to push theoretical issues forward.

ADD: sperber-beyond-2015

1.1.3.4 Digital media

Acerbi (2016) defines digital media as "media encoded in digital format, typically to be transmitted and consumed on electronic devices, such as computers and smartphones". The ubiquity of this medium, which created the ongoing avalanche of available digital traces, has opened both questions and possibilities for the study of cultural evolution over the past 15 years. Indeed digital media is both a measurement tool and an object of study, as it has become embedded in everyday life in many societies, with its own practices of interaction, mediation, or transmission, possibly impacting cultural evolution. While digital practices are different from those in physical encounters, the digital transition remains an addition to the possible range of interaction media, and the cultural evolution framework can study it as such, with increased access to the artefacts those interactions produce. Acerbi (2016) argues precisely for such an approach to digital media, and reviews relevant works that have explored that space. In what follows I present three areas of focus that have received particular attention in the literature.

A core—and somewhat canonical—challenge for digital media has been to describe the behaviours of diffusion and change of artefacts in social networks, and if possible predict their macroscopic spread and evolution. The question is far from new (see Rogers [1962] 2005) and works have historically tackled this question through analytical models, simulations and empirical studies, but the recent increase in access to digital traces and computing power to make sense of such data has boosted empirical developments. Gathering data from blogspace, for instance, has allowed studying the propagation of information topics, as Gruhl et al. (2004) did by separating topics into "chatter" and externally-triggered ("spike") subjects to model their spread over the social network formed by users. The email network is another source of digital traces, with patterns specific to it; indeed Liben-Nowell and Kleinberg (2008) showed that information diffusion along email chains has an unexpected deep tree-like structure, which they suggest is because of the asynchronous nature of email. Such studies focus on *socio-semantic systems*, that is systems made of, on one side, a collection of users whose interactions or links form a social network, and on the other side, a set of topics or subjects around which the users interact, which also features a network-like structure. The two levels of structure reciprocally influence each other, as Cointet and Roth (2009) show for blogspace (see also Cointet and Roth 2007 who explore the relative roles of social network topology and transmission rules, related to the structure of topics in the spread of information).

The scale of the study of social networks has grown considerably over the past decade, and linguistic memes in particular have received much attention. In a landmark endeavour, Leskovec, Backstrom, and Kleinberg (2009) gathered and published a data set of quotes extracted from a million blogs and news outlets over a nine month period, and developed a method to group minimally different occurrences into quotation families in order to quantify the popularity of news topics over time. The technique allowed the authors to study the evolution of the online news cycle, measuring differences in publication timings across blogs and news outlets. Simmons, Adamic, and Adar (2011) further analysed that data set, showing that transformations of quotes upon copy are frequent (contrary to what one would expect for such memes), work that Omodei, Poibeau, and Cointet (2012) then extended with a more accurate multi-level transformation model. Adamic et al. (2016) developed a similar study for the evolution of explicit memes (that include instructions asking the reader to copy and pass on the contents of the meme) in a Facebook data set of hundreds of millions of occurrences; by using a biological evolutionary model of mutation and replication where genotype corresponds to the meme's content and phenotype to the copying instructions, the authors explore the implications of pushing the biological analogy to its limits in such a paradigmatic case. The range of empirical questions, and the technical challenges involved in tackling them, are such, that the focus has moved towards developing methods for the collection and study of similar data sets. For instance, the

MemeTracker project initiated by Leskovec, Backstrom, and Kleinberg (2009) has now evolved into a fully-fledged network collection and analysis platform (Leskovec and Krevl 2014) with associated data sets (Leskovec and Sosič 2016). Another noteworthy example of this is the development by Moritz et al. (2016) of text re-use detection methods for historical works, a technique that could open the application of the above studies to digitised historical corpora.

A second research stream isolates the different processes involved in the spread and change of artefacts. In particular for transformation, separating effects of content from effects of context is a necessary step to understand the processes responsible for the changes of artefacts. Danescu-Niculescu-Mizil et al. (2012) thus studied the memorability of movie quotes by identifying features that can predict quotes marked as memorable by users of the Internet Movie Database (call these IMDb-memorable): from about 1000 movie scripts, the authors extracted around 2200 pairs of quotes, each consisting of one IMDb-memorable quote paired with the closest quote in the movie script that has the same length, is spoken by the same character, and is not IMDb-memorable. By contrasting these pairs, the authors surface the content-related features of a quote that make it memorable, and factor out the context in which the quotes appear, context which otherwise plays an important role in the memorability rating. After checking that human subjects can identify which quote in the pair is memorable (they do so with an average 78% success rate), the authors show that memorable quotes, on average, use less frequent vocabulary, more frequent grammatical categories (POS tags), and more general constructions (fewer 3rd person pronouns, more indefinite articles, etc.) that make them more adaptable to changing contexts (each of these measures, taken individually, partitions the quote pairs into two subsets containing about 40% vs. 60% of the whole set). Cancelling out context effects to develop content-related features has become a widely used approach, with adaptations ranging from the identification of linguistic markers of politeness in online content (Danescu-Niculescu-Mizil et al. 2013) to the measurement of attractiveness of famous quotes (Acerbi and Tehrani 2017). In a study reminiscent of Hall (1950), Acerbi and Tehrani (2017) compared the relative strength of content and presentational context in a sample of famous quotes that participants had to rate for attractiveness. The authors compared conditions where quotes were presented alone, versus presented with random attribution to more or less famous personalities, or versus presented with a random popularity score. They found that such minimal context has little effect if any at all: attribution, famous or not, bears no effect on the attractiveness of a quote, and popularity has little. Althoff, Danescu-Niculescu-Mizil, and Jurafsky (2014) also opened the study of context versus content to relational variables, by showing how social status and presentational features (such as showing a strong need) can affect the success of requests on Reddit.

A third related stream of research focuses more specifically on influence in social networks, and its links with attention: what network effects trigger the diffusion of a particular meme or piece of information? Among the micro-processes involved in the spread of information in networks, what is the role of influence across connected nodes? Bakshy, Karrer, and Adamic (2009) investigated the question of social influence by examining information cascades in Second Life. Information cascades, where a comparatively small initial event triggers large scale diffusion, are a well-known phenomenon in social networks, and their size distribution is well modelled by peer-pressure threshold models which link the cascade behaviours to the topology of the network in which they occur (Watts 2002; Ruan et al. 2015). Bakshy, Karrer, and Adamic (2009) thus tracked the spread of *assets* in the virtual world provided by Second Life (that is pieces of content introduced and copied by players in the game); they find that a significant part of contagion happens along the friend network, instead of in avatar-to-avatar interactions, indicating that the adoption rate of (in-game) social circles has a strong impact on a person's adoption of an asset in Second Life (see Bakshy et al. 2011 for another example study, on Twitter, separating the strength of content from the strength of social influence).

Attention in social networks is another related factor. Considering the amount and constant flow of information available, filtering and attention management is a necessary component of the diffusion of artefacts; it is usually accounted for through competition among pieces of information. Weng et al. (2012), for instance, model the spread of Twitter hashtags through agents with bounded memory and attention, and show that such simple assumptions account well for the distribution of hashtag diffusion along the social network. The relationship between attention and strength of ties has also been explored by Weng et al. (2015) in data gathered from Twitter, cell phone, and email networks. In these data sets, the authors confirm that while strong ties transport the majority of events, users devote comparable attention levels to both strong and weak ties; they suggest that strong ties play a social communication role, while users use weak ties for seeking novel information, a distinction which could explain the different attentional patterns they measure across the different media.

The empirical study of information diffusion and spread has steadily grown since the advent of digital traces; the number of factors included in analyses is growing, and the influence of core processes such as attention is gradually becoming clearer (an interesting addition would be the role of power relationships, which are also detectable through markers of interactive behaviour, Danescu-Niculescu-Mizil et al. 2011). As mentioned above, Acerbi (2016) provides a useful overview of other works that are relevant for current questions of cultural evolution.

1.1.3.5 Conclusion

As I discuss in Chapter 2, the development of data set collection and analysis methods can bring insight, as well as refined questions, to the study of the reciprocal influences between cognition and culture. # TODO: "I" or "we", for BCP? Other empirical fields in psychology and linguistics are useful to the study of CAT as a framework for cognition-culture interactions: I further introduce works in psycholinguistics relevant to the study of quotes online in Chapter 2, and Chapter 4 will return to how future works could make deeper use of "Smartphone Psychology" to contribute to the more contentious issues. # TODO: actually do that. Let me now move on to the most debated developments of cultural evolution, the criticisms opposed to the approach, the alternatives emerging from these critiques, and the possibilities of reconciliation.

1.1.4 Developments

The exact nature of evolution is subject to debate in biology and philosophy of biology, and some of the recent developments have made their way into the core of cultural evolution theory. In a parallel movement, the nature of cognition is itself debated inside philosophy of cognitive science. The aim of the next two sections is to briefly discuss the relevant parts of those debates for the cultural evolution approach, delineating first the elements that could be—or are already partly—integrated into mainstream CAT, and second the critiques which, at least in current writings, seem to call for a partial rethink of the paradigm. The questions here are more theoretical than above, as they explore both the way different disciplines studying life are best meshed together, and what core components should be at the root of such a convergence. This is not to say the debates in biology and cognitive science concern theory alone, as each debated position is well supported by empirical work; rather, up to now those works have not translated to actionable contradictory predictions in the study of cultural evolution proper. Nonetheless, I will argue in Chapter 4 that these debates provide crucial context to understand a particular practical challenge in the empirical study of CAT, namely the definition of the meaning of representations and its impact on the dimensions of attraction. I begin with developments then, that is the elements that seem possible to integrate into CAT. These also

lay some of the groundwork for the subsequent criticisms, which I believe challenge CAT closer to its foundations.

1.1.4.1 Niche construction theory

A central question in the study of evolution is the definition of what counts as heritable material, for which two broad views are competing. The debate, agreements and disagreements between both views are well documented, and I base the following discussion on the recent reviews provided by Laland et al. (2014) and Scott-Phillips et al. (2014). The standard account of biological evolution, or Standard Evolutionary Theory (SET) as termed by Scott-Phillips et al. (2014), defines evolution as "change in the frequency of DNA sequences (i.e., genes and associated regulatory regions) in a population, from one generation to the next" (Scott-Phillips et al. 2014, 1232; referring to Futuyma 2005). Such change occurs through what is known as an evolutionary process:

> "Evolutionary processes are generally thought of as processes by which these changes occur. Four such processes are widely recognized: natural selection (in the broad sense, to include sexual selection), genetic drift, mutation, and migration (Fisher 1930; Haldane 1932). The latter two generate variation; the first two sort it." (Scott-Phillips et al. 2014, 1232)

In this view, DNA sequences constitute the principle heritable material transmitted from parent to offspring across generations, and their distribution and change should be the main focus of evolutionary theory. Furthermore:

> "There are many factors that can cause these four evolutionary processes to occur, and for the skeptics [of Niche Construction Theory], niche construction is one such factor." (Scott-Phillips et al. 2014, 1233)

Niche construction is the process by which organisms engineer their own and other organisms' environment in ways that are often beneficial to them. A classic example of such niches are the dams built by beavers along the rivers they inhabit; a beaver-built dam creates a local lake, and its presence actively changes the environment in which future generations of beavers—as well as neighbouring organisms—develop. The constructed niche is inherited across generations such that it can have a lasting impact on the selection pressures under which later generations evolve. SET recognises this phenomenon and defenders of the classical account are among those who actively study it (Laland et al. 2014). Niche Construction Theory (NCT, Odling-Smee, Laland, and Feldman 2003), however, contends that increasing amounts of evidence are unsatisfactorily accounted for by SET (though not in contradiction with it), and proposes "a broadened concept of inheritance, including 'ecological inheritance,' the modified environmental states that niche-constructing organisms bequeath to their descendants" (Scott-Phillips et al. 2014, 1233). Those constructed environmental states bias the natural selection of later generations (so-called "selective niche"), and also affect the social and ecological environment in which offspring develop (so-called "ontogenetic niche"), both being processes that can lead to evolutionary feedback loops. The evolution of dairying, as analysed by O'Brien and Laland (2012), is claimed as a paradigmatic case that is well accounted for by NCT.

ADD: gilbert-eco-evo-devo:-2015 if clearly aligned

NCT is part of a broader movement in evolutionary biology that seeks to integrate a strong view of such feedback dynamics into evolutionary theory, by combining evolutionary developmental biology ("Evo-Devo") with the evolution of the environment in which development take place. These

works argue for an Extended Evolutionary Synthesis (EES, also presented as Eco-Evo-Devo by Gilbert, Bosch, and Ledón-Rettig 2015), conceiving of evolution as the co-evolution of organism and environment, a system that inherits genetic material, but also constructed selective and ontogenetic niches. The crux of the disagreement with SET lies in the importance of the dynamics that this feedback generates: SET considers it more parsimonious to define evolution as change in frequency of DNA sequences, and thus frames niche construction and other ecological inheritance processes as a cause for changes in DNA. Conversely, EES considers it more *fruitful* to define evolution as change in the whole organism-environment system, for which niche construction is a core evolutionary process, like genetic mutation or natural selection. According to Scott-Phillips et al. (2014), the current evidence does not tease the two perspectives apart unequivocally: all known phenomena can still be explained by both approaches with varying degrees of shoehorning, and predictions from each theory can be rephrased into the other one (although such inseparability might not last). However, proponents of EES argue that the study of ecological processes in evolution, while present in SET, has become systematic only thanks to the change of focus brought by the development of NCT.

The Extended Evolutionary Synthesis offers a natural framework for the study of all aspects of evolution, be they cultural or biological, and indeed the gene-culture co-evolution framework fits well with this synthesis. The communities developing those approaches overlap partially (Marcus Feldman, notably, is a core contributor to both research streams), and Sterelny (2017) argues that NCT is a core—if sometimes implicit—component of both Californian and Parisian cultural evolution. Indeed, EES is capable of integrating a non-opinionated notion of culture as part of the organism-environment system under study, and the task at hand then joins up with that of dual inheritance theory, presented above: identifying the co-evolution dynamics of genetic and environmental inheritance channels. On this view, then, culture is accounted for by a blend of ecological and cognitive-epistemic niche construction processes.

1.1.4.2 "4E" cognitive science: the extended mind

TODO: use chemero-after-2008 for this summary

In a strikingly parallel movement in cognitive science and philosophy of mind, the nature of cognition and its units of analysis have been debated along two broad dimensions (see Chemero and Silberstein 2008 for a detailed review of questions and possible answers). (1) The boundaries of cognition: are cognitive processes brain-bound, do they extend to the body, or do they include the environment or the surrounding (cognitive) organisms, and if so in what sense (Clark and Chalmers 1998; Menary 2010). (2) The role of time-dependent dynamics, and the corresponding construal of the nature of cognition: are cognitive systems best described as digital computers processing information in the form of representations (i.e. symbol processing systems), where time can often be reduced to an ordering of events, or are they best described as dynamical systems where time is important to define rates of change, flows, or dynamic couplings (van Gelder 1998; Beer and Williams 2015). In treating these questions, the extended, embedded, embodied, and enactive approaches to cognitive science (the so-called "4E") have argued to various degrees that cognition is not only (or not at all) an information-processing operation that can take place in the void, but also (or exclusively) a situated activity supported by (or a dynamic coupling with) its environment. The extended mind theory, among the less radical 4E approaches, is quite compatible with EES and thus with both Californian and Parisian cultural evolution. Indeed Sterelny (2010; 2012), building on NCT, has argued that the extended mind approach is a special case of epistemic niche construction. He suggests that the environments human beings grow in are the result of cumulative cognitive niche construction processes, that engineer the material and social environment of humans to support the

growth of everyday cognitive capacities, thus scaffolding cognition during development and life.

1.1.5 Criticisms

1.1.5.1 Developmental systems theory

A related and somewhat complementary extension to the standard evolutionary account has developed in parallel to NCT, with a more radical notion of extended inheritance: Developmental Systems Theory (DST, Oyama [1985] 2000; Oyama, Griffiths, and Gray 2001). Kim Sterelny characterises it, on one side, with three critical theses:

> "(1) We cannot simply assume that the organism/environment boundary is of theoretical significance for developmental and evolutionary biology … (2) It may be legitimate to foreground genetic structure and genetic change for specific explanatory or predictive purposes. But in general, the genes an organism carries are just one set of developmental resources among many. Genes and gene changes are important both to development and to evolution, but they are not of primary or privileged importance. (3) Developmental systems theorists are skeptical about the project of explaining intergenerational similarity by appealing to the transmission of phenotype-making information across generations." (Sterelny 2001, 335)

Crucially, DST claims that overlooked evidence in development indicates that there is a "causal parity" between genes and non-genetic development factors, such that evolutionary theories should not give greater (or smaller) importance to the former over the latter. This in turn sustains the third thesis: views of the genome as a bearer of biological information should be qualified in light of the complex interactions between developmental processes in which genes participate (for further detail, see Griffiths and Stotz 2013, who extensively review the ways in which genes can be, or historically have been, considered to bear information). Now the other side of Sterelny's characterisation of DST is its positive story:

> "The positive program of developmental systems theory is that the fundamental unit of evolution is the life cycle. In turn, the life cycle is the set of developmental resources that are packaged together and interact in such a way that the cycle is reconstructed. The most obvious life cycle is that of the organism plus its immediate environment, but developmental systems theorists are open to the idea that cycles will exist at both finer and coarser grains." (Sterelny 2001, 335)

A driving goal in DST is to recontextualise, explain, and if possible do without the conceptual divide between matter and form, that is between specification and realisation, which underlies most discussions of nature-nurture (Oyama [1985] 2000). This implies putting an endogenous account of the concept of information at the centre of its theory, by focusing on the way such information is generated through the dynamics of a system as it develops and as its resources interact.

In many aspects, NCT-EES and DST complement each other (see Griffiths and Gray 2005 on the complementarity of DST and Evo-Devo in particular). However for the current purposes I suggest we locate both approaches with respect to the two following questions:

- How far should models and theories of evolution integrate the detailed processes of development and environmental interaction to provide an accurate picture of evolution? This question is not about finding the right level of descriptive complexity, but about finding the relative importance of each level of complexity. In other words, it asks what is the shape of the cost

function representing the trade-off between parsimony and explanatory power (rather than where on that function should a theory operate). The classical account of evolution suggests development can be usefully abstracted away, such that analysing gene flow with the four recognised processes (natural selection, genetic drift, mutation, and migration) can account for all important evolutionary dynamics. NCT claims that organism-environment interactions can generate dynamics that do not fit into the standard account but have long term effects nonetheless, warranting an extension of the evolutionary processes considered. DST further claims that evolution's unit of analysis should be the full life cycle of a developmental system (organism or other), which makes it more difficult to abstract out absolute information items but guarantees an accurate view of the causal parity of developmental resources, such that developmental processes are not obviated.

- What notion of information should evolutionary theory rely on? Classical evolutionary theory and NCT rely on the idea that DNA bears the biological information that is used throughout development, in an interaction with the environment. Conversely, DST refuses to conceptually separate inherited from acquired traits, a separation it sees as unnecessary for a populational approach: instead it adopts a relational notion of information as the co-product of development, "not by special creation from nothingness, but always from the conditional transformation of prior structure—that is, by ontogenetic processes" (Oyama [1985] 2000, 4).

Notably, the stances adopted by DST put it at odds with a dual-channel account of cultural evolution. While the developmental systems approach is compatible with population thinking, and thus with darwinian approaches in general (Griffiths and Gray 2005), it sees no theoretical reason to conceptually isolate different genetic and cultural channels of inheritance. Aside from genetic material and ecological niches, developmental systems also inherit epigenetic material, behavioural patterns and communication systems (Jablonka 2001) and more generally the full matrix of resources involved in the development of the system at its next life cycle. It might be useful for practical or descriptive reasons to focus more on some resources than on others, as is done for instance in the study of phylogeny, but doing so does not change the developmental interactions and causal parity of those resources with the rest of the set. A similar point will be made, below, about representations for the study of cognition.

1.1.5.2 Convergence

The notion of developmental system, and the importance it gives to the ontogenetic niche in which organisms grow, is convergent with 4E cognitive science and Sterelny's scaffolded mind approach. Aside from providing a natural account of culture similar to that of EES, DST also integrates the role of the ontogenetic niche in development all the way up to the cognitive level (Stotz 2010). Wimsatt and Griesemer (2007) have argued that an appropriate focus on scaffolding and development could shed light on the mechanisms of inheritance across generations, a necessary step in an account of cultural evolution. In agreement with DST, however, the authors feel that fixed- or dual-channel accounts of inheritance based on a notion of information transferred to offspring[3] obviate the role of development. Conversely, grounding an account of constancy and change in developmental processes requires considering all the resources that a system inherits, and their interaction in development.

In spite of such differences, the literature indicates that many aspects of EES, DST, and 4E cognitive science seem possible to fruitfully integrate for the study of cultural evolution. Indeed, DST is

[3]The definition provided by Boyd and Richerson (2005), and adopted by Mesoudi, Whiten, and Laland (2006), is along these lines. The authors define culture as "information capable of affecting individuals' behavior that they acquire from other members of their species by teaching, imitation, and other forms of social transmission" (Boyd and Richerson 2005, 6).

compatible with the population approach necessary for a darwinist analysis of culture (Griffiths and Gray 2005; Lewens 2012, 477), which is one of the core elements on which both the Parisian and the Californian cultural evolution streams are built. Lewens (2012, 474) also remarks on the encouraging fact that Russell Gray is both one of the main authors of DST, and is now an outspoken proponent of cultural evolutionary theory. A final case in point is the successful application of these ideas in theoretical and empirical proposals: as noted above, Croft (2013) provides a compelling account of language evolution that integrates the main contentions of DST; McGraw et al. (2014) also merge an interactionist approach to mind with classical cultural evolutionary theory, relying on both research streams to analyse the products of the repeated interaction experiment they study.

The shift in focus, from intrinsic capabilities of information-processing systems to the dynamical properties of the coupling of organisms with their environment, is also aligned with parts of the criticisms addressed by anthropologists to the initial versions of Californian and Parisian cultural evolution (Fuentes 2006; 2009; Ingold 1998; 2001).

1.1.5.3 "4E" revisited: the enactive approach

TODO: use chemero-after-2008 for this summary

A second avatar of the debate on the nature of information in evolution presents itself with the nature of representations in cognitive science. There are two levels to this question. First, if cognitive systems are construed as information- or representation-processing systems, the naturalisation of the content of such representations is a non-trivial matter; indeed it has attracted much attention in philosophy of mind. Second, it is not clear that information processing, and thus an account centrally based on representations, is the best description of the nature of cognition itself (as noted above, this is one of the driving questions in the debate around 4E cognitive science). More radical streams of the 4E movement contend that a notion of representation is unnecessary to account for the vast majority—if not all—of human cognition, and are developing alternative proposals (the capacity of cognitive systems to represent, if maintained in the theory, may then be seen as a contingent property and not a constitutive part). The enactive approach, in particular, proposes such an account of cognition. As described by Varela, Thompson, and Rosch (1991), Di Paolo (2005), and Thompson (2007), the enactive account builds on the notion of autopoiesis and ties cognition to living systems, considered as networks of processes that depend on each other for continued operation and continually produce and reproduce both the boundaries separating them from their environment and the conditions of their operation. Given this definition, a living system dies, that is loses its identity, if its network of self-producing processes ceases from functioning, and as such every interaction with the environment bears intrinsic value in terms of its contribution to the maintenance of the system's identity. This value is the basis for the enactive notion of meaning, which displaces the focus on information, or content, in representationalist accounts. Since living systems use their environment to self-reproduce, they are continually coupled to it in order to maintain the organisation of their network of processes. Cognition, then, is the dynamic regulation that the system operates on its coupling with the environment, and is an intrinsically meaning-making activity.

The enactive approach leads to a notion of meaning as a property emerging from the dynamics of interaction with the environment, and can be relevant at several levels of a system's organisation. This notion, and the view of cognition as a *meaning-making* activity, are also central in enactive accounts of social cognition (De Jaegher and Di Paolo 2007) and of the basic processes underlying language (Cuffari, Di Paolo, and De Jaegher 2015). The rationale for the approach is quite similar to that of developmental systems theory: talk of representations (or of biological information, in the case of

DST) easily leads to what Thompson (2007), building on Oyama ([1985] 2000), calls an "informational dualism". The conceptual separation between matter and content creates a gap, and reifies information in a way that makes it difficult to naturalise.

These works are not explicitly directed towards the study of cultural evolution, and are by no means the only proposals competing within the debate on representations. Yet I will argue in Chapters 3 and 4 that CAT's reliance on mental representations as a unit of analysis renders this matter especially relevant, both theoretically and empirically, to the study of cultural evolution. The enactive treatment, among the most radical of the positive accounts because of its non-representationalist commitment, can serve as a useful point of reference, at the far end of the spectrum, in assessing approaches to information and meaning in the context of cultural evolution.

1.2 Open problems

Moving back into the core of the cultural evolution framework put forth by Californian and Parisian cultural evolution, we can now put the focus on a number of outstanding questions for current and future research. The following discussion far from exhausts the questions to tackle, but gives nonetheless an overview of what I consider to be the most actionable of central items in the field.

1.2.1 Attraction versus source selection

The Californian and Parisian research streams are built around two complementary processes: attraction and source selection. Attraction is the umbrella phenomenon studied mostly under the Parisian approach, and is a central concept to explain cultural constancy and change when there is no clear copying behaviour (i.e. transmission is low-fidelity): if we find constancy and gradual change in a given cultural domain in spite of interpretation, rich effects of psychology, cognitive biases, interaction, and no clear copying behaviour of agents, then cultural attraction might be a good candidate to explain its evolution. In itself, finding a cultural attractor is not an explanation, but an indication that the transmission biases are interacting in such a way that the state of culture is maintained in spite of important changes in micro-level transmission events. Source selection is the umbrella process studied mostly under the Californian approach, and is an explanatory factor of evolution in domains that feature copying behaviours (i.e. transmission is high-fidelity): when agents copy traits, or attempt to copy them, with for instance conformity, prestige, or content biases, then evolution can be usefully explored by looking at the way agents select the sources from which they copy, and the differential spread that entails. Both source selection and attraction act on multiple scales, as cultural traits or elements can be copied and transformed on several dimensions that potentially interact.

As Acerbi and Mesoudi (2015) note, attraction and source selection are not incompatible, as both are part of the overall cultural evolution process. Given their multi-level nature, both processes can also be part of one another; for instance, transformation at the level of a sentence can be analysed as selection at the level of words or concepts. Sterelny (2017) also notes that the Parisian and Californian research streams differ mostly in what they aim to explain. For the Californians, the question is how humans managed to survive and develop successful practices in the face of an opaque environment (for instance poisonous plants that, only if processed properly, can become highly nutritive). They explain this by appealing to cumulative cultural learning, which allows agents to learn from the practices of their preceding generation by copying and slightly modifying them in the process. Some

level of imitation or copy is crucial to this account, precisely because the opacity and the dangerousness of the environment makes the sort of self-confident experimenting one could observe without copy extremely risky. For the Parisians, the question is why cultural traditions with no clear utilitarian value can exist, and evolve, aside from survival-related practices such as cooking techniques. They argue that the evolution of such traditions crucially involves ostensive communication (where the recipient of the communication must recognise the intention behind the communicative act), which is fidelity-neutral but not content-neutral: not all pieces of content are communicated equally, and successful communication only rarely entails copy. The Parisians thus ask why, in the face of a low-fidelity process that introduces such variation at every step, some cultural elements maintain a level of constancy and keep being transmitted through time, thus becoming evolving traditions.

If the two approaches aim to explain different features of culture, in which domain is each approach most appropriate? Are there domains that involve both imitation and non-copying communication in important degrees? How do the two processes interact in such cases? These are the questions left open by recognising a complementarity between the Parisian and Californian approaches.

1.2.2 Interaction of cultural and genetic evolution

Opinions on the importance of gene-culture co-evolution dynamics vary widely. To what extent do genetic and cultural evolution interact, and in which cases does cultural change drive genetic change? Niche construction theorists, and more broadly the Extended Evolutionary Synthesis movement, have convincingly argued for the relevance of such interactions in some specific cases, such as lactose tolerance. Developmental systems theorists further argue that genetic material and culture in its broadest sense are part of a wider matrix of resources that participate in the growth of developmental systems, and are thus necessarily interacting. Proponents of the Californian approach to cultural evolution broadly agree with the niche construction perspective, considering that the cultural processes they study (concerned more with norms than with traditions *per se*) can have a long-term impact on the selective niche in which later generations develop. In particular, one view has it that the centrality of cultural transmission in human beings created a selective pressure for transmission capacities themselves (Sterelny 2017), such that cultural evolution drives genetic changes that enhance transmission. While CAT is broadly compatible with a gene-culture co-evolution account, and also discusses cases of change in context with downstream effects, Morin (2016) is not convinced that an impact of culture on genetic change is necessary to explain the emergence of global human traditions. Indeed, he argues that faithful transmission itself is neither necessary nor sufficient to explain the diffusion of global traditions, and consequently that scenarios that do not involve genetic adaptation to imitation are just as plausible given the current evidence.

This question is intimately related to the relative importance of imitation and non-copying communication. But answers might also lie in the importance and exact nature of the cognitive niche described by Sterelny (2010): how does the construction of such niches contribute to both ecological and psychological (through scaffolding of cognitive development) factors of attraction as the Parisian stream views them? How can the process of niche construction be usefully modelled to test hypotheses on the feedback dynamics between different channels of inheritance? Here too much is open to explore, which leads us to the two next points: formalisation and further development of empirical studies.

1.2.3 Framework versus formal theory contrasted with alternatives

Sperber (1996, 83) argues that CAT should not aim to become a "grand unitary theory", a position which is well informed by the diversity of domains CAT seeks to explain. It also seems in accord with the variety of ways one can approach a single domain, as the Cinderella example reminds us. Instead, CAT proposes a framework, a way of thinking that generates certain questions for the explanation of culture. Nevertheless, the approach relies on two fundamental elements that provide some degree of unification. First, Sperber insists on providing a clear ontology to the study of culture: that of public and mental representations, the latter relying on cognitive science. Second, part of CAT's argument for cultural attraction rests on the idea that much human communication is *ostensive* communication (Morin 2016, chap. 2), that is it works through the inference of communicative intentions, a process which involves a great deal of reconstruction. Relevance Theory (Sperber and Wilson [1986] 1995; Wilson and Sperber 2004), an approach that integrates well with CAT and proposes a framework to understand how agents select salient dimensions and implications of behaviours in concrete situations, is one contender for the detailed explanation of ostensive modes of communication. Given these two fundamentals, what prevents CAT from developing an abstract but systematic mapping of communicative process to stylised phenomenon, in a similar manner to the modelling endeavours of Boyd and Richerson (1985)? Sterelny (2017, 49) notes indeed that no such systematic formal models have developed in the Parisian research stream (although some models have been developed, for instance Claidière and Sperber 2007; Claidière, Scott-Phillips, and Sperber 2014).

Part of the challenge, at least, is the formalisation of both representations and the context in which they are communicated and evolve. While models need not reach this level of detail to be useful (indeed Claidière and Sperber 2007; and Claidière, Scott-Phillips, and Sperber 2014 propose higher-level models based on Evolutionary Causal Matrices), the richness of CAT lies principally with the recognition of an important role of cognition in cultural evolution, that is, of transformative and reconstructive processes. Thus, constructing stylised, simple and tractable models for CAT without emptying the approach of its main contribution is a challenge that has yet to be overcome. Classical cultural evolution, to the contrary, rests on imitative processes, such that models can avoid restrictions on the exact nature of a cultural trait and focus on its frequency or spread in a population, without having to worry too closely about mutation. Not so for CAT: a meaningful model of cultural attraction must account for representations and their transformations, up to a degree, but in a sufficiently general way to (1) elicit effects due to the added ingredient of transformation, and (2) be applicable to different contexts. Claidière and Sperber (2007) and Claidière, Scott-Phillips, and Sperber (2014) have developed the first steps of such an approach, with discrete or continuous one-dimensional representation spaces that already show that attraction can create behaviours not accounted for by imitation-only models. The next step, however, is much higher, and is likely to involve a simple version of Relevance Theory, that is on one side an account of the multi-level and multi-dimensional aspect of representations, and on the other side a mechanism for agents to sieve through that added complexity.

While such an endeavour seems quite ambitious given the current state of modelling, it would provide a well-defined playing field to confront accounts of cultural evolution and especially the theories of cognition on which they are based. Indeed, as we have seen in Section 1.1.5, defining the cognitive factor of cultural evolution in terms of representation processing (a view shared by CAT and Relevance Theory) is not the only option on the market; one could imagine, for instance, a modification of CAT where the cognitive-level processes have been swapped out in favour of an enactive account of interactions. Indeed, that approach is also developing a description of the way organisms make sense of their world, a description which can come to replace the information-processing

account of organisms selecting and inferring relevant information in their communications (in the terms of information processing approaches).

Overall, modelling CAT has not yet been pushed to its limits, and for good reasons: the challenges involved are considerable. But such a line, if successful, is extremely promising for the confrontation of approaches that genuinely compete for accounts of the interaction of cultural evolution and cognition. In the meantime however, the empirical investigation of cultural attractors is a workable task has much to bring to these theoretical questions.

1.2.4 Empirical attractors

Reliably defining and observing phenomena that qualify as cultural attractors in real life can be challenging, first and foremost because of the multi-level nature of culture, representations, and attraction. As Acerbi and Mesoudi (2015, 494) note, the definition of what counts as a cultural trait is not settled. Should one focus on mental processes, on artefacts, or on both? At what descriptive level of artefacts or representations does one consider a transformation to be a meaningful cultural change, worthy of being included as an instance of cultural attraction? Take the much debated example of Cinderella: any telling of that story will differ from other instances on dozens of features, which could all be taken to be significant in a particular context. A change in prosody or in the choice of words might alter the overall feel of the story, but not its narrative structure. A change in narrative style or structure, or an instantiation of the characters in a modern setting, might completely change the face of the story while maintaining the "persecuted heroine" aspect that tale classification systems attribute to Cinderella. Telling the tale in a particular context or to certain people only, say as an intimate bed-time story for one's child versus as a political metaphor in public discourse, can dramatically change the way it is received by its listeners. Looking for the information encoded in a version of the tale, as it were, or the representation it can elicit, only serves to postpone the problem one step further. The nature of these effects is well recognised by most writings on CAT, and one way of tackling them could be using Relevance Theory. Nevertheless, treating this diversity in an empirical study, even if it is only by classifying levels to evaluate their isolated relevance, is very much an open question (even though Acerbi and Mesoudi 2015 note that similar multi-level indetermination of the unit of analysis is also found in the study of genetics).

As we shall see, identifying a cultural attractor in a specific domain also opens the question of the feedback loops it emerges from or participates in. Indeed an attractor is expected to have effects on the ecological context from which it arises, since it changes the distribution of representations people are exposed to. When psychological and ecological factors are involved in the emergence of the attractor, feedback effects are to be expected. How does an attractor participate to or transform its own context? In which cases is that retroaction a factor in the emergence of the attractor, versus a side-effect? For long-lived attractors, is this process a form of niche construction?

Second, as explained in Section 1.1.2.2, the scope of the phenomena to measure makes practical data collection also quite challenging. Existing methods can be divided into the meta-analysis of large bodies of historical and anthropological works, laboratory transmission chains or micro-societies, and the analysis of digital traces, especially from social networks. The expertise necessary to uncover reliable patterns using each of those methods makes higher integration a demanding task. Nonetheless, there is much to gain by articulating these paradigms together, individually pushing them beyond their current limitations, and connecting them to studies in cognitive science. Developing empirical methods is an integral part of any theoretical endeavour, and creating better observation tools is a sure path to push theoretical issues forward and to open previously unavailable questions to exploration.

More generally, fleshing out the predictions that attraction makes in specific empirical cases and developing methods to test those situations will help bring the ideas underpinning CAT into sharper focus. Much more empirical application is needed to evaluate the framework, and only such extensive testing in varied domains will determine how fruitful that approach is for the study of cultural evolution.

The present thesis aims to contribute to this project. In the following chapters, I will present two empirical projects aimed at testing for the presence of attractors in a particular case: the evolution of short linguistic written utterances. The practical goals throughout this work have been (1) to combine disconnected but complementary disciplines to improve on current empirical techniques, and (2) to explore dimensions in which those techniques can be pushed beyond existing limitations. In doing so, we shall also face the tension between two general approaches. On one side, *in vivo* studies use passively collected, ecological data, and must face the full complexity of reality plagued with external factors. On the other side, *in vitro* (laboratory) studies control most of the situation in which their data is generated, but are hostage to those same conditions; in particular, it is often necessary to give a task—or something much like it—to participants of a laboratory study, making any results dependent on the implications of the task. The work I present has two additional final goals: (1) contribute to the empirical evaluation of CAT as a theory of the interactions of culture, evolution, and cognition, in the form of the data points that I was able to collect; (2) highlight outstanding questions that are in need of more attention to make progress in the understanding of those interactions, thanks to a discussion of the costs handled and opportunities navigated by empirical work in the linguistic domain.

Chapter 2

Brains Copy Paste

2.1 Introduction

The reciprocal influence between cognition and culture has a long history in both social science and psychology. While this question has been the subject of intense debate in the social sciences in the 20th century, today's discussion is mostly structured by proponents from cognitive science, who construe culture as an evolutionary process analogous and parallel to biological evolution. That analogy can be traced a long way back, with milestones such as Kroeber's works (1952), Dawkins' *Memetics* ([1976] 2006), and later the development of *Dual Inheritance Theory* by Boyd and Richerson (1985) and Cavalli-Sforza and Feldman (1981) among others. More recently, Dan Sperber has drawn on this principle to explicitly connect anthropology and cognitive science through the theory of *Epidemiology of Representations* (Sperber 1996), and the study of cultural evolution has been growing steadily since.

The collection of works by Aunger (2000; in particular Bloch 2000; and Kuper 2000) has shown how memetics cannot account for the levels of transformation culture undergoes as it is transmitted. Mesoudi and Whiten (2008) discuss the uses of transmission chain experiments to test what dual inheritance theory can explain about cultural evolution. Morin (2013) and Miton, Claidière, and Mercier (2015), by carefully compiling a series of anthropological works, demonstrate how cognitive biases have influenced the evolution of cultural artifacts over several centuries. Kirby, Cornish, and Smith (2008) and Cornish, Smith, and Kirby (2013) have shown how evolutionary pressures lead to the emergence of structured and expressive artificial languages in simulations and laboratory experiments. Such transmission chain experiments have also been explored in non-human primates by Claidière et al. (2014).

The theory of Epidemiology of Representations proposes a unifying framework for all these works by recasting them as questions of spread and transformation of representations: these are alternatively located in the mind ("mental representations" in Sperber's terminology), or in the outer world ("public representations") as expressions of mental representations in diverse cultural artifacts (pieces of text, utterances, pictures, building techniques, etc.). A human society is then modeled as a large dynamical system of people constantly interpreting public representations into mental representations, and producing new public representations based on what they have previously interpreted. Two key points are that (a) transmission is not reliable (representations change significantly each time they are interpreted and produced anew, as opposed to e.g. memetics), and (b) the reciprocal

influences of cognition and culture can be captured by studying the evolution of public representations themselves, which is what the above-cited studies are doing.

The theory makes an additional strong hypothesis, which this paper focuses on: as transformations accumulate, some representations evolve to be very stable and spread throughout an entire society without changing any more (they are called "cultural representations", because they characterize a given culture). This process should manifest itself as attractors (called "cultural attractors") in the dynamical system that models cultural evolution, that is: there should be areas of the representation space where cognitive effects in transformations bring representations closer to a given stable asymptotic point.[1]

This hypothesis, a cornerstone of the theory because of the intelligibility it gives to cultural evolution, has been hard to test in concrete situations as quantitative data on out-of-laboratory cultural artifacts is not easy to collect. One approach, as mentioned above, has been the meta-analysis of large bodies of anthropological studies (see Miton, Claidière, and Mercier 2015, for instance). This paper exemplifies a second approach, taking advantage of the ever-increasing avalanche of available digital footprints since the 2000's. Indeed, tools and computing power to analyze such data are now widespread, and the body of research aimed at describing online communities and content is growing accordingly. For instance, the propagation of cultural artifacts across social networks has been studied in blogspace (Gruhl et al. 2004) and in emails (Liben-Nowell and Kleinberg 2008); Cointet and Roth (2009) described the reciprocal influence between the social network topology and the distribution of issues; Leskovec, Backstrom, and Kleinberg (2009) detailed the characteristic times and diffusion cycles both within these social networks and with respect to the topical dynamics of news media, and Danescu-Niculescu-Mizil et al. (2012) studied the characteristics of particularly memorable quotes that circulate in those networks. We believe these works can connect the field of cultural evolution with psycholinguistics to advance the testing of cultural attractors.

To show this we analyze how quotes in blogs and media outlets are modified when they are copied from website to website. These public representations should normally not change as they spread on the Web (as opposed to more elaborate expressions or opinions, not identified as quoted utterances), but empirical observation shows that they are in fact occasionally transformed (Simmons, Adamic, and Adar 2011): authors spontaneously transform quotes, not only cropping them but also replacing words. For instance the quote "we will not be scared of these cowards" (a substring of a quote from former Pakistani President Asif Ali Zardari) is also found as "we will not be **afraid** of these cowards". More meaningful changes often happen too, such as the transformation of McCain's "I admire Senator Obama and his accomplishments" during the 2008 US presidential campaign, into "I **respect** Senator Obama and his accomplishments". Since authors are implicitly required to copy quotes exactly, we can assume that most transformations, especially simple ones such as those shown above, are the result of automatic (i.e. hard to control) low-level cognitive biases of the authors.

We thus ask the following question: given such representations that seem to evolve precisely because of the kind of automatic cognitive biases evoked in the theory of epidemiology of representations, do cultural attractors appear, and if so how do cognitive biases participate in them? We chose to restrict our analysis to substitutions (i.e., one word being replaced by another), both to keep the analysis tractable and because of missing information in our data set.[2] While this limits the scope of our results to the particular data set we use, the methodological point we also make is left intact.

[1] Attractors need not be points in fact, they can also be sub-areas; in that case any transformation brings representations in the area closer to (or confined to) the target sub-area.

[2] As explained further down, source-destination links between quotes must be inferred from the data set, an operation which is much more reliable if we restrict our analysis to substitutions. This also impedes us from observing the effect of accumulated transformations in the long term, limiting our results to a view of the individual evolutionary step.

By characterizing words using 6 well-studied features, we identify what makes a substitution more likely, and how a word changes when it is substituted. This exploratory approach uncovers a number of transmission biases consistent with known effects in linguistics. While the transformations we describe are not the only ones at work in this data set, our analysis also indicates that feature-specific attractors could exist because of the substitution process. This study can be viewed as analyzing part of the transmission step operating in transmission chains of artificial languages like those studied by Kirby, Cornish, and Smith (2008), yet with natural language out of the laboratory.

The next section describes our hypotheses along with a review of the psycholinguistics literature. Then, we describe the data set and detail the various assumptions that were made in order to analyze it. Next, we introduce the measures we built to observe cognitive biases operating in quote transmission. Finally, we discuss the relevance of these results for the study of cultural evolution, followed with general guidelines for further work.

2.2 Related work

The study of cultural evolution on the part of cognitive science emerged only recently. While formal models of cultural transmission appeared with the development of dual inheritance theory (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985) and have included the notion of cultural attractor since then (Claidière and Sperber 2007; Claidière, Scott-Phillips, and Sperber 2014), collecting data to test and iterate over such models has been more challenging. The first above-mentioned method consists in rebuilding the history of a given type of representation by compiling anthropological or historical works on the subject (as for instance Morin 2013; and Miton, Claidière, and Mercier 2015, have done). A second approach uses cultural evolution experiments in the laboratory, with an array of methods reviewed by Mesoudi and Whiten (2008). Transmission chains, in particular, have been used extensively to study the evolution of human language (see Tamariz and Kirby 2016 for a review). Other recent examples include studies of the evolution of simple audio loops through consumer preference (MacCallum et al. 2012), the emergence of structure in visual patterns transmitted by baboons (Claidière et al. 2014), and the amplification of risk perception through chains of casual conversation (Moussaïd, Brighton, and Gaissmaier 2015).

Research on online content points to a third approach to this question. By investigating the transformations of quotations in a large corpus of US blog posts and online news stories initially collected and studied by Leskovec, Backstrom, and Kleinberg (2009), Simmons, Adamic, and Adar (2011) and later Omodei, Poibeau, and Cointet (2012) show that even for quotations, a type of public representation that should change the least when transmitted on the Web, it is still possible to witness significant transformations. These studies focus on the influence of the quotation source (e.g. news outlet vs. blog) or of the surrounding public space (e.g. quotation frequency in the corpus), and suggest diffusion-transformation models to capture the dynamics of the population of quotations. But the cognitive features which may determine or, at least, influence these transformations, are overlooked. On the other hand cognitive and linguistic features have been used in diffusion studies not involving transformation: Danescu-Niculescu-Mizil et al. (2012), for instance, show that particularly memorable quotations (taken from movie scripts in this case) use more distinctive words and have more common syntax than less memorable quotations; they are also the quotes that adapt best to new contexts of use. One source of ideas to study the transformations of such quotes, then, might be the psycholinguistic literature studying word and sentence recall.

Potter and Lombardi (1990) suggest that immediate recall of sentences is based on the retention of an unordered list of words which is then regenerated as a sentence at the moment of production.

Priming recall with other words can lead to replacement in the recalled sentence if the primed words are consistent with the overall meaning of the sentence. Regenerated syntax can also be influenced by priming recall with another syntactic structure (Potter and Lombardi 1998), or with verbs whose category constraints call for a different structure (Lombardi and Potter 1992).

Compared to full sentences, recall of word lists provides a situation that is easier to fully explore and has been extensively studied. In particular, the Deese, Roediger, and McDermott paradigm (introduced by Deese 1959; and later popularized by Roediger and McDermott 1995) has shown that it is possible to construct lists of words which reliably create the false memory of an external word related to those in the list. This is done by using lists of words produced by free association from the target intrusion word; the intruding recall then happens with probability nearly proportional to the average semantic association strength between the intruding word and the words in the list. A sizable literature studies this type of task with varying complexities in the design of the lists, a good review of which is given by Zaromb et al. (2006). One notable effect is that the semantic relations between words greatly influence, and correlate to, the order in which words are recalled (Tulving 1962; Howard and Kahana 2002), and that this reordering of items improves subjects' repeated recalls (Tulving 1966). The frequency and type of intrusions in lists of random words are also influenced by associations created by the presentation of previous lists (Zaromb et al. 2006). Indeed, the question of how such temporal associations (contributing to contextual information retrieval in recall) interact with the prior semantic associations of subjects (contributing to associative information retrieval) is at the core of many of these studies.

These effects do not transpose simply to sentence recall however, as not only syntax but also effects of attention come into play for both retrieval and encoding. Jefferies, Lambon Ralph, and Baddeley (2004), for instance, show that attention is central to the encoding and retention of unrelated propositions, on top of more automatic syntactic and semantic processes. This involvement of executive resources also seems to contribute to the much greater memory span subjects exhibit for sentences compared to word lists (see Jefferies, Lambon Ralph, and Baddeley 2004 again, for more details).

Given this complexity we decided to focus on more aggregate measures, where variations of the conditions in which sentences are read and produced have a chance of being statistically smoothed out.[3] If a cognitive bias in the substitution of words manifests itself with simple measures, then it will be worth applying predictive models of the substitution process in further research.

Lexical features, then, are obvious well-studied word measures that can be analyzed in aggregate. Indeed word frequency (see Yonelinas 2002 for a review), age-of-acquisition (Zevin and Seidenberg 2002), number of phonemes (see for instance Rey et al. 1998; Nickels and Howard 2004), and phonological neighborhood density (Garlock, Walley, and Metsala 2001) to name a few, all have known effects on word recognition or production. More complex features based on word networks built from free association or phonological data have also been analyzed: Nelson et al. (2013) for instance, show the importance of clustering coefficient in such a semantic network by studying the role it plays in a variety of recall and recognition tasks (extralist and intralist cuing, single item recognition, and primed free association). Chan and Vitevitch (2010) show that pictures are named faster and with fewer mistakes when they have a lower clustering coefficient in an underlying phonological network. Griffiths, Steyvers, and Firl (2007) analyze a task where subjects are asked to name the first word which comes to their mind when they are presented with a random letter from the al-

[3] Aside from our lack of control on the precise conditions of encoding and recall in our data set, the analysis techniques mentioned above are better suited to data consisting of a high number of measures over a smaller number of lists (in which case it makes sense to ask e.g. what proportion of intrusions come from prior lists). As is explained further down however, our data set is shaped the opposite way: a great number of sentences, with only very few to no measures at all on each sentence.

phabet. The authors show that there is a link between the ease of recall of words and their authority position (pagerank) in a language-wide semantic network built from external word association data (Austerweil, Abbott, and Griffiths 2012 further develop this tool to give a parsimonious account of the fact that related words are often retrieved together from memory).

On the whole, research on lexical features hints towards two antagonistic types of effects (also known as the 'word-frequency paradox', Mandler, Goodman, and Wilkes-Gibbs 1982). On one hand, part of the literature shows that recall is easier for the least "awkward" words; those whose age of acquisition is earlier, length is smaller, semantic network position is more central — this is particularly true in retrieval, that is in tasks where participants are asked to form spontaneous associations or utter a word in response to a given signal. On the other hand, when the task consists in recognizing a specific item in a list, "awkward" words are actually more easily remembered, possibly as they are more informative and plausibly more discernible (see again Yonelinas 2002 for a review). The jury is still out as to whether reformulation alteration, that is spontaneous replacement of words when asked to rewrite a given utterance, is rather of the former or latter sort. We also aim to shed some light on this debate, considering oddness as a dimension of the purported fitness of utterances.

2.3 Methods

We rely on a text corpus made of quotations extracted from online blog posts, and focus on their evolution. Quotations appeared to be a perfect candidate to propose a first measure of automatic cognitive bias in cultural transmission. First, they are usually cleanly delimited by quotation marks, which greatly facilitates their detection in text corpora. Second, they stem from a unique original version, and are ideally traceable back to that version. Third, and most importantly, their duplication should *a priori* be highly faithful, apart from cases of cropping: not only should transformations be of moderate magnitude, but when specific words are not perfectly duplicated, it is safe to assume that the variation is due to involuntary cognitive bias — as writers may expect any casual reader to easily verify, and thus criticize, the fidelity to the original quotation.

We could therefore study the individual transformation process at work when authors alter quotations, by examining the modified words in each transformation. Since our approach is exploratory however, we do not know at the outset which precise effect of cognitive bias we are looking for. Indeed, the data we use does not come from a controlled experiment in the laboratory, designed to elicit a particular effect: they are recordings of real life interactions, with all the complexity and uncertainty of conditions this entails. Our goal, therefore, is to show that cognitive biases have measurable effects in this setting even if they are part of a larger complexity (the detailed prediction and deconstruction of the cognitive processes responsible for them being left to further research). If this is confirmed, we will have successfully tested fundamental cognitive biases with out-of-laboratory data, opening a path to explanations of actual (vs. simulated) cultural evolution with tools from cognitive science. Aiming to exhibit such subtle biases in complex data is the main reason we chose to use aggregate measures that have a chance of smoothing out the possible variations of experimental conditions in the data set.

To keep the analysis tractable, we focused on quotation transformations consisting of the *substitution* of a word by another word (and only those cases) in order to unambiguously discuss single word replacements. This restriction also allows us to more reliably infer the information that is missing in our data set, as explained in the "Substitution model" section. To quantify these substitutions we decided to associate a number of features to each word, the variation of which we can statistically study.

The next subsections describe the data set and the measures we used to assess this cognitive bias.

2.3.1 Corpus-based utterances

We used a quotation data set collected by Leskovec, Backstrom, and Kleinberg (2009), large enough to lend itself to statistical analysis. This data set consists of the daily crawling of news stories and blog posts from around a million online sources, with an approximate publication rate of 900k texts per day, over a nine-month period of time from August 2008 to April 2009 (Leskovec, Backstrom, and Kleinberg 2009}).[4] The authors automatically extracted quotations from this corpus. Each quotation is a more or less faithful excerpt of an utterance (oral or written) by the quoted person; for instance:

> The Bank of England said, "these operations are designed to address funding pressures over quarter-end."

Then, the authors gathered quotations in a graph and connected each pair that differed by no more than one word or that shared at least ten consecutive words (they tested this procedure with a number of different parameters, see Leskovec, Backstrom, and Kleinberg 2009, for more details). We find for example the following variation of the above quote:

> "these operations are **intended** to address funding pressures over quarter-end."

Next, they applied a community detection algorithm to that quotation graph to detect aggregates of tightly connected, that is sufficiently similar, groups of quotations (see again Leskovec, Backstrom, and Kleinberg 2009, for more details). This analysis yielded the final data we had access to, with a total of about 70,000 sets of quotations; each of these sets ideally contains all variations of a same parent utterance, along with their respective publication URLs and timestamps (since the procedure cannot be perfect, sets of quotations contain occasional rogue unrelated variations that should have been discarded or assigned to another set).

Manual inspection of this data set revealed that it contains a significant number of everyday language quotations (such as "it was much better than I expected", "did that just happen", as well as many simple expletive-based sentences). Their presence is largely due to random variations around casual expressions, while we are interested in transformations of news-related quotes causally linked to an original, identifiable utterance. To filter them out, we exclude quotes with less than 5 words or whose occurrences span more than 80 days (indicating causally unrelated occurrences), as well as quotes not written in English. Clusters that are emptied by this procedure are therefore excluded. If, after this screening, a cluster's occurrences still span more than 80 days (because of short-lived but unrelated quotes far apart in time), we also exclude it. We eventually keep 50,427 clusters (out of 71,568; i.e. 70.5%), containing a total of 141,324 unique quotes (out of 310,457; i.e. 45.5%) making up about 2.60m occurrences (out of 7.67m; i.e. 33.9%).[5] Even if we lose some real event-related utterances which are present in clusters lasting more than 80 days (one such lost quote, for instance, is "the city is tired of me and the organization and I have run our course together"), we check that our filtering approach fulfills its goals by coding a random sub-sample of 100 clusters: 35 of them are rejected by the filter, with 15 false negatives (rejected clusters that should have been kept) and 9 false positives (clusters kept when they should have been rejected), giving a precision score of 0.862 and a recall score of 0.789. Furthermore, all but one of the 9 false positives are left with a single

[4] The original article (Leskovec, Backstrom, and Kleinberg 2009) does not provide further details on the source selection methodology.

[5] The significantly larger loss in occurrences indicates that, on average, the clusters we lose contain more occurrences than those we keep, which is to be expected for everyday language utterances.

non-rejected quote, meaning those clusters are ignored by our substitution analysis; this brings the effective precision of our filter to 0.982.[6]

2.3.2 Word-level measures

2.3.2.1 Lexical features

We first introduce some lexical measures on words.

- **Word frequency**: the frequency at which words appear in our data set, known to be relevant for both recognition and recall (Gregg 1976),
- **Age of Acquisition**: the average age at which words are learned (obtained from Kuperman, Stadthagen-Gonzalez, and Brysbaert 2012), known to have different effects than word frequency (Morrison and Ellis 1995; Dewhurst, Hitch, and Barry 1998),
- **Phonological** and **Orthographic Neighborhood Density** (obtained from Marian et al. 2012), also known to be relevant for word production (Garlock, Walley, and Metsala 2001),
- The average **Number of Phonemes** and **Number of Syllables** for all pronunciations of a word (obtained from the Carnegie Mellon University Pronouncing Dictionary, Weide 1998)[7], as well as **Number of Letters**, as a proxy to word production cost,
- The average **Number of Synonyms** for all meanings of a word (obtained from WordNet 2010) as an *a priori* indicator of how easy it would be to replace a word.

We also consider grammatical types within quotations by detecting *Part-of-Speech* (POS) categories with TreeTagger (Schmid 1994); we distinguish between verbs, nouns, adjectives, adverbs, and closed class-like words.

Aside from these raw features, the systemic dimension of vocabulary (Cornish, Smith, and Kirby 2013) has led authors to develop measures based on the full topology of networks built from free association data or phonological similarity. Several such measures have been shown to be involved in recall, recognition, and naming tasks (Nelson et al. 2013; Chan and Vitevitch 2010; Griffiths, Steyvers, and Firl 2007).

To compute these features we relied on the free association (FA) norms collected by Nelson, McEvoy, and Schreiber (2004), which record the words that come to mind when someone is presented with a given cue. As Nelson, McEvoy, and Schreiber (2004) explain, "free association response probabilities index the likelihood that one word can cue another word to come to mind with minimal contextual constraints in effect." Similar to what Griffiths, Steyvers, and Firl (2007) did, we first considered the directed weighted network formed by association norms, where nodes are words and edges are directed from cue to target word, with a weight equal to the association strength (that is the probability of that target word being produced when this particular cue is presented). This network is of particular interest since it lets us define features that reflect the associations driving false memories in word lists (Deese 1959), a phenomenon which may be involved in the transformation of quotations.

We used three standard measures on the FA network:

[6] A similar analysis was made for language detection, which is part of the cluster filtering: out of 100 randomly sampled quotes, 17 are rejected because their detected language is not English, with no false positives and 6 false negatives, giving a precision score of 1 and a recall score of 0.933. Of the 6 false negatives, 4 had less than 5 tokens and would have been excluded by the cluster filter anyway.

[7] The CMU Pronouncing Dictionary is included in the NTLK package (Bird, Klein, and Loper 2009), the natural language processing toolkit we used for the analysis.

- **Incoming degree centrality**, measured by the number of cues for which a given word is triggered as a target, and a corresponding generalized measure, node **Pagerank** (Page et al. 1999), which has already been used on the FA network by Griffiths, Steyvers, and Firl (2007). In the present case these two polysemy-related measures are quasi-perfectly correlated.[8]
- **Betweenness centrality**, another measure of node centrality describing the extent to which a node connects otherwise remote areas of the network (Freeman 1977). This quantity tells us if some words behave as unavoidable waypoints on association chains connecting one word to another.[9]
- **Clustering coefficient**, which measures the extent to which a node belongs to a local aggregate of tightly connected nodes (Watts and Strogatz 1998), computed on the undirected weighted version of the FA network.[10] This tells us if a word belongs more or less to a group of equivalent words (from a free association point of view).

2.3.2.2 Variable correlations

Several of these features are strongly related and can be grouped together. To make correlation values as well as future comparisons more reliable, we log-transformed features that have marked exponential distributions (i.e. a few words valued orders of magnitude higher than the vast majority of other words).

The pairwise correlations in the initial set of features appears in Fig. 2.1. By looking at absolute values, three subsets of highly correlated features can be easily identified: (a) number of letters, phonemes, and syllables with pairwise correlations greater than .75; (b) orthographic and phonological neighborhood densities, with a correlation of .8; (c) age of acquisition, betweenness, degree, and pagerank centralities, with absolute pairwise correlations at .41, .59, .6, .61, .63 and .85. Applying a feature agglomeration algorithm targeted at 6 groups refined this observation by producing identical (a) and (b) groups, a (c) group without betweenness centrality which was instead assigned to a group (d) with clustering coefficient, and the remaining features (frequency and number of synonyms) as singletons.[11]

Since our data is about written transformations, number of letters and orthographic neighborhood density are the natural representatives of groups (a) and (b) respectively. Given the importance of age of acquisition in the lexical feature literature, we chose it to represent group (c). Finally we used clustering coefficient to represent group (d) since it has already been used in previous studies. The final set of features we will discuss in the rest of the paper, as well as their cross-correlations, can be seen in Fig. 2.2.[12]

[8] Note that in-degree does not take the weights of links into account, as it counts 1 for each incoming link. Pagerank on the other hand, does take the weights into account.

[9] For this measure, weights are interpreted as inverse cost: the stronger a link, the easier it is to travel across it. A stronger link will be favored over weaker links in the computation of the shortest path between two words.

[10] he Clustering coefficient is formally defined as the ratio between the number of actual versus possible edges between a node's neighbors; this is poorly defined in the case of directed networks, which led us to ignore the direction of links in the network for this measure (if two words are connected in both directions, the weights of both links are added to make the final undirected link's weight).

[11] Agglomerating into less than 6 groups merged groups (a) and (b), which we excluded to keep neighborhood densities in their own group; agglomerating into more than 6 groups separated age of acquisition from group (c), which we excluded given its high correlation values to the rest of group (c). We used scikit-learn's FeatureAgglomeration class for this procedure (Pedregosa et al. 2011).

[12] Note that feature values stem from different data sets which do not always encode the same words. Indeed, we have data on frequency for about 33.5k words, on age of acquisition for 30.1k words, on clustering coefficient for 5.7k words, number of synonyms 111.2k, and orthographic density 17.8k words. Quite often then, not all features are available for a given word in our data set; however this is not problematic since the analysis is done on a per-feature basis, and not all words need be

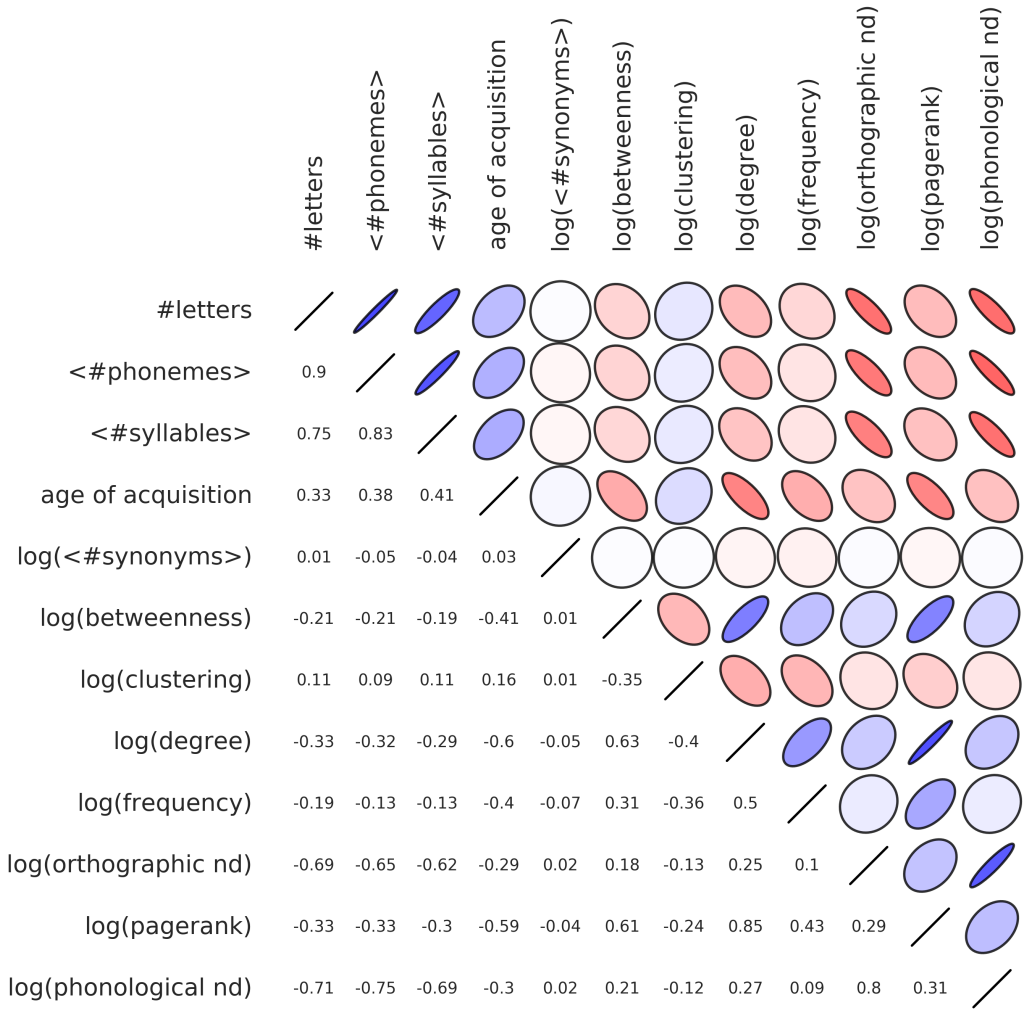


Figure 2.1: Spearman correlations in the initial set of features

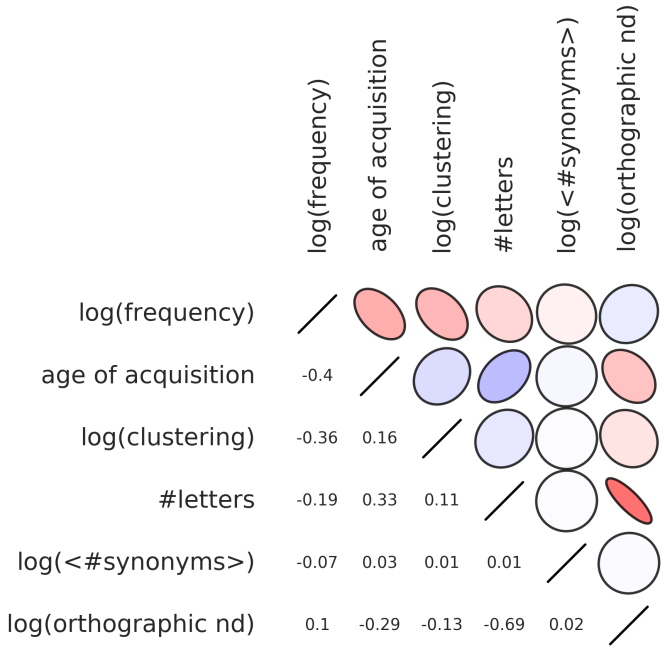


Figure 2.2: Spearman correlations in the filtered set of features

2.3.3 Substitution model

We finally need a substitution detection model, for the quotation data we use presents a challenge: quote-to-quote transformations and substitutions are not explicitly encoded in the data set. More precisely, each set of quotations bears no explicit information about either the authoritative original quotation, or the source quotation(s) each author relied on when creating a new post and reproducing (and possibly altering) that source. In other words we face an inference problem where, given all quotations and their occurrence timestamps, we must estimate which was the originating quotation for each instance of each quotation.

We therefore model the underlying quotation selection process by making a few additional assumptions. Given a particular occurrence of a quotation, the first issue is deciding whether that occurrence is a strict copy of an earlier occurrence, or a substitution from another quotation, or maybe a substitution or copy from quotes appearing outside the data set, that is from a source external to the data collection perimeter. The second issue is deciding which source originated such a substitution when several candidate sources are available.

Let us give an example: say the quotation "These accusations are false and **absurd**" (q) appears in two different blogs on January 19, and the slightly different quotation "These accusations are false and **incoherent**" (q') appears in another blog on the 20th of January. The second occurrence of q can safely be assumed to be a faithful copy of the first one the same day. And since q is fairly prominent when q' first appears, we could assume that the author of q' on the 20th based herself on q, as is shown with a dashed line in Fig. 2.3. Now say a third version, "These **allegations** are false and **incoherent**" (q'') also appears once on January 19 and once on January 20 after q'. Here, q and q'' differ by two substitutions, so we discard the possibility that one was written based on the other (see below for further details). q'' is only one substitution away from q' however, so we could also consider the first occurrence of q'' as a potential source for q' on the 20th. Conversely, the occurrence of q'' on the 20th could be considered as a substitution from q', or as a faithful copy from its initial occurrence on January 19. (Options shown in Fig. 2.3.)

One way to settle these questions is the following: group quote occurrences into fixed bins spanning Δt days (1 day in the implementation), each one representing a unit of time evolution; when a quotation q' appears in bin $t+1$, it is counted as a substitution if it differs from the most frequent quote of the preceding bin t (or a substring thereof) by only one word; if not, q' is not considered to be an instance of substitution. Fig. 2.4a shows the inferences made by such a model. The assumptions it embeds, however, are a subset of a much wider set of possibilities, each leading to alternative inferences.

We identified four binary parameters that differentiate potential models, such that the resulting 16 combinations cover most of the reasonable answers to inference uncertainties. The first two parameters define the preceding time bin from which authors could have drawn a source when producing a new occurrence: (1) **bin positions**, which can be aligned to midnight (as in the model presented above) or kept sliding (for each occurrence, use a bin that ends precisely at that occurrence); (2) **bin span**, which can be $\Delta t = 1$ day (as in the model above) or can be extended up to the very first occurrence in the quotation family. The other two parameters configure rules on the selection of source and destination quotes of a substitution: (3) **candidate sources** can be restricted to the most frequent quotations in the preceding time bin (as in the model above), or not (in which case all quotations in the preceding bin are candidate sources); (4) **candidate destinations** can be restricted to quotations that do not appear in the preceding bin, or without restriction (as in the model above).

---

encoded in all features.

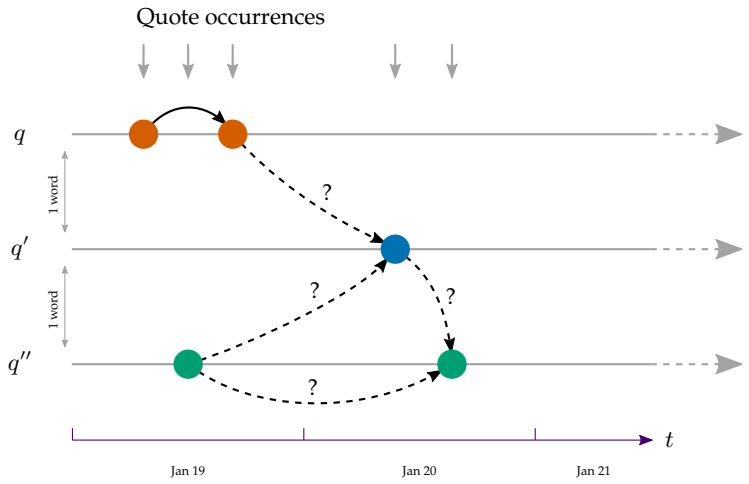


Figure 2.3: **Possible paths from occurrence to occurrence.** q, q' and q'' are three quotation variants belonging to the same cluster. q and q'' differ by two words, but q' differs from both q and q'' by one word. The second occurrence of q can safely be considered a faithful copy of the first, but the occurrences of q' and q'' are uncertain: while the first occurrence of q' is most likely a substitution from q, it could also stem from q''; conversely, the second occurrence of q'' could also be a substitution from q' instead of being a faithful copy of its first occurrence.

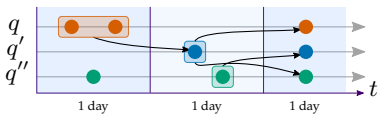
A substitution model, then, is the given of a value for each of those parameters; it considers valid all the substitutions (and only those) where the source and destination follow the rules set out by the parameters. If a destination has substitutions from multiple sources we count a single effective substitution where, for each feature, the value for the effective source word is the average of the values of the candidate source words.

Put shortly a model defines how many times, and under what source and destination conditions, quote occurrences can be counted as substitutions. Fig. 2.4 shows the inferences made by the four models that use bins spanning 1 day aligned to midnight: later occurrences of q' and q'' are counted as substitutions in Fig. 2.4a and Fig. 2.4c, whereas in Fig. 2.4b and Fig. 2.4d they are not.

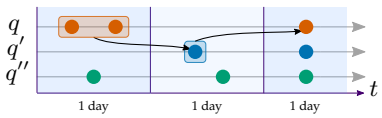
The results reported and discussed in the following sections are valid for all 16 models, and the graphics we present were produced by the model first introduced above. Finally, note that this inference procedure is one of the reasons we restricted our analysis to single-substitutions: looking for more complex transformations would (a) exponentially increase the number of candidate sources for a destination occurrence, which correspondingly reduces the confidence in inferences made, and (b) greatly increase the complexity of the transformation models used to make these inferences.[13]

In practice for the model first introduced above, from the 2.60m initial occurrences spread over 50,427 quotation families, with significant redundancy (many quotes are indeed simple duplicates), we mine 40,868 substitutions. From these substitutions we remove those featuring stopwords, minor spelling changes (e.g. center/centre), abbreviations (e.g. November/Nov or Senator/Sen), spelled

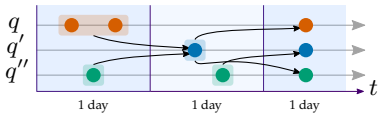
[13]We checked that this restriction does not bias the results discussed below by extending our protocol to two-substitution transformations; the results were unchanged.



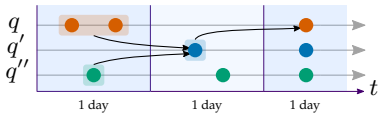
(a) Source must be majority in preceding bin, destination can be anything



(b) Source must be majority in preceding bin, destination must not appear in preceding bin



(c) Source can be anything, destination can be anything



(d) Source can be anything, destination must not appear in preceding bin

Figure 2.4: **Substitution models.** Substitutions inferred by four models in the situation introduced by Fig. 2.3. Each of these models uses bins spanning 1 day aligned to midnight (see the main text for a complete description of parameters). In the top left panel (a), q holds the majority in the first bin and is considered the unique basis for q' in bin 2. q' and q'' have equal maximum frequency in bin 2 however, so both are sources of substitutions towards bin 3. In the top right panel (b), quotes that appear in the preceding bin cannot be the target of a substitution; this removes two substitutions compared to panel (a). In the bottom left panel (c), the majority constraint is lifted compared to panel (a), making q'' in bin 1 a candidate source for q' in bin 2. In the bottom right panel (d), the majority constraint is also lifted compared to panel (a) (adding the same $q'' \rightarrow q'$ substitution as in panel (c)), and the excluded-past constraint is added as in panel (b) (removing the two same substitutions from bin 2 to bin 3 as in panel (b)). If the bins were extended to the beginning of the quotation family, the excluded-past constraint would also remove the $q' \rightarrow q$ substitution from bin 2 to bin 3. In all four panels, a background rectangle or square indicates the quotation is the source of a substitution. A thick border on that rectangle or square indicates the quotation was selected because it has maximum frequency.

out numbers, words unknown to WordNet, and deletions in substrings (which can appear as substitutions of non-deleted words); this eventually yields 6,177 valid substitutions.[14]

2.4 Results

Substitutions usually replace a word with another semantically related word: manual observation of a random subset of 100 substitutions shows that, compared to the word it replaces, the new word often achieves a similar meaning in the context of its sentence while still slightly changing the implications or the attitude expressed by the author.[15] The following examples illustrate this phenomenon:

[14]Manually coding a random subset of 100 substitutions to evaluate this last filter showed that 84 were true negatives, 5 were false positives, and 11 true positives, giving a recall score of .688. Precision was evaluated over a random subset of 100 *kept* substitutions, showing a score of .87. Finally, note that excluding minor spelling changes does not bias our use of orthographic neighborhood density as a feature: out of the first 100 substitutions coded for recall, those with Levenshtein distance equal to 1 (which is what orthographic neighborhood density codes, Marian et al. 2012) were all typos or UK/US spelling changes, neither of which are relevant for this study.

[15]However, the substituted and substituting words are not so often direct synonyms: only a third of all substitutions travel less than 3 hops on the hyponym-hypernym network defined by WordNet (direct synonyms count as 0 hops on this network), meaning that at least two thirds involve non-synonyms. A similar phenomenon is observed on the FA network, where about

- "This is {socialism → welfare} for the rich,
- [The] "perverse logic of {clashes → confrontation} and violence,
- "This {crisis → problem} did not develop overnight and it will not be solved overnight.

Our question concerns the low-level properties of these substitutions: we ask (a) which words are targets of the substitutions and (b) what change these words are subjected to. To this end, we build the following two observables for each word feature. First, we measure which word features are more or less substituted compared to how often they would be if the process were random, in order to capture the susceptibility for words to be the target of a substitution in a quote. Second, we measure the change in word feature upon substitution, looking at the variation of a given feature between start and arrival words. Since sentence context is also central to this process, we extend these two observables by applying them to feature values relative to the distribution of feature values in the sentence in which a word appears.

Note that since we only consider substitutions and not faithful copies, we measure the features of an alteration *knowing that there has been an alteration*, that is we do not take invariant quotations into account. Indeed, in the former case we know there has been a human reformulation, whereas in the latter case we cannot know whether there has been perfect human reformulation or simply digital copy-pasting of a source ("Ctrl-C/Ctrl-V"). Moreover, perfect human reformulation possibly involves different practices than those involved in alteration — for instance drafting before publishing, double-checking sources, proof-reading — and may not be representative of the cognitive processes at work during alteration. The two situations are different enough to be studied separately, and we focus here on the latter.

2.4.1 Susceptibility

We say that a word is *substitutable* if it appears in a quote which undergoes a substitution, whether the substitution operates on that word or on another one. For a given group of words g, say all nouns, or all words in a particular range of values for a feature (e.g. words 2 to 4 letters long), susceptibility is computed as the ratio of s_g, the number of times words of that group are substituted, to s_g^0, the number of times words of that group would be substituted if substitutions fell randomly on substitutable words.[16] That is:

$$\sigma_g = \frac{s_g}{s_g^0}$$

In other words, susceptibility measures how much more or less a group of words g actually gets substituted compared to picking targets at random in quotes undergoing substitutions. By applying this measure to Part-of-Speech (POS) categories and feature bins (e.g. for a feature ϕ and a bin $[a;b]$, $g = \{w|\phi(w) \in [a;b]\}$), susceptibility measures the bias in the selection of start words involved in substitutions, i.e. it measures the preferential selection of some word properties for substitution.

104 clusters have substitutions traveling only 1 hop, 110 traveling 2 hops, 137 traveling 3, 72 traveling 4, and 13 traveling 5.

[16] s_g^0 is computed by summing, over all quotes undergoing a substitution, the ratio of the number of words in a quote that are from group g to the number of words in the same quote that could have been the substitution's target. If, for instance, half the content words of a given quote are nouns, such a quote contributes .5 to the total s_{nouns}^0. Further, to avoid possible autocorrelation effects due to substitutions belonging to the same cluster (which are likely not statistically independent and may lead to overly optimistic confidence intervals), we scale s_g and s_g^0 to count one for each cluster. That is, each quote cluster has a maximum contribution of 1, computed as the average contribution of all substitutions in that cluster.

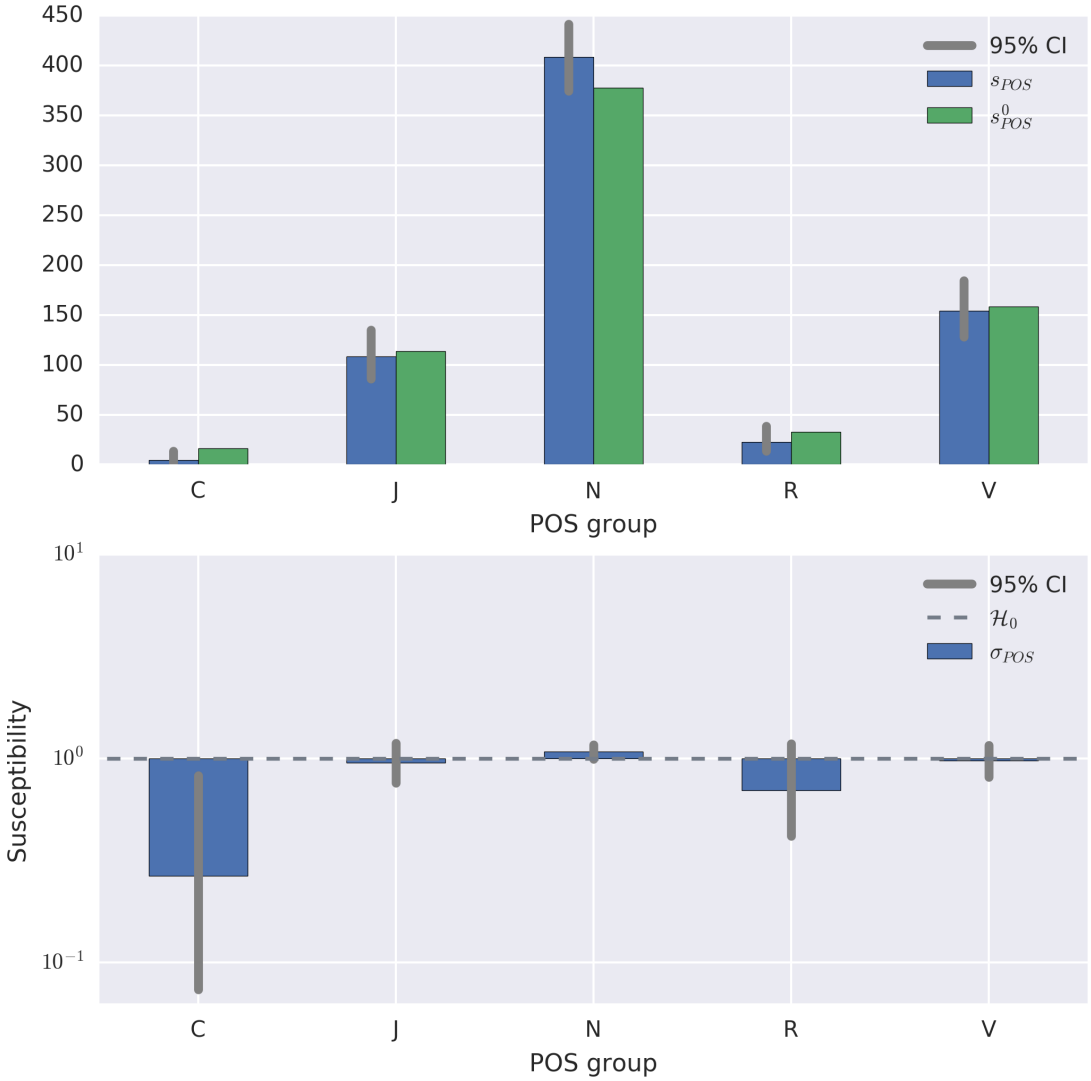


Figure 2.5: **POS-related results:** categories are simplified from the TreeTagger tag set: *C* means *Closed class-like* (see main text for details), *J* means adjective, *N* noun, *R* adverb, and *V* means verb. The top panel shows the actual s_{POS} and s^0_{POS} counts. The bottom panel shows the substitution susceptibility σ_{POS}, which is the ratio between the two previous counts. Confidence intervals are computed with the Goodman (1965) method for multinomial proportions.

Fig. 2.5 gathers the results for POS groups. A Goodman-based multinomial goodness-of-fit test (Goodman 1965) shows that these categories have a significant effect on susceptibility ($p < .05$ in all substitution models), but this seems mostly due to the *Closed class-like*[17] and *Adverb* categories. Indeed, detailing which categories are out of their confidence region under $\mathcal{H}_0$ shows that susceptibility for closed class-likes is significantly below 1 (confirmed in all substitution models), as is that for adverbs in 13 of the 16 substitution models; none of the other categories are significantly different from $\mathcal{H}_0$ (except nouns which appear significantly above 1 in a single substitution model). While we acknowledge the low susceptibilities of adverbs and closed class-likes, these categories concern less than 7% of all substitutions under $\mathcal{H}_0$ (and even less in the actual data); it seems, then, that POS categories do not capture any strong bias in the selection of substitution targets.

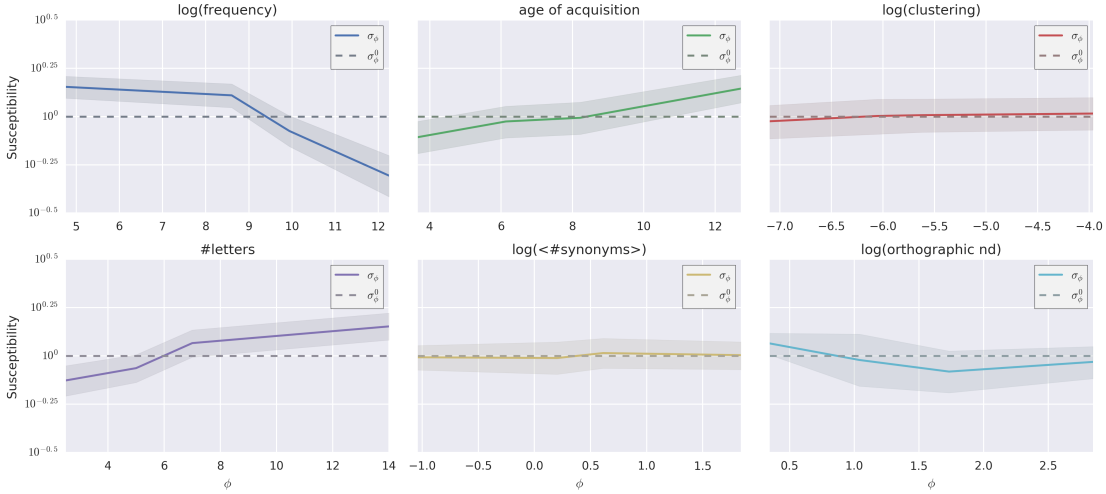


Figure 2.6: **Substitution susceptibility for feature values:** susceptibility to substitution versus feature value of a candidate word for substitution (binned by quartiles), with 95% asymptotic confidence intervals (Goodman-based multinomial).

The results for word features presented in Fig. 2.6, on the other hand, show marked effects for several features. Word frequency, Age of acquisition, and Number of letters each exhibit significant susceptibility variations (Goodman goodness-of-fit with $p < .05$ in all substitution models, $p < .001$ in most) consistent with known effects of those features on recall. High-frequency words, much easier to recall, are substituted about half as much as they would be at random; conversely low-frequency words, harder to recall, are substituted about 50% more than random. Age of acquisition and Number of letters show the opposite pattern, consistent with their negative correlation to word frequency ($-.4$ and $-.19$): words learned before 5 or 6 years old, or made of less than 5 letters, are substituted less than random, whereas words learned after 10 years old, or made of more than 8 letters, are substituted far more than random. Orthographic neighborhood density also shows a slight effect (significant at $p < .05$ in 15 of the 16 substitution models): words with very sparse

[17] The *Closed class-like* category gathers all the POS groups representing closed class words (coordinating conjunctions, prepositions, subordinating conjunctions, modals and possessive endings). These groups, essentially made of stopwords, feature very low counts for both s (substitutions falling on stopwords are filtered out) and s^0 (stopwords are never counted as substitutable). While the susceptibility reported for the remaining words is left unbiased (as s and s^0 are equally affected), they represent a very small portion of all substitutions, which led us to group them together. Finally, we added to this meta-category the few POS groups that cover words entirely excluded from the analysis (foreign words, punctuation symbols and interjections), only sporadically present because of tagging fluctuations; hence the name *Closed class-like*.

neighborhoods are more substituted than random (which may seem counter-intuitive, but is probably because over 70% of those words have 7 letters or more). Clustering coefficient shows no effect on susceptibility, and neither does Number of synonyms: in particular, words with many synonyms do not attract substitutions more than random (in fact, half the substitution models show they have a slight tendency to be substituted less than random).

On the whole, the trends observed are consistent with known effects of word frequency, age of acquisition, and number of letters, indicating that the triggering of a substitution could behave quite similarly to word recall in standard tasks.

2.4.2 Variation

We now examine how words are modified when they are substituted, that is how their features change upon substitution. Considering a word w substituted for w', we measure how a feature ϕ of w varies when it is replaced with w', that is we look at $\phi(w')$ as a function of $\phi(w)$. Averaging this value over all start words such that $\phi(w) = f$ yields the mean variation for that feature value f, that is:[18]

$$\nu_\phi(f) = \langle \phi(w') \rangle_{\{w \to w' | \phi(w) = f\}}$$

We are interested in comparing the value of $\nu_\phi(f)$ to f itself, as this shows whether there is an attraction (or a repulsion) effect towards (respectively from) some values of each feature. In other words, plotting the $y = x$ line, we can see if substitutions tend to attract words towards some typical feature value or not — a standard procedure in the study of dynamical systems.

We also introduce two null hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_{00}$, to compare the actual variation of a word's feature to expected variations under unbiased transformations. $\mathcal{H}_0$ models the situation where the arrival word w' is randomly chosen from the whole pool of words available in the data set for that feature.[19] In this case, since $\phi(w')$ becomes a constant value in the above averaging (by definition w' does not depend on w anymore), the baseline variation under $\mathcal{H}_0$ may be rewritten as:

$$\nu^0_\phi(f) = \langle \phi \rangle$$

$\mathcal{H}_{00}$ models the situation where the arrival word w' is chosen *among immediate synonyms of the start word* w, i.e. an arrival word chosen among semantically plausible though still random words. In this case ν^{00}_ϕ does depend on f:[20]

$$\nu^{00}_\phi(f) = \left\langle \langle \phi(w') \rangle_{w' \in syn(w)} \right\rangle_{\{w | \phi(w) = f\}}$$

This approach yields a fine-grained view of how word features evolve upon substitution, on average, with respect to (a) the original feature (vs. $y = x$), (b) a random arrival (vs. ν^0_ϕ), and (c) an unbiased semantically plausible arrival (vs. ν^{00}_ϕ).

[18] Similarly to what we do for susceptibility, we avoid possible autocorrelation effects by averaging start- and arrival-word features of substitutions from the same cluster into a single aggregate substitution per cluster.

[19] For instance, when considering the feature "Clustering coefficient, the arrival word is randomly chosen among words present in the data set of FA norms.

[20] The actual implementation has an additional level of averaging since WordNet, used to get a word's synonyms, defines several meanings for a single given word, which we have no means of disambiguating. Therefore:

$$\nu^{00}_\phi(f) = \left\langle \left\langle \langle \phi(w') \rangle_{w' \in syn(m)} \right\rangle_{m \in meanings(w)} \right\rangle_{\{w | \phi(w) = f\}}$$

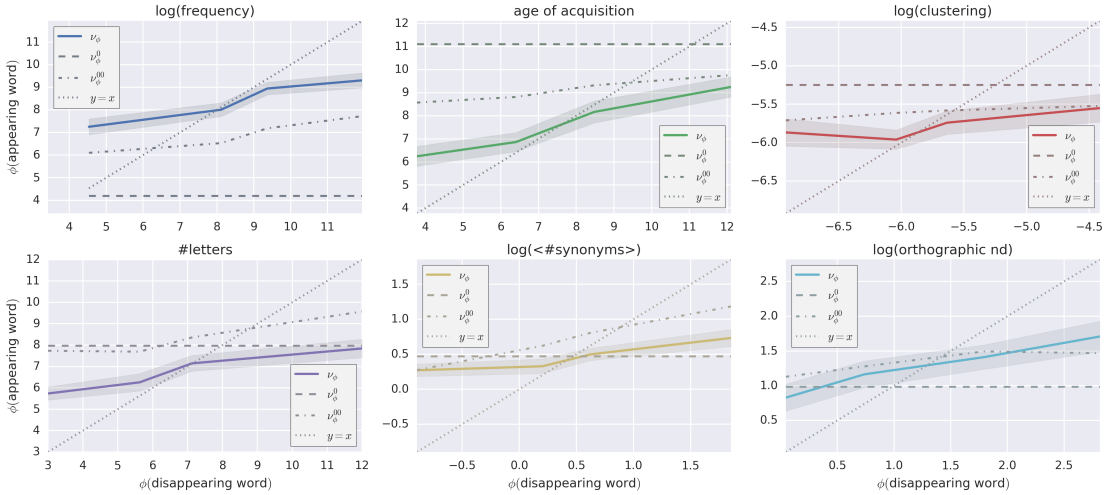


Figure 2.7: **Feature variation upon substitution:** ν_ϕ, average feature value of the appearing word as a function of the feature value of the disappearing word in a substitution (binned by quartiles), with 95% asymptotic confidence intervals based on Student's t-distribution. The overall position of the curve with respect to the dashed line representing $\mathcal{H}_0$ (constant ν_ϕ^0) indicates the direction of the cognitive bias compared to a purely random variation. The position with respect to the dash-dotted line representing $\mathcal{H}_{00}$ (ν_ϕ^{00}) indicates the bias compared to a semantically plausible random variation obtained by choosing a random synonym of the disappearing word. The intersection with $y = x$ marks the attractor value. The fact that all curves have slopes smaller than 1 in absolute value means that the substitution operation is contractile on average: it brings each feature closer to its own specific asymptotic range.

Results are gathered in Fig. 2.7. A first observation is that all graphs show the existence of a unique intersection of ν_ϕ with $y = x$, and the slope of ν_ϕ is smaller than 1 in absolute value, independently of the feature considered. This means that for each feature ϕ, whichever the value $\phi(w)$ of the disappearing word, the appearing word's feature value $\phi(w')$ will, on average, be closer to that feature's intersection of ν_ϕ with $y = x$.[21] In other words, beyond individual variation patterns, the substitution process exhibits a unique attractor for each feature. Note that this is also true under $\mathcal{H}_0$ or $\mathcal{H}_{00}$ (both null hypothesis curves have single intersections with $y = x$ with slopes smaller than 1): the substitution process naturally leads to an attraction even under reasonable random conditions.

Second, the comparison with ν_ϕ^0 and ν_ϕ^{00} shows that there are two classes of attractors, depending on whether: * there is a triple intersection (of $y = x$, ν_ϕ, and ν_ϕ^0 or ν_ϕ^{00}); * or ν_ϕ always remains above or below ν_ϕ^0 and ν_ϕ^{00}.

The first class (Number of synonyms and Orthographic neighborhood density) are features for which the substitution process only brings words slightly closer to ν_ϕ^0 (for Number of synonyms) or ν_ϕ^{00} (for Orthographic neighborhood density), and no uniform bias can be observed. The second class (comprising Word frequency, Age of acquisition, Clustering coefficient, and Number of letters) are features for which the substitution process has a clear bias, positive or negative, with respect to both the purely random situation ($\mathcal{H}_0$) and the semantically plausible random situation ($\mathcal{H}_{00}$).

Word frequency, with ν_ϕ always significantly above ν_ϕ^0 and ν_ϕ^{00}, exhibits a strong bias towards more frequent words. This, in turn, is consistent with the hypothesis that substitution is a recall process, since common words are favored over awkward ones. Age of acquisition, Clustering coefficient and Number of letters, on the other hand, exhibit a clear negative bias for the substitution process (except for high clustering values or very high number of letters). The three curves are significantly below their respective ν_ϕ^0 and ν_ϕ^{00} curves for most start values, which is consistent with the literature on recall: words learned earlier, with lower clustering coefficient or with fewer letters are easier to produce than average (Nelson et al. 2013; Zevin and Seidenberg 2002; Baddeley, Thomson, and Buchanan 1975). All these effects are significant with two-tailed t-tests at $p < .05$ (and more often $p < .001$) and were verified across the 16 substitution models.

To make sure our observations are not the product of correlations or interactions, we model the variations of the 6 features as a linear function of the start word's feature values:

$$\boldsymbol{\phi}(w') - \boldsymbol{\phi}(w) = \boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{\phi}(w)$$

where $\boldsymbol{\phi}$ is the vector of all 6 features of a word, $\boldsymbol{A}$ is an intercept vector, and $\boldsymbol{B}$ is a 6×6 coefficients matrix. This regression achieves an overall R^2 of .33. The corresponding matrix of coefficients $\boldsymbol{B}$ is shown in Fig. 2.8: aside from Age of acquisition and Clustering coefficient on which word frequency has a slight effect, the variation of all other features depends solely on the disappearing word's same feature. In other words there is little to no interaction between a disappearing word's features in determining the variations that that word will undergo when substituted.

To make things concrete, here is an example substitution taking place in the data set. Around mid-November 2008 several media websites reported the following quote from Burmese poet Saw Wai (arrested for one of his poems),

[21] This reasoning is standard in the analysis of dynamical systems (where the same transformation is applied to the whole system over and over), and becomes obvious when one manually simulates a substitution on the graph by picking a start value, using the ν_ϕ curve to obtain the corresponding arrival value, and comparing it to the start value: the arrival value is always closer to the intersection with $y = x$, meaning that that intersection is an attractor point for the substitution process. If the slope of ν_ϕ were greater than one (in absolute value), the arrival value would always be *farther* from the intersection than the start value was, making the intersection with $y = x$ a *repulsor* point. This is how the number of intersections with $y = x$ and the slope of ν_ϕ at those intersections characterize the behavior of substitutions.

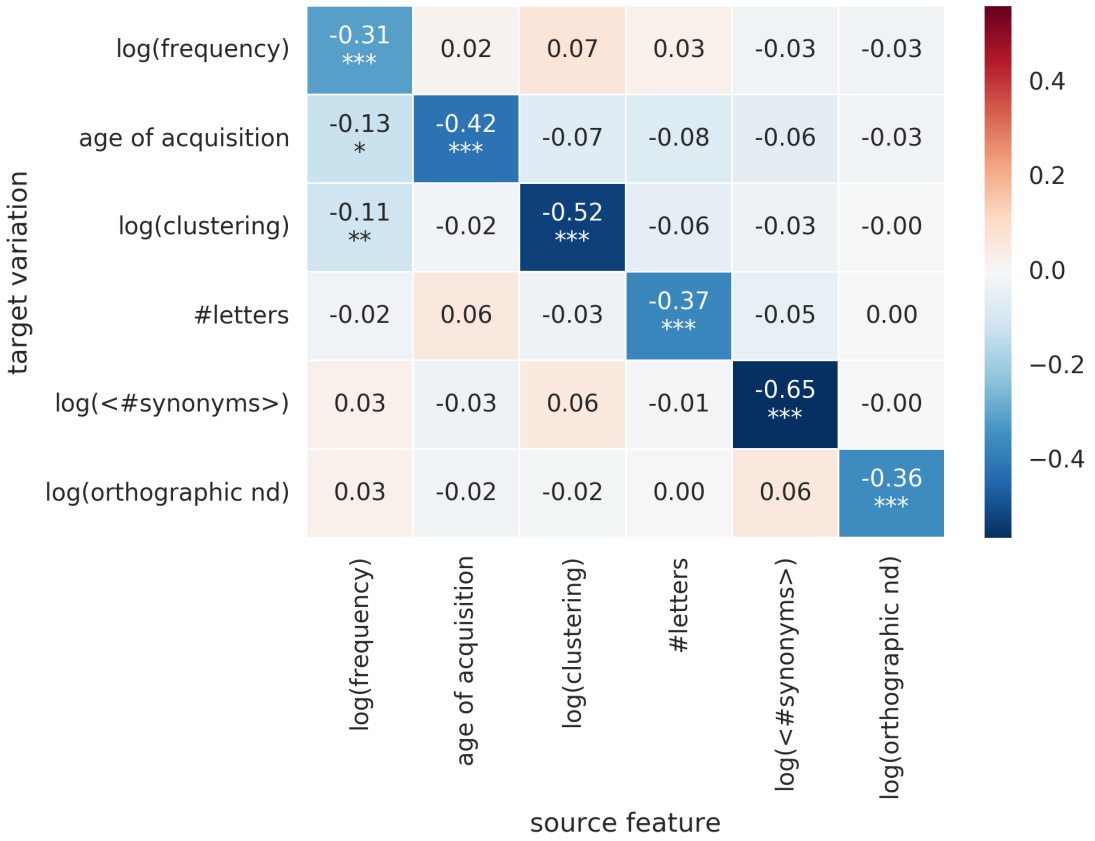


Figure 2.8: **Feature variations regression coefficients**: source feature values (columns) and feature variations (rows) were normalized to $[0; 1]$ to ensure the coefficients are comparable. Significance levels for individual t-tests against the hypothesis of a null coefficient are denoted by stars below the corresponding coefficient (*** for $p \leq .001$, ** for $p \leq .01$, * for $p \leq .05$, and nothing when $p > .05$). Frequency has a slight effect on Age of acquisition and Clustering coefficient, with small coefficients compared to the respective diagonal ones. Aside from those two, only diagonal values are significantly non null.

"Senior general Than Shwe is foolish with power.

and a smaller number of media websites, and blogs, reported the following,

"Senior general Than Shwe is **crazy** with power.

The word *foolish* is acquired at an average of 8.94 years old, appears 675 times in the data set, has a Clustering coefficient of 8.2×10^{-3} and is 7 letters long. The word it was replaced with, *crazy*, is acquired on average at 5.22 years old, appears about 4.1k times in the data set, has a Clustering coefficient of 1.7×10^{-3}, and is 5 letters long. Such a change, though minor in appearance, is a typical example of alteration along the lines shown by our results.

2.4.3 Sentence context

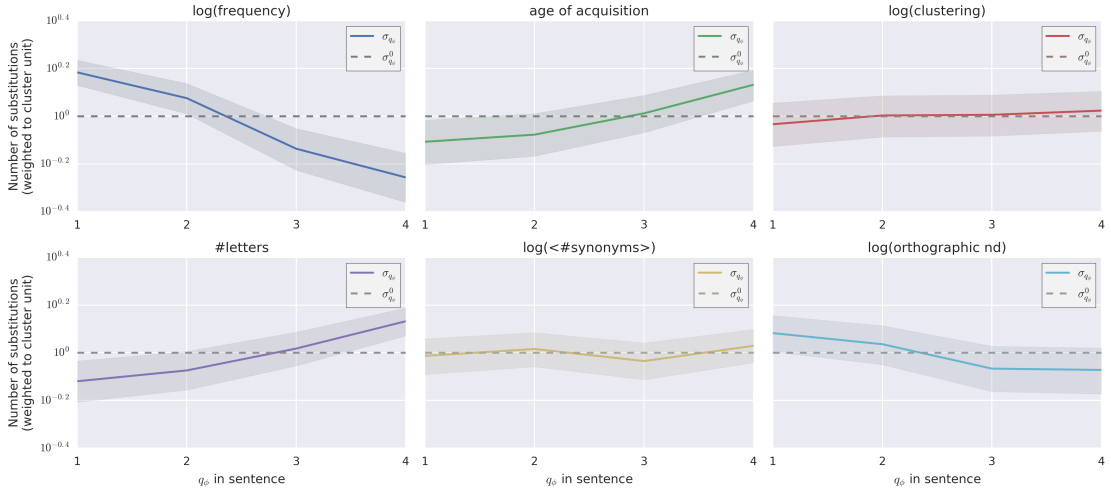


Figure 2.9: **Substitution susceptibility for in-quote feature quartiles:** susceptibility to substitution versus quartile of the feature distribution in the originating quote, with 95% asymptotic confidence intervals (Goodman-based multinomial).

The alterations we study are always made in a context, that is while the author is writing. We wish to ask, therefore, if taking that context into account can provide more insight into the substitution process. To do so we adapt the two observables presented above to capture some of the relationships between a word and the sentence it appears in.

Let us start with the first one: given a feature ϕ, we define the context-relative susceptibility to substitution with the following three steps. (1) For each quote in which a substitution appears, compute the distribution of ϕ values in that quote (excluding stopwords) and divide it into quartiles. (2) Count how many times each quartile (first, second, third or fourth) contains a word that is substituted. This procedure tells us, for $i \in \{1; 2; 3; 4\}$, how many times substitutions fall in the i-th quartile of each in-quote distribution of ϕ; in other words it gives us the numerator s_{q_i} for the computation of susceptibility, where q_i represents the i-th quartile of the distributions of ϕ in the quotes. (3) Finally divide each quartile count by its corresponding $s^0_{q_i}$, that is the number of times the i-th quartile would receive substitutions if targets were picked at random; since the random situation would equally distribute a fourth of all substitutions to each quartile, we divide by the number of

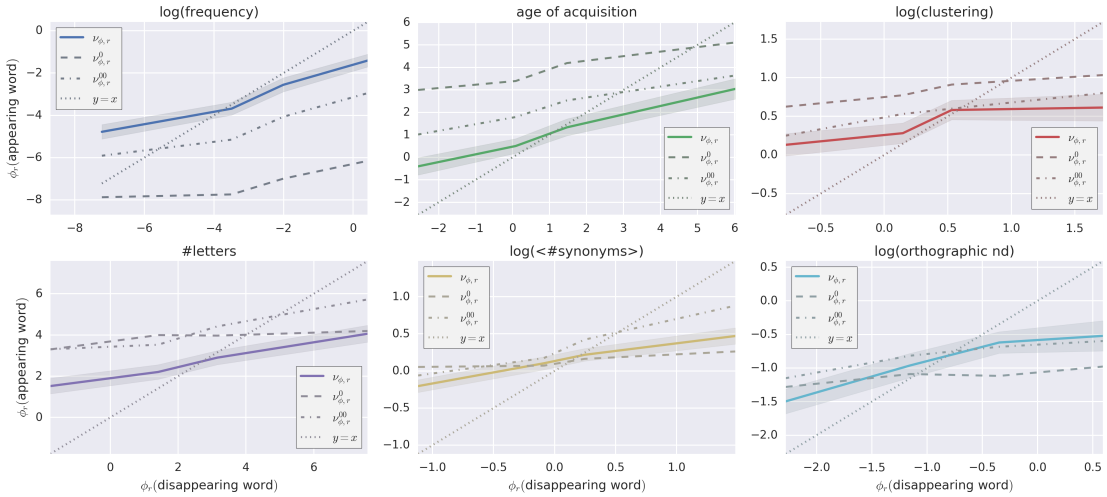


Figure 2.10: **Sentence-relative feature variation:** ν_{ϕ_r}, average sentence-relative feature value of the appearing word as a function of the sentence-relative value of the disappearing word (binned by quartiles), with 95% asymptotic confidence intervals based on Student's t-distribution. $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ are similarly converted to be sentence-relative. Attraction, magnitude and direction of bias with respect to null hypotheses are similar to Fig. 2.7. However, attractors are always positioned between sentence median ($y = 0$) on one side and $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ on the other side. Clustering coefficient, Number of synonyms and Orthographic neighborhood density are limit cases, with triple intersections with one of the null hypothesis curves.

substitutions divided by 4. Taking for instance word frequency, this measure tells us if words that have high- or low-frequency *compared to the quote they appear in* are more or less substituted than at random.

Surprisingly the results for this measure are no different from the context-free measure, as can be seen in Fig. 2.9: low-frequency words compared to the sentence are substituted much more than higher-frequency words, words learned earlier than the rest of the sentence are substituted less than words learned later, shorter words less than longer words, and words with scarce neighborhoods slightly more than words with denser neighborhoods. Clustering coefficient and Number of synonyms are, here again and across all substitution models, not significantly different from $\mathcal{H}_0$: with or without context, they do not seem relevant to the selection of substitution targets.

Feature variation is more easily extended to the context-relative case. To do so we consider all feature values relative to the median word feature in the sentence. That is, in all the equations of the previous subsection we replace $\phi(w)$ with:

$$\phi_r(w) = \phi(w) - \text{median}\,\{\phi(w)|w \in \text{sentence}\}$$

ν_ϕ, ν^0_ϕ and ν^{00}_ϕ each transpose to ν_{ϕ_r}, $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ (note that $\nu^0_{\phi_r}$ now depends on w since it is sentence-relative, whereas ν^0_ϕ did not).

The results for sentence-relative feature variations are gathered in Fig. 2.10. Here too, the behavior is strikingly similar to the context-free measure: a single attractor is visible for each feature, and the magnitude and direction of biases are near-identical to those for the previous measure. The values of the appearing words give an additional insight into the process, however: the attractor value

of a feature, at the intersection of ν_{ϕ_r} and $y = x$, is always between the sentence median, on one side, and $\nu^0_{\phi_r}$ and $\nu^{00}_{\phi_r}$ on the other side (for Number of synonyms it is a triple intersection with $\nu^0_{\phi_r}$; for Clustering coefficient and Orthographic neighborhood density, a triple intersection with $\nu^{00}_{\phi_r}$). Substitutions, therefore, seem to attract words closer to the sentence median than what a random process would do. This is true with respect to both null hypotheses (semantically plausible or not) for Frequency, Age of Acquisition and Number of letters, and true with respect to at least one of the two null hypotheses for the remaining features.

On the whole, we observe a clear attraction pattern for each feature, with two different classes corresponding to the psychological relevance of each feature for the substitution process. More awkward words along relevant features (less frequent, learned later, or made of more letters), both globally and with respect to the sentence they appear in, are substituted more often than what would happen if targets were picked randomly in the sentences; conversely, more common words are substituted less. Finally, across all features, substituted words are attracted towards a point closer to the sentence median than what a random process, semantically plausible or not, would do.

2.5 Discussion

We initially aimed to connect the field of cultural evolution with psycholinguistics by asking if cultural attractors appear in a corpus of online news-related quotes gradually transformed by low-level biases. The data set we used imposed a few constraints on our analysis: first, it was necessary to infer source-destination links, an operation made more reliable when restricting the scope of transformations to very simple cases, which we did by focusing on single word substitutions. Second, contrary to laboratory experiments which produce data made of many repeated measures on a small number of cases (e.g. a given list of words), we have a great number of different cases (one case per cluster in which substitutions are found, i.e. 698 cases), with very few measures on each of them (average 9, median 5). This rendered the prediction of individual words impractical: if we cannot compute a percentage of explained data for a given case, any approximate prediction will be heavily underestimated. This last factor, added to the potential for variation of external conditions when authors wrote the quotes, led us to use word features to analyze the transformations by aggregating over individual cases.

By characterizing substitutions with 6 features on the disappearing word, we show that authors preferentially substitute words known for being harder to recall: most prominently words with low frequency (Gregg 1976), learned later (Dewhurst, Hitch, and Barry 1998), or made up of more letters (Nickels and Howard 2004), both globally and in comparison to the sentence they appear in. Further characterizing the substitutions by examining the variation of word features from disappearing to appearing words, we show: (a) that the operation is contractile on average, that is words are brought closer to an attractor point on each feature; (b) that authors produce words that are easier to remember than the average of synonyms of the disappearing word (a fact that is reflected in the position of the attraction point).

We do not actually observe quotes converging on a global scale towards attractors in their various dimensions. Indeed the limits of the data set do not allow us to infer chains of substitutions, and substitutions themselves are not the only type of transformation at work in the data set. Nonetheless, these findings (a) bring light to this simple type of transformation, and (b) are consistent with known psycholinguistic effects, with the hypothesis of cultural attractors in representations from everyday life, and with the lineage specificity discussed in the iterated learning literature (Claidière

et al. 2014; Cornish, Smith, and Kirby 2013). They are obtained by successfully applying knowledge from cognitive science to real-life complex data, a task that remains a challenge in the study of cultural evolution. More broadly, we believe that applying such data mining tools to manage the complexity of real-life data is a promising approach for the joint analysis of cognitive science and culture.

In the simple case presented here, however, much remains to be explored. Since it is clear that observing cognitive biases in such data is now possible, questions addressed in controlled laboratory situations could be opened by further research. One question concerns the influence of the context surrounding a quote, be it in terms of other quotes preceding it temporally or of text surrounding it in a post. A first step could be the application of results from Zaromb et al. (2006) who have shown, in the simpler task of recall of random word lists, that the source of prior-list intrusions can be predicted based on the associations those preceding lists have formed: in our case, a substitution could be triggered and directed by associations formed by preceding context. A further step would be to follow what Cornish, Smith, and Kirby (2013) have shown about reciprocal influences between context and transformations (in their case, with transmission chains of artificial content). Indeed substitutions, and more generally all transformations, also participate in creating the context for later quotes. One can ask, therefore, what are the reciprocal effects between, on one side, the corpus-level evolution of quotes through iterated transformations, and on the other side, a gradual change in the properties of transformations operated because of the evolution of surrounding context. Such interactions have been shown to underlie the lineage specificity observed in transmission chains (Claidière et al. 2014). Exploring how similar loop interactions happen in real-life data could indeed be the next step in understanding the coevolution of cultural content and the ways in which it is transformed. In our particular case, such insight could shed some light on how the feature attractors examined in this paper actually emerge, and help assess their potential role on this coevolution.

2.6 Concluding remarks

The theory of Epidemiology of Representations proposes a unifying framework for the study of cultural evolution. One of its core claims, the existence of cultural attractors, has been both a challenge to test empirically and a fruitful line to pursue in the study of cultural evolution. We aimed to contribute to testing this hypothesis by studying a simple everyday-life task where individuals are implicitly trying to reproduce quotations. To some extent, our work amounts to an out-of-laboratory experiment where we examine the influence of well-known word features on the accuracy of reproduction of short sentences. Our analysis of substitutions shows that words are attracted, in each dimension, to feature-specific values. Furthermore, the features' known effects in psycholinguistic experiments are reflected in the biases of these attraction points, meaning that the evolution of such quotations can be partially explained by known low-level cognitive biases. We believe that such an approach, which combines psycholinguistic knowledge and data mining tools, can be fruitfully developed to improve the study of cultural attractors and explore the reciprocal influences of cognition and culture.

Let us conclude by noting that the question of short- and long-term cultural evolution, and the approaches to study them, are becoming increasingly relevant to other fields. In biology in particular, work on evo-devo and non-genetic inheritance has accumulated evidence that is poorly accounted for by the modern synthesis of biological evolution, and is creating a demand for new or extended approaches to joint cultural and biological evolution (see Gilbert, Bosch, and Ledón-Rettig 2015 for instance). Such an approach has long been called upon by anthropologists like Ingold (2004; 1998),

in line with Mauss' initial works (Mauss 1936), and the question is not entirely foreign to the enactive-representational debate in cognitive science. The study of cultural evolution will most likely benefit greatly from the growing interactions between these disciplines.

Acknowledgements

We are warmly grateful to Ana Sofia Morais for her precious feedback and advice on this research, and to Telmo Menezes, Jean-Philippe Cointet, Jean-Pierre Nadal, Sharon Peperkamp, Nicolas Claidière and Nicolas Baumard for useful suggestions and comments.

This work has also been partially supported by the French National Agency of Research (ANR) through the grant Algopol (ANR-12-CORD-0018).

Software colophon

Finally, this paper was developed using Python's scientific computing ecosystem (Millman and Aivazis 2011). In particular, we directly used NumPy and SciPy (van der Walt, Colbert, and Varoquaux 2011), Matplotlib (Hunter 2007), Pandas (McKinney 2010), scikit-learn (Pedregosa et al. 2011), NetworkX (Hagberg, Schult, and Swart 2008), NLTK (Bird, Klein, and Loper 2009), IPython (Pérez and Granger 2007), and many other libraries from the Python ecosystem. The software and analyses written for the paper are documented and published under a Free Software license. They can be found at github.com/wehlutyk/brainscopypaste.

Chapter 3

Gistr

Chapter 4

Discussion

Chapter 5

Conclusion

References

Acerbi, Alberto. 2016. 'A Cultural Evolution Approach to Digital Media'. *Frontiers in Human Neuroscience* 10 (December). doi:10.3389/fnhum.2016.00636.

Acerbi, Alberto, and Alex Mesoudi. 2015. 'If We Are All Cultural Darwinians What's the Fuss About? Clarifying Recent Disagreements in the Field of Cultural Evolution'. *Biology & Philosophy* 30 (4): 481–503. doi:10.1007/s10539-015-9490-2.

Acerbi, Alberto, and Jamshid J. Tehrani. 2017. 'Did Einstein Really Say That? Testing Content Versus Context in the Cultural Selection of Quotations'. *SocArXiv Preprints*, April. https://osf.io/preprints/socarxiv/x8db2.

Acerbi, Alberto, and Claudio Tennie. 2016. 'The Role of Redundant Information in Cultural Transmission and Cultural Stabilization'. *Journal of Comparative Psychology* 130 (1): 62–70. doi:10.1037/a0040094.

Adamic, Lada A., Thomas M. Lento, Eytan Adar, and Pauline C. Ng. 2016. 'Information Evolution in Social Networks'. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 473–82. New York, NY, USA: ACM. doi:10.1145/2835776.2835827.

Althoff, Tim, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. 'How to Ask for a Favor: A Case Study on the Success of Altruistic Requests', May. http://arxiv.org/abs/1405.3282.

Aunger, Robert. 2000. *Darwinizing Culture: The Status of Memetics as a Science*. Oxford, NY: Oxford University Press.

Austerweil, Joseph L., Joshua T Abbott, and Thomas L. Griffiths. 2012. 'Human Memory Search as a Random Walk in a Semantic Network'. In *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 3041–9. Red Hook, NY: Curran Associates, Inc. http://papers.nips.cc/paper/4761-human-memory-search-as-a-random-walk-in-a-semantic-network.pdf.

Baddeley, Alan D., Neil Thomson, and Mary Buchanan. 1975. 'Word Length and the Structure of Short-Term Memory'. *Journal of Verbal Learning and Verbal Behavior* 14 (6): 575–89. doi:10.1016/S0022-5371(75)80045-4.

Bakshy, Eytan, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. 'Everyone's an Influencer: Quantifying Influence on Twitter'. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, edited by Irwin King, Wolfgang Nejdl, and Hang Li, 65–74. WSDM '11. New York, NY, USA: ACM. doi:10.1145/1935826.1935845.

Bakshy, Eytan, Brian Karrer, and Lada A. Adamic. 2009. 'Social Influence and the Diffusion of User-Created Content'. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, 325–34. New York,

NY, USA: ACM. doi:10.1145/1566374.1566421.

Bangerter, Adrian. 2000. 'Transformation Between Scientific and Social Representations of Conception: The Method of Serial Reproduction'. *British Journal of Social Psychology* 39 (4): 521–35. doi:10.1348/014466600164615.

Barrett, Justin L., and Melanie A. Nyhof. 2001. 'Spreading Non-Natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials'. *Journal of Cognition and Culture* 1 (1): 69–100. doi:10.1163/156853701300063589.

Bartlett, Sir Frederic Charles. (1932) 1995. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

Baumard, Nicolas, Jean-Baptiste André, and Dan Sperber. 2013. 'A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice'. *Behavioral and Brain Sciences* 36 (01): 59–78.

Baumard, Nicolas, Alexandre Hyafil, Ian Morris, and Pascal Boyer. 2015. 'Increased Affluence Explains the Emergence of Ascetic Wisdoms and Moralizing Religions'. *Current Biology* 25 (1): 10–15. doi:10.1016/j.cub.2014.10.063.

Beer, Randall D., and Paul L. Williams. 2015. 'Information Processing and Dynamics in Minimally Cognitive Agents'. *Cognitive Science* 39 (1): 1–38. doi:10.1111/cogs.12142.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media, Inc.

Bloch, Maurice. 2000. 'A Well-Disposed Social Anthropologist's Problems with Memes'. In *Darwinizing Culture: The Status of Memetics as a Science*, edited by Robert Aunger, 189–204. Oxford, NY: Oxford University Press.

Bourdieu, Pierre. 1980. *Le sens pratique*. Paris: Editions de Minuit.

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

———. 2005. *The Origin and Evolution of Cultures*. New York, N.Y. ; Oxford: Oxford university press.

Boyer, Pascal. 2001. *Religion Explained: The Evolutionary Origins of Religious Thought*. Basic Books.

Caldwell, Christine A., and Ailsa E. Millen. 2008a. 'Experimental Models for Testing Hypotheses About Cumulative Cultural Evolution'. *Evolution and Human Behavior* 29 (3): 165–71. doi:10.1016/j.evolhumbehav.2007.12.001.

———. 2008b. 'Studying Cumulative Cultural Evolution in the Laboratory'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1509): 3529–39. doi:10.1098/rstb.2008.0133.

Caldwell, Christine A., and Kenny Smith. 2012. 'Cultural Evolution and Perpetuation of Arbitrary Communicative Conventions in Experimental Microsocieties'. *PLOS ONE* 7 (8): e43807. doi:10.1371/journal.pone.0043807.

Carr, Jon W., Kenny Smith, Hannah Cornish, and Simon Kirby. 2017. 'The Cultural Evolution of Structured Languages in an Open-Ended, Continuous World'. *Cognitive Science* 41 (4): 892–923. doi:10.1111/cogs.12371.

Cavalli-Sforza, L. L, and Marcus W Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, N.J.: Princeton University Press.

Chan, Kit Ying, and Michael S. Vitevitch. 2010. 'Network Structure Influences Speech Production'.

Cognitive Science 34 (4): 685–97. doi:10.1111/j.1551-6709.2010.01100.x.

Chemero, Tony, and Michael Silberstein. 2008. 'After the Philosophy of Mind: Replacing Scholasticism with Science'. *Philosophy of Science* 75 (1): 1–27.

Claidière, Nicolas, and Dan Sperber. 2007. 'The Role of Attraction in Cultural Evolution'. *Journal of Cognition and Culture* 7 (1): 89–111. doi:10.1163/156853707X171829.

Claidière, Nicolas, Thomas C. Scott-Phillips, and Dan Sperber. 2014. 'How Darwinian Is Cultural Evolution?' *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1642): 20130368. doi:10.1098/rstb.2013.0368.

Claidière, Nicolas, Kenny Smith, Simon Kirby, and Joël Fagot. 2014. 'Cultural Evolution of Systematically Structured Behaviour in a Non-Human Primate'. *Proceedings of the Royal Society of London B: Biological Sciences* 281 (1797): 20141541. doi:10.1098/rspb.2014.1541.

Claidière, Nicolas, Emmanuel Trouche, and Hugo Mercier. 2017. 'Argumentation and the Diffusion of Counter-Intuitive Beliefs'. *Manuscript Submitted for Publication*. https://sites.google.com/site/hugomercier/Argumentation%20and%20counter-intuitive%20beliefs.pdf.

Clark, Andy, and David Chalmers. 1998. 'The Extended Mind'. *Analysis* 58 (1): 7–19. http://www.jstor.org.ins2i.bib.cnrs.fr/stable/3328150.

Cointet, J. P., and C. Roth. 2009. 'Socio-Semantic Dynamics in a Blog Network'. In *International Conference on Computational Science and Engineering, 2009. CSE '09*, edited by Alex Pentland, Justin Zahn, and Daniel Zeng, 4:114–21. Washington, DC: IEEE Computer Society. doi:10.1109/CSE.2009.105.

Cointet, Jean-Philippe, and Camille Roth. 2007. 'How Realistic Should Knowledge Diffusion Models Be?' *Journal of Artificial Societies and Social Simulation* 10 (3): 5. http://jasss.soc.surrey.ac.uk/10/3/5.html.

Cornish, Hannah, Kenny Smith, and Simon Kirby. 2013. 'Systems from Sequences: An Iterated Learning Account of the Emergence of Systematic Structure in a Non-Linguistic Task'. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, edited by Markus Knauff, Michael Pauen, Natalie Sebanz, and Ipke Wachsmuth. Austin, TX: Cognitive Science Society.

Croft, William. 2013. 'An Evolutionary Model of Language Change and Language Structure'. In *Explaining Language Change: An Evolutionary Approach*, Draft 2nd edition (revised). http://www.unm.edu/~wcroft/Papers/ELC2-Chap02.pdf.

Cuffari, Elena Clare, Ezequiel Di Paolo, and Hanne De Jaegher. 2015. 'From Participatory Sense-Making to Language: There and Back Again'. *Phenomenology and the Cognitive Sciences* 14 (4): 1089–1125. doi:10.1007/s11097-014-9404-9.

Danescu-Niculescu-Mizil, Cristian, Justin Cheng, Jon Kleinberg, and Lillian Lee. 2012. 'You Had Me at Hello: How Phrasing Affects Memorability'. In *ACL '12 Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers*, edited by Haizhou Li, Chin-Yew Lin, and Miles Osborne, 1:892–901. Stroudsburg, PA: ACM. http://arxiv.org/abs/1203.6360.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg. 2011. 'Echoes of Power: Language Effects and Power Differences in Social Interaction', December. http://arxiv.org/abs/1112.3670.

Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. 'A Computational Approach to Politeness with Application to Social Factors', June.

http://arxiv.org/abs/1306.6078.

Dawkins, Richard. (1976) 2006. *The Selfish Gene*. Oxford; New York: Oxford University Press.

De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. 'Participatory Sense-Making'. *Phenomenology and the Cognitive Sciences* 6 (4): 485–507. doi:10.1007/s11097-007-9076-9.

Deese, James. 1959. 'On the Prediction of Occurrence of Particular Verbal Intrusions in Immediate Recall'. *Journal of Experimental Psychology* 58 (1): 17–22. doi:10.1037/h0046671.

Dewhurst, Stephen A., Graham J. Hitch, and Christopher Barry. 1998. 'Separate Effects of Word Frequency and Age of Acquisition in Recognition and Recall'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24 (2): 284–98. doi:10.1037/0278-7393.24.2.284.

Di Paolo, Ezequiel A. 2005. 'Autopoiesis, Adaptivity, Teleology, Agency'. *Phenomenology and the Cognitive Sciences* 4 (4): 429–52. doi:10.1007/s11097-005-9002-y.

Durkheim, Emile. (1897) 2012. *Le Suicide: Étude de Sociologie*. Project Gutenberg. http://www.gutenberg.org/ebooks/40489.

Eriksson, Kimmo, and Julie C. Coultas. 2012. 'The Advantage of Multiple Cultural Parents in the Cultural Transmission of Stories'. *Evolution and Human Behavior* 33 (4): 251–59. doi:10.1016/j.evolhumbehav.2011.10.002.

———. 2014. 'Corpses, Maggots, Poodles and Rats: Emotional Selection Operating in Three Phases of Cultural Transmission of Urban Legends'. *Journal of Cognition and Culture* 14 (1-2): 1–26. doi:10.1163/15685373-12342107.

Fay, Nicolas, Simon Garrod, Leo Roberts, and Nik Swoboda. 2010. 'The Interactive Evolution of Human Communication Systems'. *Cognitive Science* 34 (3): 351–86. doi:10.1111/j.1551-6709.2009.01090.x.

Fisher, Ronald Aylmer. 1930. *The Genetical Theory of Natural Selection*. Oxford, UK: Clarendon Press. http://www.biodiversitylibrary.org/title/27468.

Fodor, Jerry A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.

Freeman, Linton C. 1977. 'A Set of Measures of Centrality Based on Betweenness'. *Sociometry* 40 (1): 35–41. doi:10.2307/3033543.

Fuentes, Agustín. 2006. 'Evolution Is Important but It Is Not Simple: Defining Cultural Traits and Incorporating Complex Evolutionary Theory'. *Behavioral and Brain Sciences* 29 (04): 354–55.

———. 2009. 'A New Synthesis. Resituating Approaches to the Evolution of Human Behaviour'. *Anthropology Today* 25 (3): 12–17.

Futuyma, Douglas J. 2005. *Evolution*. Sunderland, MA, USA: Sinauer Associates.

Galantucci, Bruno. 2005. 'An Experimental Study of the Emergence of Human Communication Systems'. *Cognitive Science* 29 (5): 737–67. doi:10.1207/s15516709cog0000_34.

Galantucci, Bruno, Simon Garrod, and Gareth Roberts. 2012. 'Experimental Semiotics'. *Language and Linguistics Compass* 6 (8): 477–93. doi:10.1002/lnc3.351.

Garlock, Victoria M., Amanda C. Walley, and Jamie L. Metsala. 2001. 'Age-of-Acquisition, Word Frequency, and Neighborhood Density Effects on Spoken Word Recognition by Children and Adults'.

Journal of Memory and Language 45 (3): 468–92. doi:10.1006/jmla.2000.2784.

Garrod, Simon, Nicolas Fay, John Lee, Jon Oberlander, and Tracy MacLeod. 2007. 'Foundations of Representation: Where Might Graphical Symbol Systems Come from?' *Cognitive Science* 31 (6): 961–87. doi:10.1080/03640210701703659.

Gauld, Alan, and Geoffrey M. Stephenson. 1967. 'Some Experiments Relating to Bartlett's Theory of Remembering'. *British Journal of Psychology* 58 (1-2): 39–49. doi:10.1111/j.2044-8295.1967.tb01054.x.

Giddens, Anthony. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge, UK; Oxford, UK: Polity Press & Blackwell.

Gilbert, Scott F., Thomas C. G. Bosch, and Cristina Ledón-Rettig. 2015. 'Eco-Evo-Devo: Developmental Symbiosis and Developmental Plasticity as Evolutionary Agents'. *Nature Reviews Genetics* 16 (10): 611–22. doi:10.1038/nrg3982.

Goodman, Leo A. 1965. 'On Simultaneous Confidence Intervals for Multinomial Proportions'. *Technometrics* 7 (2): 247–54. doi:10.1080/00401706.1965.10490252.

Gregg, Vernon. 1976. 'Word Frequency, Recognition and Recall'. In *Recall and Recognition*, edited by John Brown, 183–216. Oxford, UK: John Wiley & Sons.

Griffiths, Paul E., and Russell D. Gray. 2005. 'Discussion: Three Ways to Misunderstand Developmental Systems Theory'. *Biology and Philosophy* 20 (2-3): 417–25. doi:10.1007/s10539-004-0758-1.

Griffiths, Paul, and Karola Stotz. 2013. *Genetics and Philosophy: An Introduction*. http://dx.doi.org/10.1017/CBO9780511744082.

Griffiths, Thomas L., and Michael L. Kalish. 2007. 'Language Evolution by Iterated Learning with Bayesian Agents'. *Cognitive Science* 31 (3): 441–80. doi:10.1080/15326900701326576.

Griffiths, Thomas L., Brian R. Christian, and Michael L. Kalish. 2008. 'Using Category Structures to Test Iterated Learning as a Method for Identifying Inductive Biases'. *Cognitive Science* 32 (1): 68–107. doi:10.1080/03640210701801974.

Griffiths, Thomas L., Michael L. Kalish, and Stephan Lewandowsky. 2008. 'Theoretical and Empirical Evidence for the Impact of Inductive Biases on Cultural Evolution'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363 (1509): 3503–14. doi:10.1098/rstb.2008.0146.

Griffiths, Thomas L., Mark Steyvers, and Alana Firl. 2007. 'Google and the Mind: Predicting Fluency with PageRank'. *Psychological Science* 18 (12): 1069–76. doi:10.1111/j.1467-9280.2007.02027.x.

Gruhl, Daniel, R. Guha, David Liben-Nowell, and Andrew Tomkins. 2004. 'Information Diffusion Through Blogspace'. In *Proceedings of the 13th International Conference on World Wide Web*, edited by Stuart I. Feldman, Mike Uretsky, Marc Najork, and Craig E. Wills, 491–501. New York, NY: ACM. doi:10.1145/988672.988739.

Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart. 2008. 'Exploring Network Structure, Dynamics, and Function Using NetworkX'. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, edited by Gaël Varoquaux, Travis Vaught, and Jarrod Millman, 11–15. Pasadena, CA.

Haldane, John Burdon Sanderson. 1932. *The Causes of Evolution*. London, UK; New York, NY, USA; Toronto, Canada: Longmans Green & Co.

Hall, K. R. L. 1950. 'The Effect of Names and Titles Upon the Serial Reproduction of Pictorial and

Verbal Material'. *British Journal of Psychology. General Section* 41 (3-4): 109–21. doi:10.1111/j.2044-8295.1950.tb00269.x.

Heath, Chip, Chris Bell, and Emily Sternberg. 2001. 'Emotional Selection in Memes: The Case of Urban Legends'. *Journal of Personality and Social Psychology* 81 (6): 1028–41. doi:10.1037/0022-3514.81.6.1028.

Howard, Marc W, and Michael J Kahana. 2002. 'When Does Semantic Similarity Help Episodic Retrieval?' *Journal of Memory and Language* 46 (1): 85–98. doi:10.1006/jmla.2001.2798.

Hunter, John D. 2007. 'Matplotlib: A 2d Graphics Environment'. *Computing in Science & Engineering* 9 (3): 90–95. doi:10.1109/MCSE.2007.55.

Ingold, Tim. 1998. 'From Complementarity to Obviation: On Dissolving the Boundaries Between Social and Biological Anthropology, Archaeology and Psychology'. *Zeitschrift Für Ethnologie* 123 (1): 21–52.

———. 2001. 'From the Transmission of Representations to the Education of Attention'. In *The Debated Mind: Evolutionary Psychology Versus Ethnography*, edited by Harvey Whitehouse, 113–53. Oxford & New York: Berg. http://lchc.ucsd.edu/MCA/Paper/ingold/ingold1.htm.

———. 2004. 'Beyond Biology and Culture. the Meaning of Evolution in a Relational World'. *Social Anthropology* 12 (2): 209–21. doi:10.1111/j.1469-8676.2004.tb00102.x.

———. 2007. 'The Trouble with "Evolutionary Biology"'. *Anthropology Today* 23 (2): 13–17. doi:10.1111/j.1467-8322.2007.00497.x.

Jablonka, Eva M. 2001. 'The Systems of Inheritance'. In *Cycles of Contingency: Developmental Systems and Evolution*, edited by Susan Oyama, Paul Griffiths, and Russell D Gray, 99–116. Cambridge, Mass.: MIT Press.

Jefferies, Elizabeth, Matthew A. Lambon Ralph, and Alan D. Baddeley. 2004. 'Automatic and Controlled Processing in Sentence Recall: The Role of Long-Term and Working Memory'. *Journal of Memory and Language* 51 (4): 623–43. doi:10.1016/j.jml.2004.07.005.

Kalish, Michael L., Thomas L. Griffiths, and Stephan Lewandowsky. 2007. 'Iterated Learning: Intergenerational Knowledge Transmission Reveals Inductive Biases'. *Psychonomic Bulletin & Review* 14 (2): 288–94. doi:10.3758/BF03194066.

Kashima, Yoshihisa. 2000a. 'Maintaining Cultural Stereotypes in the Serial Reproduction of Narratives'. *Personality and Social Psychology Bulletin* 26 (5): 594–604. doi:10.1177/0146167200267007.

———. 2000b. 'Recovering Bartlett's Social Psychology of Cultural Dynamics'. *European Journal of Social Psychology* 30 (3): 383–403. doi:10.1002/(SICI)1099-0992(200005/06)30:3<383::AID-EJSP996>3.0.CO;2-C.

Kirby, Simon, Hannah Cornish, and Kenny Smith. 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): 10681–6. doi:10.1073/pnas.0707835105.

Kirby, Simon, Mike Dowman, and Thomas L. Griffiths. 2007. 'Innateness and Culture in the Evolution of Language'. *Proceedings of the National Academy of Sciences* 104 (12): 5241–5. doi:10.1073/pnas.0608222104.

Kirby, Simon, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. 'Compression and Communication in the Cultural Evolution of Linguistic Structure'. *Cognition* 141 (August): 87–102.

doi:10.1016/j.cognition.2015.03.016.

Kroeber, A. L. 1952. *The Nature of Culture*. Chicago, IL: University of Chicago Press.

Kuper, Adam. 2000. 'If Memes Are the Answer, What Is the Question?' In *Darwinizing Culture: The Status of Memetics as a Science*, edited by Robert Aunger, 175–88. Oxford, NY: Oxford University Press.

Kuperman, Victor, Hans Stadthagen-Gonzalez, and Marc Brysbaert. 2012. 'Age-of-Acquisition Ratings for 30,000 English Words'. *Behavior Research Methods* 44 (4): 978–90. doi:10.3758/s13428-012-0210-4.

Laland, Kevin, Tobias Uller, Marc Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, et al. 2014. 'Does Evolutionary Theory Need a Rethink?' *Nature News* 514 (7521): 161. doi:10.1038/514161a.

Leskovec, Jure, and Andrej Krevl. 2014. 'SNAP Datasets: Stanford Large Network Dataset Collection', June. http://snap.stanford.edu/data.

Leskovec, Jure, and Rok Sosič. 2016. 'SNAP: A General-Purpose Network Analysis and Graph-Mining Library'. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8 (1): 1.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 2009. 'Meme-Tracking and the Dynamics of the News Cycle'. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, edited by John Elder, Françoise Fogelman-Soulié, Peter A. Flach, and Mohammed J. Zaki, 497–506. New York, NY: ACM. doi:10.1145/1557019.1557077.

Lewens, Tim. 2012. 'Cultural Evolution: Integration and Skepticism'. In *The Oxford Handbook of Philosophy of Social Science*, edited by Harold Kincaid, 458–80. Oxford; New York: Oxford University Press.

Lewontin, R. C. 1982. 'Cultural Transmission and Evolution: A Quantitative Approach'. *American Journal of Human Genetics* 34 (5): 831–32. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1685425/.

Liben-Nowell, David, and Jon Kleinberg. 2008. 'Tracing Information Flow on a Global Scale Using Internet Chain-Letter Data'. *Proceedings of the National Academy of Sciences* 105 (12): 4633–8. doi:10.1073/pnas.0708471105.

Lombardi, Linda, and Mary C Potter. 1992. 'The Regeneration of Syntax in Short Term Memory'. *Journal of Memory and Language* 31 (6): 713–33. doi:10.1016/0749-596X(92)90036-W.

MacCallum, Robert M., Matthias Mauch, Austin Burt, and Armand M. Leroi. 2012. 'Evolution of Music by Public Choice'. *Proceedings of the National Academy of Sciences* 109 (30): 12081–6. doi:10.1073/pnas.1203182109.

Mandler, George, George O. Goodman, and Deanna L. Wilkes-Gibbs. 1982. 'The Word-Frequency Paradox in Recognition'. *Memory & Cognition* 10 (1): 33–42. doi:10.3758/BF03197623.

Marian, Viorica, James Bartolotti, Sarah Chabal, and Anthony Shook. 2012. 'CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities'. *PLOS ONE* 7 (8): e43230. doi:10.1371/journal.pone.0043230.

Mauss, Marcel. 1936. 'Les Techniques Du Corps'. *Journal de Psychologie* 32 (3-4): 271–93. http://classiques.uqac.ca/classiques/mauss_marcel/socio_et_anthropo/6_Techniques_corps/Techniques_corps.html.

Maxwell, R. S. 1936. 'Remembering in Different Social Groups'. *British Journal of Psychology. General*

Section 27 (1): 30–40. doi:10.1111/j.2044-8295.1936.tb00814.x.

McGraw, John J., Sebastian Wallot, Panagiotis Mitkidis, and Andreas Roepstorff. 2014. 'Culture's Building Blocks: Investigating Cultural Evolution in a LEGO Construction Task'. *Frontiers in Psychology* 5 (September). doi:10.3389/fpsyg.2014.01017.

McKinney, Wes. 2010. 'Data Structures for Statistical Computing in Python'. In *Proceedings of the 9th Python in Science Conference (SciPy 2009)*, edited by Stéfan van der Walt and Jarrod Millman, 445:51–56. Pasadena, CA. http://conference.scipy.org/proceedings/scipy2010/mckinney.html.

Menary, Richard. 2010. *The Extended Mind*. Cambridge, MA, USA: MIT Press. http://public.eblib.com/choice/publicfullrecord.aspx?p=3339152.

Mercier, Hugo, and Dan Sperber. 2011. 'Why Do Humans Reason? Arguments for an Argumentative Theory'. *Behavioral and Brain Sciences* 34 (2): 57–74. doi:10.1017/S0140525X10000968.

Mesoudi, Alex, and Andrew Whiten. 2004. 'The Hierarchical Transformation of Event Knowledge in Human Cultural Transmission'. *Journal of Cognition and Culture* 4 (1): 1–24. doi:10.1163/156853704323074732.

———. 2008. 'The Multiple Roles of Cultural Transmission Experiments in Understanding Human Cultural Evolution'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363 (1509): 3489–3501. doi:10.1098/rstb.2008.0129.

Mesoudi, Alex, Andrew Whiten, and Robin Dunbar. 2006. 'A Bias for Social Information in Human Cultural Transmission'. *British Journal of Psychology* 97 (3): 405–23. doi:10.1348/000712605X85871.

Mesoudi, Alex, Andrew Whiten, and Kevin N. Laland. 2006. 'Towards a Unified Science of Cultural Evolution'. *Behavioral and Brain Sciences* 29 (04): 329–47. doi:10.1017/S0140525X06009083.

———. 2007. 'Science, Evolution, and Cultural Anthropology. A Response to Ingold (This Issue)'. *Anthropology Today* 23 (2): 18–18. doi:10.1111/j.1467-8322.2007.00498.x.

Millman, K. Jarrod, and Michael Aivazis. 2011. 'Python for Scientists and Engineers'. *Computing in Science & Engineering* 13 (2): 9–12. doi:10.1109/MCSE.2011.36.

Mitkidis, Panagiotis, John J. McGraw, Andreas Roepstorff, and Sebastian Wallot. 2015. 'Building Trust: Heart Rate Synchrony and Arousal During Joint Action Increased by Public Goods Game'. *Physiology & Behavior* 149 (October): 101–6. doi:10.1016/j.physbeh.2015.05.033.

Miton, Helena, Nicolas Claidière, and Hugo Mercier. 2015. 'Universal Cognitive Mechanisms Explain the Cultural Success of Bloodletting'. *Evolution and Human Behavior* 36 (4): 303–12. doi:10.1016/j.evolhumbehav.2015.01.003.

Morin, Olivier. 2013. 'How Portraits Turned Their Eyes Upon Us: Visual Preferences and Demographic Change in Cultural Evolution'. *Evolution and Human Behavior* 34 (3): 222–29. doi:10.1016/j.evolhumbehav.2013.01.004.

———. 2016. *How Traditions Live and Die*. New York: Oxford University Press. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1074945.

Moritz, Maria, Andreas Wiederhold, Barbara Pavlek, Yuri Bizzoni, and Marco Büchler. 2016. 'Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and Its Application to Bible Reuse'. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language*

Processing, edited by Jian Su, Xavier Carreras, and Kevin Duh, 1849–59. Austin, TX: Association for Computational Linguistics.

Morrison, Catriona M., and Andrew W. Ellis. 1995. 'Roles of Word Frequency and Age of Acquisition in Word Naming and Lexical Decision'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (1): 116–33. doi:10.1037/0278-7393.21.1.116.

Moussaïd, Mehdi, Henry Brighton, and Wolfgang Gaissmaier. 2015. 'The Amplification of Risk in Experimental Diffusion Chains'. *Proceedings of the National Academy of Sciences* 112 (18): 5631–6. doi:10.1073/pnas.1421883112.

Nelson, Douglas L., Kirsty Kitto, David Galea, Cathy L. McEvoy, and Peter D. Bruza. 2013. 'How Activation, Entanglement, and Searching a Semantic Network Contribute to Event Memory'. *Memory & Cognition* 41 (6): 797–819. doi:10.3758/s13421-013-0312-y.

Nelson, Douglas L., Cathy L. McEvoy, and Thomas A. Schreiber. 2004. 'The University of South Florida Free Association, Rhyme, and Word Fragment Norms'. *Behavior Research Methods, Instruments, & Computers* 36 (3): 402–7. doi:10.3758/BF03195588.

Nickels, Lyndsey, and David Howard. 2004. 'Dissociating Effects of Number of Phonemes, Number of Syllables, and Syllabic Complexity on Word Production in Aphasia: It's the Number of Phonemes That Counts'. *Cognitive Neuropsychology* 21 (1): 57–78. doi:10.1080/02643290342000122.

Norenzayan, Ara, Scott Atran, Jason Faulkner, and Mark Schaller. 2006. 'Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives'. *Cognitive Science* 30 (3): 531–53. doi:10.1207/s15516709cog0000_68.

Northway, Mary L. 1936. 'The Influence of Age and Social Group on Children's Remembering'. *British Journal of Psychology. General Section* 27 (1): 11–29. doi:10.1111/j.2044-8295.1936.tb00813.x.

O'Brien, Michael J., and Kevin N. Laland. 2012. 'Genes, Culture, and Agriculture: An Example of Human Niche Construction'. *Current Anthropology* 53 (4): 434–70. doi:10.1086/666585.

Odling-Smee, F. John, Kevin N Laland, and Marcus W Feldman. 2003. *Niche Construction: The Neglected Process in Evolution*. Princeton: Princeton University Press.

Omodei, Elisa, Thierry Poibeau, and Jean-Philippe Cointet. 2012. 'Multi-Level Modeling of Quotation Families Morphogenesis'. In *Proceedings of the 4th ASE/IEEE International Conference on Social Computing*, edited by Anton Nijholt, Alessandro Vinciarelli, and Dirk Heylen, 392–401. Washington, DC: IEEE Computer Society. http://arxiv.org/abs/1209.4277.

Oyama, Susan. (1985) 2000. *The Ontogeny of Information: Developmental Systems and Evolution*. Durham, NC: Duke University Press.

Oyama, Susan, Paul Griffiths, and Russell D Gray, eds. 2001. *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge, Mass.: MIT Press.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. 'The PageRank Citation Ranking: Bringing Order to the Web'. Stanford InfoLab.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. 'Scikit-Learn: Machine Learning in Python'. *Journal of Machine Learning Research* 12 (Oct): 2825–30. http://www.jmlr.org/papers/v12/pedregosa11a.html.

Pérez, Fernando, and Brian E. Granger. 2007. 'IPython: A System for Interactive Scientific Comput-

ing'. *Computing in Science & Engineering* 9 (3): 21–29. doi:10.1109/MCSE.2007.53.

Perfors, Amy, and Daniel J. Navarro. 2014. 'Language Evolution Can Be Shaped by the Structure of the World'. *Cognitive Science* 38 (4): 775–93. doi:10.1111/cogs.12102.

Potter, Mary C, and Linda Lombardi. 1990. 'Regeneration in the Short-Term Recall of Sentences'. *Journal of Memory and Language* 29 (6): 633–54. doi:10.1016/0749-596X(90)90042-X.

Potter, Mary C., and Linda Lombardi. 1998. 'Syntactic Priming in Immediate Recall of Sentences'. *Journal of Memory and Language* 38 (3): 265–82. doi:10.1006/jmla.1997.2546.

Purzycki, Benjamin Grant, and Aiyana K. Willard. 2016. 'MCI Theory: A Critical Discussion'. *Religion, Brain & Behavior* 6 (3): 207–48. doi:10.1080/2153599X.2015.1024915.

Reali, Florencia, and Thomas L. Griffiths. 2009. 'The Evolution of Frequency Distributions: Relating Regularization to Inductive Biases Through Iterated Learning'. *Cognition* 111 (3): 317–28. doi:10.1016/j.cognition.2009.02.012.

Rey, Arnaud, Arthur M Jacobs, Florian Schmidt-Weigand, and Johannes C Ziegler. 1998. 'A Phoneme Effect in Visual Word Recognition'. *Cognition* 68 (3): B71–B80. doi:10.1016/S0010-0277(98)00051-1.

Risjord, Mark. 2012. 'Models of Culture'. In *The Oxford Handbook of Philosophy of Social Science*, edited by Harold Kincaid, 387–408. Oxford; New York: Oxford University Press.

Roberts, Gareth, and Bruno Galantucci. 2017. 'Investigating Meaning in Experimental Semiotics'. *Psychology of Language and Communication* 20 (2): 130–53. doi:10.1515/plc-2016-0008.

Roediger, Henry L., and Kathleen B. McDermott. 1995. 'Creating False Memories: Remembering Words Not Presented in Lists'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (4): 803–14. doi:10.1037/0278-7393.21.4.803.

Rogers, Everett M. (1962) 2005. *Diffusion of Innovations*. New York, NY: Free Press.

Ruan, Zhongyuan, Gerardo Iñiguez, Márton Karsai, and János Kertész. 2015. 'Kinetics of Social Contagion'. *Physical Review Letters* 115 (21): 218702. doi:10.1103/PhysRevLett.115.218702.

Schmid, Helmut. 1994. 'Probabilistic Part-of-Speech Tagging Using Decision Trees'. In *Proceedings of International Conference on New Methods in Language Processing*, edited by Daniel B. Jones, 12:44–49. Manchester, UK: UMIST.

Scott-Phillips, Thomas C. 2017. 'Pragmatics and the Aims of Language Evolution'. *Psychonomic Bulletin & Review* 24 (1): 186–89. doi:10.3758/s13423-016-1061-2.

Scott-Phillips, Thomas C., and Simon Kirby. 2010. 'Language Evolution in the Laboratory'. *Trends in Cognitive Sciences* 14 (9): 411–17. doi:10.1016/j.tics.2010.06.006.

Scott-Phillips, Thomas C., Simon Kirby, and Graham R. S. Ritchie. 2009. 'Signalling Signalhood and the Emergence of Communication'. *Cognition* 113 (2): 226–33. doi:10.1016/j.cognition.2009.08.009.

Scott-Phillips, Thomas C., Kevin N. Laland, David M. Shuker, Thomas E. Dickins, and Stuart A. West. 2014. 'The Niche Construction Perspective: A Critical Appraisal'. *Evolution* 68 (5): 1231–43. doi:10.1111/evo.12332.

Silvey, Catriona, Simon Kirby, and Kenny Smith. 2015. 'Word Meanings Evolve to Selectively Preserve Distinctions on Salient Dimensions'. *Cognitive Science* 39 (1): 212–26. doi:10.1111/cogs.12150.

Simmons, Matthew P., Lada A. Adamic, and Eytan Adar. 2011. 'Memes Online: Extracted, Subtrac-

ted, Injected, and Recollected'. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, edited by Nicolas Nicolov, James G. Shanahan, Lada A. Adamic, Ricardo Baeza-Yates, and Scott Counts, 353–60. Menlo Park, CA: The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2836.

Slingerland, Edward G. 2008. *What Science Offers the Humanities: Integrating Body and Culture*. Cambridge; New York: Cambridge University Press.

Smith, Kenny, and Elizabeth Wonnacott. 2010. 'Eliminating Unpredictable Variation Through Iterated Learning'. *Cognition* 116 (3): 444–49. doi:10.1016/j.cognition.2010.06.004.

Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford, UK; Cambridge, Mass.: Blackwell.

Sperber, Dan, and Deirdre Wilson. (1986) 1995. *Relevance: Communication and Cognition*. Oxford UK & Cambridge USA: Blackwell.

Sterelny, Kim. 2001. 'Niche Construction, Developmental Systems, and the Extended Replicator'. In *Cycles of Contingency: Developmental Systems and Evolution*, edited by Susan Oyama, Paul Griffiths, and Russell D Gray, 333–49. Cambridge, Mass.: MIT Press.

———. 2010. 'Minds: Extended or Scaffolded?' *Phenomenology and the Cognitive Sciences* 9 (4): 465–81. doi:10.1007/s11097-010-9174-y.

———. 2012. *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA, USA; London, UK: The MIT Press.

———. 2017. 'Cultural Evolution in California and Paris'. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 62 (April): 42–50. doi:10.1016/j.shpsc.2016.12.005.

Stotz, Karola. 2010. 'Human Nature and Cognitive–developmental Niche Construction'. *Phenomenology and the Cognitive Sciences* 9 (4): 483–501. doi:10.1007/s11097-010-9178-7.

Tamariz, Monica, and Simon Kirby. 2016. 'The Cultural Evolution of Language'. *Current Opinion in Psychology* 8 (April): 37–43. doi:10.1016/j.copsyc.2015.09.003.

Tamariz, Mónica, and Simon Kirby. 2015. 'Culture: Copying, Compression, and Conventionality'. *Cognitive Science* 39 (1): 171–83. doi:10.1111/cogs.12144.

Tamariz, Monica, T. Mark Ellison, Dale J. Barr, and Nicolas Fay. 2014. 'Cultural Selection Drives the Evolution of Human Communication Systems'. *Proceedings of the Royal Society of London B: Biological Sciences* 281 (1788): 20140488. doi:10.1098/rspb.2014.0488.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Tulving, Endel. 1962. 'Subjective Organization in Free Recall of "Unrelated" Words'. *Psychological Review* 69 (4): 344–54. doi:10.1037/h0043150.

———. 1966. 'Subjective Organization and Effects of Repetition in Multi-Trial Free-Recall Learning'. *Journal of Verbal Learning and Verbal Behavior* 5 (2): 193–97. doi:10.1016/S0022-5371(66)80016-6.

van der Walt, Stéfan, S. Chris Colbert, and Gaël Varoquaux. 2011. 'The NumPy Array: A Structure for Efficient Numerical Computation'. *Computing in Science & Engineering* 13 (2): 22–30.

doi:10.1109/MCSE.2011.37.

van Gelder, Tim. 1998. 'The Dynamical Hypothesis in Cognitive Science'. *Behavioral and Brain Sciences* 21 (5): 615–28. https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/the-dynamical-hypothesis-in-cognitive-science/C121F1B65A534F3E7A27075EE489AD1E.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Verhoef, Tessa, Simon Kirby, and Bart de Boer. 2014. 'Emergence of Combinatorial Structure and Economy Through Iterated Learning with Continuous Acoustic Signals'. *Journal of Phonetics* 43 (March): 57–68. doi:10.1016/j.wocn.2014.02.005.

Wallot, Sebastian, Panagiotis Mitkidis, John J. McGraw, and Andreas Roepstorff. 2016. 'Beyond Synchrony: Joint Action in a Complex Production Task Reveals Beneficial Effects of Decreased Interpersonal Synchrony'. *PLOS ONE* 11 (12): e0168306. doi:10.1371/journal.pone.0168306.

Ward, T. H. G. 1949. 'An Experiment on Serial Reproduction with Special Reference to the Changes in the Design of Early Coin Types'. *British Journal of Psychology. General Section* 39 (3): 142–47. doi:10.1111/j.2044-8295.1949.tb00213.x.

Watts, Duncan J. 2002. 'A Simple Model of Global Cascades on Random Networks'. *Proceedings of the National Academy of Sciences* 99 (9): 5766–71. doi:10.1073/pnas.082090499.

Watts, Duncan J., and Steven H. Strogatz. 1998. 'Collective Dynamics of "Small-World" Networks'. *Nature* 393 (6684): 440–42. doi:10.1038/30918.

Weide, Robert. 1998. 'The CMU Pronouncing Dictionary'. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Weng, L., A. Flammini, A. Vespignani, and F. Menczer. 2012. 'Competition Among Memes in a World with Limited Attention'. *Scientific Reports* 2 (March). doi:10.1038/srep00335.

Weng, Lilian, Márton Karsai, Nicola Perra, Filippo Menczer, and Alessandro Flammini. 2015. 'Attention on Weak Ties in Social and Communication Networks', May. http://arxiv.org/abs/1505.02399.

Whiten, Andrew, Christine A Caldwell, and Alex Mesoudi. 2016. 'Cultural Diffusion in Humans and Other Animals'. *Current Opinion in Psychology*, Culture, 8 (April): 15–21. doi:10.1016/j.copsyc.2015.09.002.

Wilson, Deirdre, and Dan Sperber. 2004. 'Relevance Theory'. In *The Handbook of Pragmatics*, edited by Laurence R Horn and Gregory Ward, 607–32. Oxford, UK: Blackwell.

Wimsatt, William C., and James R. Griesemer. 2007. 'Reproducing Entrenchments to Scaffold Culture: The Central Role of Development in Cultural Evolution'. In *Integrating Evolution and Development: From Theory to Practice*, edited by Roger Sansom and Robert N Brandon, 227–323. Cambridge, MA, USA: MIT Press.

Winters, James, Simon Kirby, and Kenny Smith. 2015. 'Languages Adapt to Their Contextual Niche'. *Language and Cognition* 7 (3): 415–49. doi:10.1017/langcog.2014.35.

WordNet. 2010. 'Princeton University "About WordNet"'. https://wordnet.princeton.edu/wordnet/.

Xu, Jing, Mike Dowman, and Thomas L. Griffiths. 2013. 'Cultural Transmission Results in Convergence Towards Colour Term Universals'. *Proc. R. Soc. B* 280 (1758): 20123073.

doi:10.1098/rspb.2012.3073.

Yonelinas, Andrew P. 2002. 'The Nature of Recollection and Familiarity: A Review of 30 Years of Research'. *Journal of Memory and Language* 46 (3): 441–517. doi:10.1006/jmla.2002.2864.

Zaromb, Franklin M., Marc W. Howard, Emily D. Dolan, Yevgeniy B. Sirotin, Michele Tully, Arthur Wingfield, and Michael J. Kahana. 2006. 'Temporal Associations and Prior-List Intrusions in Free Recall'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32 (4): 792–804. doi:10.1037/0278-7393.32.4.792.

Zevin, Jason D, and Mark S Seidenberg. 2002. 'Age of Acquisition Effects in Word Reading and Other Tasks'. *Journal of Memory and Language* 47 (1): 1–29. doi:10.1006/jmla.2001.2834.