EXPLORATIONS IN CULTURAL EVOLUTION:

METHODOLOGICAL CHALLENGES AND CASE STUDIES

By

Helena Miton

Submitted to

Central European University

Department of Cognitive Science

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Science

Supervisors: Christophe Heintz & Dan Sperber

Advisor: Olivier Morin

Budapest, Hungary

2019

Declaration of Authorship

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

Helena Miton

The present thesis includes work that appears in the following publications:

Miton, H., & Charbonneau, M. (2018). Cumulative culture in the laboratory: Methodological and theoretical challenges. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20180677.

Miton, H., & Morin, O. (2019). When iconicity stands in the way of abbreviation: No Zipfian effect for figurative signals. *PloS one*, *14*(8), e0220793.

Other parts of the present thesis shall be submitted for publication with the following co-authors:

- Chapter 3 with Hans-Jörg Bibiko and Olivier Morin
- Chapter 4 with Mikolaj Hernik and Dan Sperber
- Chapter 5 with Thomas Wolf, Cordula Vesper, Gunther Knoblich and Dan Sperber.

Abstract

This thesis aims at developing a framework for empirical research in cultural evolution, drawing on cultural attraction theory. This framework is outlined in the Introduction. The five chapters of the thesis demonstrate the robustness of this framework across different cultural domains and diverse types of causal factors relevant to explaining the emergence, success, and evolution of cultural types.

Chapter 1 reviews the use of cultural transmission experiments (transmission chains, replacement, closed groups and seeded groups) in studying cumulative cultural evolution. Cumulative cultural evolution is usually defined as the process by which traditions are gradually modified. This chapter identifies several mismatches between theoretical definitions of cumulative culture and their implementation in cultural transmission experiments, and suggests possible solutions to reduce these mismatches.

Chapter 2 documents an exception to Zipf's law of abbreviation (which relates more frequent signals to shorter signal lengths) by observing two large corpus of European heraldic motifs (total N = 25115). Our results suggest that lacking –or at least losing- iconicity may be a precondition for Zipf's Law of Abbreviation to obtain in a graphic code.

Chapter 3 tests hypotheses on possible determinants of visual complexity in characters, using a standardized collection of 47,880 pictures from 133 writing systems, and two measures of visual complexity (algorithmic and perimetric). This chapter provides evidence that (1) the size of a script's inventory influences character complexity, (2) one of the main determinant of character complexity is the script's type (e.g., alphabetic, syllabic), and (3) there is a surprising lack of evolutionary change in character complexity.

Chapter 4 provides evidence of the existence of a forward bias in human profile-oriented portraits: there is a widespread tendency (total N = 1833, from 582 unique painters) to represent sitters with more space in front of them than behind them. It also suggests that this bias became more frequently and more strongly expressed over time.

Chapter 5 shows that different physical affordances can influence the rhythms naïve participants produce in a transmission chain experiment. Rhythmical sequences produced by participants having to adapt to use different movements reflected such constrains in both their structure and timing.

Two shorter introductions to chapter 2 and 3 and to chapter 4 and 5 outline the commonalities between the two chapters they each introduce. The conclusion revisits the question raised and the framework outlined in the introduction in the light of the five chapters of the thesis.

Acknowledgments

This PhD has been a wild ride, but also a fantastic one. I am reaching the end of this adventure feeling grateful for both the experience and the company I shared it with.

First and foremost, my endless gratitude goes to Dan, for being not only an amazing supervisor and collaborator, but also an all-around-the-clock mentor. I am grateful to all three of my supervisors for their wisdom, support and patience, Dan, Olivier and Christophe. It has been an honour being your student. Hugo, although not a supervisor of my thesis, deserves credit for introducing me to cultural evolution a few years ago, and remaining a great collaborator ever since.

I have had two academic homes during my thesis - the OSCAR group in Budapest, and the Mint (Minds and Traditions) in Jena. The Minds and Traditions research group (Mint) has been my research team at heart, and would you like to address special thanks to Barbara, James, Piers, Oleg and Thomas. From the Oscar group, I would especially like to thank Mathieu, for your never-ending patience and enthusiasm while working with me, and teaching me equally well about Quebec's idioms and theoretical precision. For the projects I worked on with in Budapest, Thomas, Mikolaj and Cordula have been particularly outstanding collaborators.

The Cognitive Science department has also been a friendly environment thanks to Oana, Thomas, Nazli, Neha, Ieva, Thom, Laura, Cameron, Pavel, all my cohort, and everyone I've shared a good moment with.

For helping me with data collection of the various chapters, I am forever indebted to Anna, Robi, David, Liliana, Giang, Olena, Noro and Julia. To Reka, Eszter & Andi, thank you so much for your help with all administrative aspects.

Thanks for scientific conversations and useful criticisms of my work, and invitations to develop my own views over the years to Moshe, Radu, Vanessa, and Santa Fe institute folks.

Now, for more personal aspects - the last few years have been a bumpy road, and I feel incredibly lucky in that I never felt lonely along it. For making home still feel like home after having left for years, Mathilde & Vincent deserve some special mention. And so do my parents!

For all the laughter, encouragements, good times and support, I'd like to thank, Manu, Anthony, Adeline, Alex, Léo and Tazzio. Very special thanks to Matyas, who has never missed a call since we met.

Jordan, thank you for having tried your best to support me and be here for me. And also for all the drives to and from the airport, no matter how early or late my flights were.

Finally, to Mia: Thanks for… Everything. Your friendship and support has always been and is overwhelming enough to leave me wordless. You are the one person who has been here through this adventure each and every step of the way, from the very beginning to the very end, in the good days and even more so in the bad days. I will never be able to thank you enough for all you brought into my life. I can't imagine any of this thesis would even exist if it wasn't for you being the most amazing partner-in-crime.

One last word, to my sister, Cecilia – trying to make you proud is and has always been an endless source of motivation.

Table of contents

Introduction

This thesis, which aims to contribute to the study of cultural evolution, can be read in two ways. The first is to read each chapter in relative isolation of the others, for its contribution to the question it delimitates. The second possible reading is to understand all chapters as illustrations of a particular methodology and as evidence that this methodology can be useful for empirical research. This introduction outlines the framework used in all four case studies included in this thesis, which largely build on cultural attraction theory.

In this introduction, (1) I briefly illustrate the heterogeneity of cultural things and consider what it means for the study culture and how the issue has been tackled in the field of cultural evolution; (2) I present a few key concepts used in the framework used in the thesis; (3) I detail some of the assumptions of this framework, (4) I explain how this framework may guide empirical research, and finally, (6) I outline the thesis' chapters.

1. Studying culture is a problem of Causal Inference

Some examples of cultural things

Let us start with three observations (see Figure 1):

- Between 1900 and 1980, Teddy bears' faces evolved. The ratio of the vertical distance between their eyes and their crown to the distance between their eyes and the base of their head increased, and the ratio of the distance between the tip of their snout and the back of their head to the distance between the top of their head and its base decreased (Gould, 1979; Hinde & Barden, 1985).

- An enigmatic picture of a woman, sometimes along with patches of colours or shades of grey, would appear for a few frames in the reel leader of 15 or 36mm films. Used for calibration purposes, these so-called « China girls » have been a feature of films reels familiar to technicians for decades, from the 1920s to the 1990s. Later, they even played a central part in some structural films (e.g., Morgan Fisher's *Standard Gauge*), an experimental film movement which emerged in the 1960s—see (Yue, 2015).

- Baku dream-eater demons, from the Chinese and Japanese folklore, are chimeras made of leftover pieces after gods created all the other animals. They almost always include, at least, elephant and tiger body parts, and are said to devour nightmares. They have been part of the Japanese folklore since at least the 15th century (Hori, 2005). A modern-day variant of this mythological creature can be found in the Pokemon Drowzee, which is, in line with its dream-eater ancestry, associated to hypnotic capacities.

Figure 1. Examples of three cultural types: from left to right, a teddy bear from the early 1900s, the LAD Kodak China girl on a film, and a Baku dream-eater (in its traditional version, as painted by Hokusai).

Teddy bear by the Smithsonian Museum of Natural History - https://www.flickr.com/photos/23165290@N00/7237653442/ , CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=41208714, Kodak LAD girl by Rosa Menkman, https://www.flickr.com/photos/r00s/24247343538, and Baku dream-eater by Hokusai, Public Domain, https://commons.wikimedia.org/w/index.php?curid=52867864

We can have a look at what explains these three examples:

- Teddy bears' traits became more *neotenic*, i.e., more similar to traits of infants and toddlers than to traits of adults (Gould, 1979; Hinde & Barden, 1985). Acquiring and preserving this feature made Teddy bears look particularly cute and secured their cultural success.
- China girls were pictures used for quality control: they allowed technicians to notice any major problem with the colours of a new copy of the film. Although the origin of the name 'China girl' is uncertain, it might have been due to the first China girls having been pictures of Chinese dolls. One particularly famous instance being the Kodak LAD (for Laboratory Aim Density) girl, for which John Pytlak received an Academy Award. The cultural success

of China girls was based in their particular aptness for colour calibration of film because, it was assumed, unwanted colour variations would be particularly likely to be noticed in pictures of women, which attracted attention anyhow (Yue, 2015).

- Finally, Baku dream-eaters join traits that make them easy to remember: they are monsters combining body parts from different animals (Sperber, 1996b).They also have supernatural abilities: they can devour dreams and nightmares. Such supernatural abilities—i.e., eating something immaterial—are counterintuitive, which provides them with a mnemonic advantage (Boyer, 2007; Harmon-Vukić, Upal, & Sheehan, 2012; Johnson, Bishop, & Kelly, 2010; Norenzayan, Atran, Faulkner, & Schaller, 2006).

Why this triplet of trivia? All three are cultural phenomena. They fall unproblematically under several common definitions of what is cultural. The occurrences of these three types are relatively spread in time and space, which echoes Olivier Morin's definition ("Culture, as defined here, consists in stable traditions that travel far, thanks to cultural transmission." (Morin, 2016, p. 12). Culture has also been defined as « information capable of affecting individuals' behaviour that they acquire from other members of their species through teaching, imitation, and other forms of social transmission » (Richerson & Boyd, 2005, p. 5)—all the examples above affect individuals' behaviour and are socially transmitted, at least to some extent. Another definition of culture is « anything that replicates by passing through cognition » (Ferdinand, 2015, p. 5). All three examples are transmitted through cognitive systems. All the same, they are quite diverse, and so are the factors involved in their evolution.

Culture has two main aspects: social transmission (or propagation) and stability. To be cultural, an item must (a) propagate in a population through many episodes of transmission, and (b) be stable—that is, keeps its characteristic features—at a populational level. Given the diversity of the kind of items that can be found in culture, and also of the causes of their cultural success, what does it mean to explain culture? What aspects of culture do we want to explain? Which types of causal inferences can we do?

Studying culture

A framework for studying culture should ideally account for the diversity of both cultural types and causal factors. Studying culture requires a flexible framework. We should also ask, what would be a *good* explanation for culture? It should, I suggest, help explain in some detail characteristic features of the cultural contents we are studying. Explaining culture should help explain both cultural success (spread in space and time) and the non-randomness of cultural types (their specific characteristics).

Cultural evolution is an interdisciplinary field of study that emerged on the assumption that studying culture could best be done by using an evolutionary approach. The field seems to be unified by the assumption that cultural phenomena can be studied using Darwinian principles (Godfrey-Smith, 2012; Mesoudi, 2011b, 2017; Mesoudi, Whiten, & Laland, 2006). Citing the Cultural Evolution Society's website, 'cultural change constitutes an evolutionary process that shares fundamental similarities with – but also differs in key ways from- genetic evolution' ("Cultural Evolution Society," n.d.). It has also developed specific methodological tools, ranging from mathematical modelling to computer simulations and from the exploitation of large-scale databases to experimental designs mimicking the passing of cultural generations.

Cultural evolution, as a field, has branched out into different frameworks, which in turn have developed different research agendas, and produced different type of case studies. These frameworks are ways to select questions and systems of causal inference relevant to the study of culture. Following Kim Sterelny (Sterelny, 2017) and Cecilia Heyes (Heyes, 2018), and for the sake of brevity, I focus mostly on the contrast between dual inheritance theory (DIT) and cultural attraction theory (CAT), also known respectively as the Californian school and the Paris school. This distinction is a quite common way to present the field of cultural evolution (see Lewens, 2015; Mesoudi, 2016b, 2016a, 2017). The work presented here is closer to the Paris school. These "schools" differ in their research agendas and in what they take to be the crucial questions for cultural evolution. Those differences, in turn, make research programs and their respective contributions rather difficult to compare—see (Sterelny, 2017) for a discussion. Here, I focus on situating the approach used in the chapters of this thesis among the different frameworks in cultural evolution.

One major difference among various frameworks is the degree to which they assume that the human species has evolved specific psychological adaptations for culture and cultural transmission, that is, evolved mechanisms that have the transmission of cultural contents as their biological function. Such mechanisms have been one of the focus of dual inheritance theory (Richerson & Boyd, 2005) and include transmission biases and social learning strategies (Kendal et al., 2018; Muthukrishna & Henrich, 2019). In such approaches, cultural phenomena are best explained by considering which adaptations allow for relatively faithful cultural transmission and for selection among available candidates for transmission.

Conversely, other approaches put less causal weight on the existence of psychological adaptations that evolved specifically for culture. Rather than focusing on such psychological adaptations, some approaches highlight the degree to which culture is 'adapted' to humans. They insist more on the fit between cultural contents and human cognitive processes that didn't necessarily evolve for cultural transmission (e.g., Sperber, 1996a; Sperber & Hirschfeld, 2004). In this perspective, cultural contents should have features that make them easy for human minds to acquire or learn, to store and retrieve from memory, or to produce (with ease of production involving a range of factors beyond psychology, from motor abilities to material affordances).

To the extent that one assumes that what explains human culture is the existence in humans of adaptations for cultural transmission then a main research goal is to establish what those adaptations are, how and under which conditions they conferred an advantage to their bearers. To the extent, on the other hand, that one assumes that human culture has evolved by adapting itself to human minds, this orients the research agenda towards linking specific features of cultural contents to constraints and opportunities presented by the organisation of human minds and by the reiterated interactions of organisms with such minds. Both perspective can be fruitful and, in the end, their contributions should be integrated. Here however, I adopt the second perspective.

How to identify relevant causal factors?

What is considered to require an explanation (explanandum), and what is considered a satisfying explanation (explanans) in the study of cultural evolution? Let's consider, for the sake of illustration, the set of cultural phenomena (teddy bears, China girls and Baku dream-eaters) we started this chapter with. Do they all fall under a general type of explanation or are several types of explanation needed? How many exactly? Should they be clustered and how?
We can assume that it is not possible to get *a priori* the right number of causes to look for. In such a situation, there are two possible starting points, at opposite ends of a continuum. Consider the whole set of phenomena as a relatively homogeneous one, so that all members of the set can ultimately be understood (and predicted) using the same type of explanation (**option 1**); or, consider each of these cultural phenomena as *sui generis*, to the point of requiring its own explanation (**option 2**).

Option 1 (possibly undershooting the number of causes) can be found, for instance, in research strategies focusing on the idea of culture being an adaptation, for the acquisition and transmission of which humans themselves possess dedicated psychological adaptations. Then, most, if not all, of human cultural phenomena can be explained by this set of adaptations (whether they are social learning strategies, transmission biases, or other mechanisms), which emerged in virtue of culture providing fitness-enhancing information in a more cost-effective way than individual learning (Henrich & McElreath, 2003; Laland, 2004; Rendell et al., 2010; Richerson & Boyd, 2005). Cost refers here both to risks encountered in exploring different possible solutions by oneself, and to the resources (including opportunity costs) used in the search. Option 2 (possibly overshooting the number of causes) can be found in frameworks (among them, Cultural Attraction) that focus on the way cultural contents evolve to fit human cognition and cultural transmission mechanisms. In this perspective, it is possible to overshoot the number of causes – i.e., include more causal factors than strictly necessary, or have excessively fine-grained explanations that apply perfectly only to the cultural phenomenon studied—part 5 of this introduction consider ways to avoid this problem.

Both fitting a model to a set of data points and building or choosing a theoretical framework to explain cultural phenomena face the same challenge. Two main aspects of the trade-off between the two options I described make it akin to the tension between over-fitting and under-fitting in statistical modelling: (1) the number of causes used to explain human cultural phenomena, and (2) the possibility (and aspiration) to generalize to new occurrences.

Studying culture is mostly a problem of causal inferences. A framework to study culture should be able to account for diverse cultural phenomena and causal factors and, for this, should be quite flexible. The framework I introduce and use here is an attempt at using a few concepts to build a set of causal inferences, which can, in turn, be put to use in empirical case studies. This approach is only one among many possibilities. It puts to use concepts from cultural attraction theory in empirical research (see next part), and adopts assumptions that may not be shared by the whole field (see part 3).

2. Framework's key concepts

Cultural attraction (as presented in Claidière, Scott-Phillips, & Sperber, 2014; Scott-Phillips, Blancke, & Heintz, 2018; Sperber, 1996a) is centered around three operational concepts: cultural causal chains, cultural attractors, and factors of attraction. I here present these concepts in relation to how they can be used in empirical research, in relation, that is, to what can be observed, predicted, and/or inferred.

Cultural causal chains

Cultural causal chains are defined as « chains of causally related events in which (a) mental representations (beliefs, knowledge, intentions, etc.) cause public productions (speech, artefacts, behaviour, etc.), which in turn cause further mental representations in other individuals, and so forth and (b) some items in the chain exhibit a degree of similarity and thus constitute relatively stable distributions of similar items in the population and its habitat, and across time and space» (Scott-Phillips et al., 2018). The links in such causal chains are transmission mental and behavioural episodes alternate: mental processes in a source lead to an observable behaviour (or products of behaviour) leading to mental processes in a recipient, and so on. The behavioural components in such a chain can be directly observed and provide evidence of the mental processes in the source and in the recipient. Cultural causal chains provide a schematic description of the different steps or events involved in cultural transmission. These chains are a useful tool: charting out these steps in social transmission of cultural contents and asking which systems or means these steps recruit provides a list of potential factors of attraction.

Cultural attractors

Attractors as defined in Cultural Attraction Theory are statistical regularities. Cultural attraction theory highlight the fact that transformations in episodes of transmission, provided that they are non-random, can be a factor of stability. More specifically, in the case of convergent transformations towards some point in the space of possibilities, then cultural items will tend to cluster and stabilize around this point of convergence, which is then called an attractor.

Attractors can be observed at the macro-level of distribution of cultural items in a population over time. Observation of cultural phenomena at a populational scale reveal clusters of close-variants around an attractor. In this perspective, the distribution of cultural items in a population reveals the location of attractors. Historical evidence may reveal change in the location and density of such clusters and hence of the movement and relative force of attractors over time.

This characterization of attractors differs from other formulations that have been proposed. (Scott-Phillips et al., 2018) define a cultural attractor as « A type of item whose frequency is relatively high and stable as an effect of cultural attraction ». (Claidière, Scott-

Phillips, et al., 2014) suggested that « any type whose relative frequency tends to increase over time ». This last formulation is inadequate for empirical research: whether an increase in frequency indicates the presence of an attractor would depend heavily on the time frame chosen to observe a given cultural practice. In particular, if the system is observed at equilibrium, frequencies might not show any more change.

For our purpose, cultural attractors are (often) observable in a cultural environment: i.e., an (evolving) distribution of items. More often than not, if we are able to record a cultural phenomenon, this implies that the phenomenon has already reached some critical level of success: most unlucky (unsuccessful) variants don't stay around long enough or spread enough to be observed. Attractors are thus, for all empirical matters, tentatively identified by looking at distributions and successful variants and assuming that successful variants cluster around attractors. So understood, they are a form of fixation point, in the variation space, around which more tokens cluster: it is a part of the variation space with higher density, relative to the rest of the variation space.

Cultural attractors, are, to some extent, like, perfect storms: events reaching criticality (here, cultural success) through a rare combination of circumstances. In human social lives and interactions, most socially transmitted informational contents never reach a properly cultural level of spread and stability. Becoming cultural is relatively rare. The cultural success of anti-vaccination beliefs exemplify this point: their cultural success is the result of a combination of causal factors, from the universal psychological reluctance to getting a dangerous substance inside one's own body to historically situated social factors such as lack of trust in pharmaceutical companies (Miton & Mercier, 2015).

Factors of attraction

Attractors, as we have described them, are a kind of statistical patterns. They are abstract concepts (like "orbit" or "barycentre" in astronomy) that can be used to describe concrete phenomena rather than concrete phenomena themselves. What explains these statistical patterns are concrete processes and mechanisms that play a causal role and are called "factors of attraction." They have been defined as « factors that probabilistically bias how mental representations cause public productions (and vice versa), and which hence cause cultural attraction to occur. Factors of attraction are thought of as whether inside the mind (cognitive/psychological) or outside of it (ecological)» (Scott-Phillips et al., 2018). Andrew Buskell (2017) mapped what constitutes, in his view, three different definitions of factors of attraction previously used by cultural attraction theorists: reconstructive learning, motivational factors and ecological factors. Far from being *definitions* of factors of attraction, these are three *types* of factors of attraction.

Factors of attraction in a cultural attraction framework are causal factors that help explain patterns in the distribution of cultural items and in particular, cultural success. Here, explaining culture involves elucidating which factors of attraction are involved in cultural attractors' success, including their emergence, their stabilization, and their decline. It also includes understanding how these factors interact. Factors of attraction have observable effects but are not directly observable. They have, most often, to be inferred. In a situation of perfect information (i.e., information on factors of attraction *and* initial distribution), it would be possible to actually infer the distribution of cultural items in a variation space, including clusters around attractors.

Explaining a cultural phenomenon in this framework means, first and foremost, determining which factors of attraction sustain it. In turn, this means that factors of attraction usually provide at least partial answers two questions at the same time: (1) why is there a cluster of practices around a given point in the variation space? And (2) what specific characteristics makes this point in the variation space an attractor? These two questions are not independent from one another: the characteristics of cultural attractors are assumed to be informative as to why cultural practices cluster in their vicinity.

3. Framework's assumptions

I adopt the following assumptions, which I discuss in more details below: (1) In general, a chain of cultural transmission events is best approached as a series of transformations of variable amplitude (including zero-amplitude) than as a series of replications with occasional mutations. (2) Cultural stability is best thought of as an emergent effect and needs to be explained, and (3) not all types of contents that get transmitted have the same probability of becoming cultural. These assumptions are central to cultural attraction but they are not incompatible with other approaches. They also, to some extent, depart from standard definitions of culture, and from other definitions commonly used in cultural evolution: for instance, the first assumption departs from Ferdinand's definition in rejecting the idea that cultural contents replicate.

Transformation as the basic operation of cultural transmission

What is the most basic operation of cultural transmission? The answer to this question has consequences for one's research agenda and in particular for what is considered as requiring an explanation. It is similar to choosing a baseline. Different basic operations have different population-level consequences. Theories that assume high-fidelity transmission securing inheritance as the 'default' operation of cultural transmission have stability as a straightforward populational-level consequence (Enquist, Strimling, Eriksson, Laland, & Sjostrand, 2010; Laland, 2018, p. 152) . They have the problem of establishing that mechanisms of imitation and communication do provide a degree of fidelity that they presuppose in the theory, and this is hard to reconcile with empirical work on imitation (Hurley & Chater, 2005) and communication (Noveck & Sperber, 2004) as it occurs in ordinary interactions. By contrast, Cultural Attraction Theory assumes that events of cultural transmission are best thought of as transformations of greater or lesser amplitude (Claidière & Sperber, 2007; Sperber, 1996a) . In this perspective, faithful copying being seen as a limiting case of zero transformation. What requires specific explanation, i.e., what is not the logical large-scale consequence of the basic operations of transmission—is thus stability. Stability, in other terms, cultural success over time, is a kind of exception to what happens to most information transmitted among human beings, namely decay or transformation beyond recognition. Most transmitted information is quite unstable and has little or no cultural success. Cultural change is often best explained simply as changes in the factors responsible for stability: the availability of oral contraceptive to unmarried women in the US radically altered, for instance, the already evolving equilibrium in men-women relationships and was a major cause of social, economic, and cultural changes in the lives of both women and men, destabilising earlier ideas and practices and stabilising new ones (Goldin, 2006; Goldin & Katz, 2002).

In this perspective, it cannot be the case that high fidelity copying is a necessary and sufficient for securing the level of distribution in space and time that makes an item cultural; some transformations at the micro-level of episodes of transmission must somehow contribute to stability at the population level. Cultural stability is achieved in part because of not all transformations are equally likely. Different transformations have different likelihoods and amplitudes. This can make some specific points in the space of possibilities more or less likely to have variants clustering around them. Such transformations may bring about, at the populational level, some form of stability of content, some degree of robustness to cultural phenomena. Each transformation is a departure from fidelity and from reproduction from one step to another in a chain of transmission. Still, some transformations have a greater probability to converge towards a given content (say, a "canonical" version of a song or of a tale) than to diverge from it. In such a case, this content, though partly lost at one transmission step is likely to be reconstructed through succeeding steps. This phenomenon was already illustrated in Frederic Bartlett (1932) pioneering chain of transmission studies: in the transmission of folktales, there were cases where one participant in the chain made a copying error by forgetting a detail and another participants later in the chain made a copying error that reintroduced the detail in question moving back towards the original version. Such examples have been a source of inspiration for the development of the idea of cultural attraction (Sperber, 1996a). Greater chances of transformations toward specific points (or attractors), in a variation space, can ensure populational stability notwithstanding low transmission fidelity. This means that loss of information, and faulty transmission is *not* necessarily a problem, as stability can in many cases be achieved without the kind of mechanisms of faithful copying that have been assumed to have evolved specifically for cultural transmission. Stability, in this perspective, is an emergent effect, a phenomenon to be explained, and that can be explained, at least in part, in terms of non-random transformation towards attractors.

Cultural stability as an emergent effect

We can understand a stable cultural tradition as consisting in a lineage of highly similar cultural traits throughout multiple episodes of cultural transmission such that two occurrences of a cultural traits separated by multiple transmission episodes are highly similar to one another (Charbonneau, 2018b). Stability is thus observable at populational level, while fidelity refers to the similarity between the input and output of each transmission episode (i.e., fidelity is observable at a micro-level). Cultural attraction theory doesn't deny that faithful transmission occurs and, in some cases, plays an important role, but it does not assume that it is frequent enough across all aspects of culture to provide the basic explanation of cultural stability. Cultural stability can be secured without faithful transmission (which, in any case, would have, to explain stability, to reach a degree of fidelity that is rarely observed; see Claidière & Sperber, 2010), but through transformations, in particular when they are convergent (Acerbi, Charbonneau, Miton, & Scott-Phillips, n.d.; Claidière, Amedon, et al., 2018; Claidière, Smith, Kirby, & Fagot, 2014), and also in other cases (Morin & Miton, 2018). A similar assumption is found, at least implicitly, in iterated learning (transmission chain) paradigms: observed fidelity tends to increase as the fit between contents and participants' priors increases. In some transmission chain experiments, increases in similarity between input and output (also called learnability, and equivalent to a decrease in error rates) occur as the contents that are transmitted between participants. This increase in stability is associated to the contents becoming easier to either learn or produce, as they match participants' priors (Kalish, Griffiths, & Lewandowsky, 2007; Xu, Dowman, & Griffiths, 2013).

If fidelity at each transmission episode is *not* assumed to explain stability, fidelity can be re-purposed: from a quasi-postulate, it becomes a useful variable to track. Variation in fidelity can productively be used to test for underlying processes (Acerbi et al., n.d.). Another illustration of how variation in fidelity can actually be informative can be found in the "Shuffling model" used by (Morin & Miton, 2018) to quantify the impact of high-fidelity transmission over the diffusion of heraldic designs. Heraldic designs (coats of arms) have a combinatorial structure which binds a motif (i.e., a symbol or shape) and colours (called tinctures). In this model, deviations from the model's predictions are informative: they indicate that some designs (combinations of elements – motif and tinctures) are copied with higher fidelity (as whole combinations) than most designs (which are recombined element by element, picking two colours and a motif separately).

Not all contents are born equal

Contents are not equipotential: they don't have the same initial likelihood of being transmitted at all, nor of being transmitted with little or no modifications. Equipotentiality is relatively rare. Most processes involved in cultural transmission are likely to be much more effective on some contents that on others (even controlling for complexity). Processes of information transmission shape the contents they transmit. The means used to transmit a content, including the means for reproduction, can drastically change how faithfully the transmission can be: whenever scribbles can be directly copied rather than first remembered and then reproduced, they are more stable over repeated transmission, and can be more complex (Scott-Phillips, 2017; Tamariz & Kirby, 2015). In other words, which contents are stable depends on how they are transmitted: complex scribbles were stable only when directly copied from an example and would quickly transform into simpler forms when reproduced from memory, while simple scribbles were stable when copied, either from memory or directly. Oral transmission also brings about some specific characteristics: contents transmitted orally depend to a large extent on human mnemonic abilities (Rubin, 1995). In short, oral traditions enjoy high-stability just like their written counterparts, but, in most cases, it comes at the expense of some of their characteristics: their orality constraints the contents they can have. In the absence of institutions favouring the conservation of relatively hard to memorize contents, inherently high memorability is a condition of cultural success as exemplified by folktales and urban legends for instance (Bartlett, 1932; Stubbersfield, Flynn, & Tehrani, 2017). Some contents features, for instance, rhymes and other prosodic feature can also favour memorisation, as illustrated in the case of children's oral traditions (Morin, 2016; Rubin, 1995). In other words, specific mechanisms of cultural transmission may favour or hinder specific contents (they are content-sensitive).

One relevant aspect to take into account is the degree to which human minds contains domain-specific mechanisms (which may combine innate and acquired features in different proportions) such as face recognition or reading. In turn, the existence of such specialized mechanism implies that that not all contents have an equal prior probability of stabilizing in a population of human minds. An important question to ask about each cultural practice is whether it is supported not just by general transmission and learning abilities but also, and in some case crucially, by domain-specific abilities (Sperber & Hirschfeld, 2004).

4. Application to empirical research

The interplay between attractors and factors of attraction is at the core of the use of cultural attraction in empirical studies, which I now detail.

Testing for attraction

Under the framework I am sketching here, testing hypotheses on what drives a cultural phenomenon involves several steps: (1) delimitate the system of cultural items to consider, (2) make predictions, and (3) test those predictions on different types of data.

Preconditions: delimiting a system and defining its state

The first question to ask is whether the system / distribution of cultural system that we are observing is at equilibrium or not, as this might influence which characteristics can be observed and exploited. The importance of considering whether a cultural system is at equilibrium or not is recognised in other approaches to cultural evolution as well, and non-equilibrium models are sometimes more appropriate (Kandler & Shennan, 2013). Properties of cultural attractors, and how to diagnose factors of attraction also depend on a cultural system and its state: i.e., a distribution of variants (defined by their features or characteristics) and their associated frequencies. Whether the system is at equilibrium or out of equilibrium in turn implies different behaviours or distributions and methods that can be exploited to detect or test for attraction and what causes it.

At equilibrium, cultural attractors can be described as specific points around which variants cluster, i.e., they can be identified by considering the distribution of variants. Out of equilibrium, attractors can additionally be diagnosed either by changes in frequency in clusters (they are points around which frequency of variants increase) or by the type of transformations that bring about this increase in frequency (i.e., they are directed rather than random transformations).

Making Predictions: Why attraction is not (content) biases

In the framework presented here, precision in the suspected causal factors turns into precision in the predictions. Causal factors—i.e., hypothesized factors of attraction—can be more or less specific to the cultural item studied. Testing for attraction requires to identify a set of constraints and their effect (i.e. a factor of attraction) and infer its consequences at a population-level. This approach includes a relatively high degree of freedom in deciding where and how to include a large diversity of causal factors. Choosing the appropriate model for a cultural item comes first from the characteristics of the cultural content itself and how the content is transmitted (what is the cultural causal chain in which it is embedded).

Attraction has sometimes been understood as a version of content biases. Content biases refer to cases in which some contents are more often copied than others in virtue of one of their intrinsic properties – for instance, including social information (Mesoudi, Whiten, & Dunbar, 2006). There are two main differences between content biases and attraction: they are not assumed to act by the same mechanism, and they make different predictions.

(1) What is the underlying operation? Content biases (social learning strategies) act by *selection*: some contents are selected over others, in virtue of them presenting a given characteristic (for instance, having better pay-offs). Attraction acts by *transformation*: at each time step, any variant is more likely to be transformed into another variant that is closer to the attractor. Occurrences at the attractor or close to it may still transform at each step, but even so all these transformation tend to result into variants close to the attractor. While transformations away from the attractor occur, a long series of such transformations all moving in the direction opposite to that of the attractor is improbable.

(2) The two approaches make predictions that differ in two main ways: (a) in whether they predict evolution towards, or stability at, a relatively precise (fixation) point or just in a general direction, and (b) in how they consider interaction with other constraints. An attractor is a 'fixation point' around which occurrences are likely to cluster. Attraction also has room for interaction between factors of attraction with the location of an attractor resulting from this interaction.

This approach is similar to the logic behind determining an efficiency optimum for semantic systems. This is usually done by first representing the space of possible alternatives (a variation space of possible semantic systems, defined by their position on an x axis – cognitive costs – and on a y axis – communicative costs). The trade-off between communicative and cognitive costs delimitates an area in (the efficiency optimum) along which semantic systems are expected to be found. It is then possible to hypothesize and test which factor would make some variants more likely than others. For instance, the importance of a given domain drives up the requirements for informative content, and moves the semantic system accordingly, towards higher cognitive costs but also higher informativeness (Kemp, Xu, & Regier, 2018). By contrast, content biases usually do not predict an optimal point – they suggest an advantage for a type of content, in a more 'unbounded' fashion. A bias in favour of contents related to disgust, for instance, predicts success for gruesome content over non-gruesome contents. It would not predict that there is such a thing as *too gruesome*.

A cultural attraction-based take, by contrast, predicts the existence of an 'optimal' level of emotionality. Instead of predicting evolution towards more emotional content, it would predict an increase in emotionality **only as long as** the content is less emotional than the attractor point. If the content is actually more emotional than the attractor point, it would predict a decrease in emotionality. Another example is in the relation of a cultural variant to intuitions. Minimally counter-intuitive variants are known to have a mnemonic advantage – they are more easily remembered than variants that do not clearly violate expectations due to intuitions based on ontological categories. Variants that violate too many intuitions don't, however, have such an advantage(Harmon-Vukić et al., 2012; Norenzayan et al., 2006). This type of non-linear relationship falls easily within the purview of the cultural attraction framework.

The second aspect in which (content) biases and attractor-based approaches differ is in how many constraints or factors they can integrate, and in the way in which they do so. Biases approaches usually do not specify how biases are supposed to interact, if more than one is at play. By contrast, attractor-based approaches assume an array of relevant factors of attraction and their likely interactions. Finally, the term 'bias' may carry some ambiguity. Especially when considered across several disciplinary fields relevant to cultural evolution, it tends to refer to quite different concepts. It can freely move between bias meant as an observation (i.e., a statistical bias), or as a cause (i.e., in the behaviour or cognition). Factors of attraction, especially when localised on cultural chains, avoid such ambiguities.

On different types of data

We now detail how it is possible to test for cultural attraction/ the existence of cultural attractors on (1) experimental data, (2) corpus (cultural, real-world) data. Different types of data have their own respective upsides and downsides. Cultural data (corpus data, often large-scale, not experimental) has ecological validity – in some cases, it can provide *natural experiments*, and it offers a range of naturally occurring variation. By contrast, experimental methods offer the possibilities to create alternative versions and to observe transformations but they raise issues of ecological validity.

On cultural (real-world) data

In order to use cultural data to test for cultural attraction, we start from what is known of the phenomenon that we study (for instance, which cognitive processes it might recruit), and extrapolate to predict what the distribution of cultural variants would look like if it were to show the influence of the factors of attraction we have identified. Conversely, cultural data can also be used to test hypothesis *on cognitive processes.* As long as there is an assumed factor of attraction, it is possible to predict what the distribution of cultural variants would be if it were to reflect that factor. For instance, characters from writing systems have features that match the hypothesis of reading being based on a recycling of our visual system (Dehaene, 2010; Dehaene & Cohen, 2007): they follow patterns of occlusion from natural scenes (Changizi, Zhang, Ye, & Shimojo, 2006) and use a disproportionate amount of cardinally oriented strokes (Morin, 2018). Both those features reflect sensitivities of the human visual system that were set before, and independently of, writing.

Most case studies included in this thesis, and especially chapters 2, 3 and 4, respectively focusing on coats of arms (heraldry), scripts (writing systems), and portraits, use cultural data and follow a similar logic.

On experimental data

Experiments, in particular transmission chains (also known as serial reproduction) are one of the methodological tools developed in cultural evolution and can be used in two main ways relating to cultural attraction. As evoked earlier, transmission chain experiments can show that stability and complexity can be acquired *via* change, with no high-fidelity transmission (Claidière, Amedon, et al., 2018; Claidière, Smith, et al., 2014). Alternatively, they offer the possibility to create variants that are predicted to be more or less attractive and thus stable: one way to test for attraction is to compare which variants change (or not), and by how much. Creating such less attractive variants is akin to generating an *out of equilibrium* system: the population of items such a system includes is not one that could naturally be sustained. It can thus be expected that it will move toward equilibrium through successive learning and recall and transmission events.

This is effectively what was done on the case of the cultural practice of bloodletting in (Miton, Claidière, & Mercier, 2015) : studies 2 and 3 of this paper start chains with artificially "less attractive" versions of the remedy, mainly 'non-colocalised' (the bleeding, used as a cure, is done on a different part of the body that the one showing symptoms) and 'accidental' (the bleeding is done by accident, not on purpose) variants of bloodletting narratives. Both those less attractive variants transformed into what was identified as an attractor: a version of bloodletting that is both intentional, and colocalised. This is also the rationale behind chapter 5, in which we vary how well-fitted are the motor constraints to the sound pattern that participants are asked to reproduce. In some conditions, the motor constraints don't match the rhythmical sequence, thus making the rhythmical sequence particularly hard to reproduce faithfully. We expect (and observe) the rhythmical sequence to change into more attractive variants (i.e., rhythms that are easier to produce as they match the motor constraints) through successive episodes of reproduction.

Another transmission chain experiment (Kalish et al., 2007) also operates in the same way. In this experiment, participants had to learn a function relating two lengths on a screen and different chains were started with different functions: some with the function assumed to be a prior in participants' mind, i.e., a linear and positive relationship, and some other with other functions assumed to not match participants' priors (such as non-linear, or linear but negative). By the end of 10 generations of participants, all chains had converge to the linear positive x = y function, independently of the function they started with.

Inferences in empirical case studies

Given the key concepts we presented and the ways in which it is possible to test for attraction, there are two main logics that empirical research can follow: either from factors of attraction to attractors, or from attractors to factors of attraction. In other words, causal inferences can run in both directions between the concepts we presented (see Figure 2).

From hypothesized factors of attraction to observing attractors

First, whenever factors of attraction are already known and evidenced, it is possible to derive predictions on attractors. One way of determining which factors of attraction might influence one cultural type is to reconstitute the causal chain that brings them about. This implies charting out the main characteristics of the cultural type. In which modality is this type expressed? I.e., by which sensory system does it have to be processed? Whether a content has to be perceived and processed by the human visual or auditory perceptual systems means that this content is not submitted to the same pressures. Visual contents are exposed, for instance, to pressures on colours (Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998) or complexity (see chapters 2 and 3)—and by the visual system's resolution. Contents mediated by the auditory system have to face, among others, pressures related to pitch (Bendor & Wang, 2006; McDermott & Oxenham, 2008) and temporal fine structures (Moore & Sęk, 2009). In particular when it comes to such human perception systems, cognitive science has now accumulated a large amount of results with regard to how they process information (including differential sensitivity to different stimuli). Similarly, having to be stored or produced through different systems also exposes different contents to different pressures.

Having identified a potential factor of attraction allows one to predict how this causal factor might cause a cultural item's shape *and* success, or at least which precise kinds or features of a given cultural item would be favoured. The 'recipe' here is as follows: First, (1) identify a potential causal factor. Ideally, it should be reliable and well-understood. It does not need to belong to any discipline in particular. (2) Find a type of cultural items on which this causal factor would have an effect, (3), Use what we know of the effect to derive predictions upon the spatial and temporal distribution of relevant cultural items (i.e., the prevalence of certain types over others). Finally, (4) test those predictions.

An example of this can be found in the study of orientation of strokes in writing systems (Morin, 2018). Because writing exploits the human visual system (Dehaene, 2010; Dehaene & Cohen, 2007), it should reflect the human visual system's sensitivity – for instance, to particular orientations, in this case, cardinals. This is a hypothesized factor of attraction. It predicts that cardinal orientations should be over-represented (i.e., more than could be expected by chance). This prediction is in turn tested - and confirmed - on a large dataset of 116 scripts: a large majority of those scripts over-represent cardinal orientations (Morin, 2018).

This approach is used in most chapters of the present dissertation (all empirical case studies) here – i.e., chapters 2 to 4. It requires only one type of data in order to test the relationship between a factor of attraction and attractors– i.e., it is does not have the additional requirement of having an independent dataset to 'spot' potential attractors beforehand. In cases in which there is no readily available (i.e., documented and evidenced) known factor of attraction, it is possible to start from the other end, i.e, from the attractors, rather than from factors of attraction.

From attractors to factors of attraction: inferring factors of attraction

The other possibility is to reverse-engineer possible factors of attraction. It is a more agnostic approach than the previous one: it does not require *a priori* knowledge of factors of attraction. It is mostly an add-on to the previous logic in order to find potential factors of attraction, which then require to be tested independently.

Here, the first step is to get information on attractors, i.e., the distribution of variants in the population. This task can be best tackled by harnessing social sciences, including historical and anthropological sources. The social and human sciences abound with data that can be used for testing hypotheses in cultural evolution: heraldry (Miton & Morin, 2019; Morin & Miton, 2018), folklorists' folk tales indexes (Tehrani, 2013), database of ethnographic excerpts like the HRAF – Human Relations Area Files (Hruschka, 2010; Miton et al., 2015; Murray, Fessler, Kerry, White, & Marin, 2017; White, Marin, & Fessler, 2017) and now D-Place (K. R. Kirby et al., 2016), and archaeological datasets (Bentley, Hahn, & Shennan, 2004; Crema, Kandler, & Shennan, 2016; Jordan & Shennan, 2003; Shennan & Wilkinson, 2001).

Harnessing this type of knowledge solves one practical problem of empirical research– granularity: relevant types and tokens have already been identified by scholars with the relevant expertise. This is exemplified in the chapters on both heraldic data (chapter 2) and scripts (chapter 3). Motifs (and their occurrences) have been defined and reported by scholars at the time (Renesse, 1894; Rolland, 1909; Rolland & Rolland, 1969), and by historians at the present time (Clemmensen, 2017). As far as characters and scripts are concerned, we relied on the classifications established by the Unicode Consortium (*The Unicode Standard, Version 11.0, (Mountain View, CA: The Unicode Consortium, 2018. ISBN 978-1-936213-19-1)*, n.d.) to a large extent, and on other resources, including (Daniels & Bright, 1996).

In understanding one specific cultural practice, one must determine what is the extent of the variation in the shapes it takes, and whether there are any associated traits that tend to be particularly recurrent. (Relatively) well-spread and stable variants are attractors.

This approach, from attractors to factors of attraction, is somehow more 'data-driven' than the previous one. Here, large-scale data is first examined to extract regular patterns out of it, then one uses those patterns to get insights into what might drive the success/stability, i.e., into *suspected* factors of attraction. Those hypothetical factors of attractions are then tested independently. In a few words, this approach can be summarized as reverse-engineering which factors might support a given cultural item's stability.

The way to proceed here goes as follow: First, (1) pick up a cultural type with relatively clear and uncontroversial principles of inclusion, Then, (2) Collect data, if possible on a large scale, define an array of parameters that might be of relevance - i.e., dimensions that might show patterns of variation (but stability as well). In other words, adjust the granularity. (3) Evaluate how many variants are theoretically possible — establish a way to recode the relevant parameters, recode those relevant parameters, and check them for regularities. (4) From these regularities, reverse-engineer what could have caused them, i.e. form hypotheses about which peculiarities are actually necessary to stability. (5) Test whether those regularities play a role in sustaining the cultural items' stability/success (in a similar way as sketched in the previous part, but on a dataset that is independent from the one used to identify attractors). Finally, (6), Repeat steps 1 to 4 of the previous part (how to test for the effect of factors of attraction).

One case study, bearing on bloodletting (Miton et al., 2015), illustrates those different steps: (1) The cultural type to study was bloodletting, defined as the medical practice of bleeding as a medical cure, i.e., with the goal of curing some ailment or disease, (2) the eHRAF (Human Relations Area Files) was searched the terms bloodletting, bleeding, phlebotomy, venesection and cupping. We then filtered the results to keep only the explicitly medical uses of the practice. (3) The ethnographic excerpts were recoded on criteria including, in that case, tool used/ manner, practitioner, prescriptions, location on the body, which allows to (4) observe regularities on these data: Co-localisation, lack of systematic prescriptions, recurrent 'bad blood' explanation. (5) Transmission chain experiment with less attractive versions (i.e., versions for which one of the suspected factors of attraction is modified, 'de-activated'), e.g., non-colocalised or accidental bloodletting, were used to test for the effect of those factors of attraction.

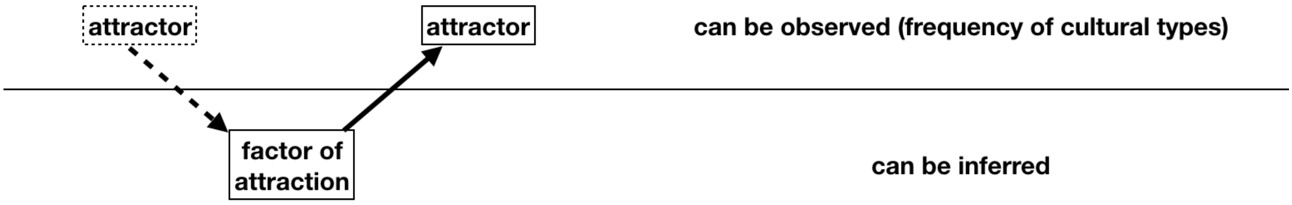


Figure 2. The dotted box and arrow represent the additional step in cases in which there are no factors of attraction known beforehand, while the plain arrow and boxes represent the step going from suspected factor of attraction to predicting (and testing for) attractors. Both attractors (above) can refer to the same attractor, but, in such a case, would have to be identified in distinct and independent datasets.

5. Generalizing from case studies to middle-ground theory

I have now outlined ways in which cultural attraction – understood as a set of possible inferences - can be productively used to scaffold empirical research in cultural evolution.

There has been a healthy scepticism about whether cultural attraction is able to generalize and reach so-called 'middle-ground' theory (Buskell, 2019; Hedström & Bearman, 2011, pp. 25–47; Heintz, Blancke, & Scott-Phillips, 2019). I here present two possible avenues for generalization. Factors of attraction, i.e., types of causes offer two such possibilities: (1) by exploiting similarities in causal chains and (2), by exploiting domain-specificity. Both ways to generalize are based on identifying similarities between cultural types that suggest similarities in the causal factors that sustain them. Such similarities emerge in at least two cases: when different cultural types have similar causal chains, and when different cultural types depend on the same processes, especially if such processes are domain-specific.

First, causal chains provide a template to determine which events and transformations a given cultural type goes through as it gets transmitted. Whenever some steps are common to a large number of cultural items biases relating to these steps can be expected to influence large arrays of culture. Whenever a cultural type requires storing in memory, effects due to this storage in memory, such as simplifications, are to be expected (Bartlett, 1932). Distinguishing different phases within the transmission of cultural contents and mapping which type of biases can occur at each of these phases is an approach that has been adopted in a few studies. For instance, (Eriksson & Coultas, 2014) have distinguished three phases in cultural transmission (choose-to-receive, encode-and-retrieve, and choose-to-transmit), and have been followed by others since (e.g., Stubbersfield, Tehrani, & Flynn, 2015). Similarly, in *Natural causes of language* (Enfield, 2014). Nick Enfield uses such chains in relation to the acquisition of language. Language evolution is particularly lends itself well to isolating which types of evolutionary pressures emerge from which process involved in communication or from which features of communication – e.g., isolating pressures coming from the interaction between speakers (Kanwal, Smith, Culbertson, & Kirby, 2017).

It may be possible to generalize from case studies to other cases within their domain. All visual cultural items, for instance, are subjected to biases that pertain to the functioning of the human visual system, such as being especially tuned to specific orientations (Girshick, Landy, & Simoncelli, 2011).

Finally, to some extent, this introduction and thesis aims at some form of generalization: it suggests how cultural attraction can provide a framework, a system of causal inferences, to study cultural phenomena. It offers systematic ways to study cultural phenomena through this lens and suggests how to test for relevant causal links.

Overall, this thesis aims to illustrate a method, and more specifically a set of rules for inferences between different levels of explanation that would be robust across different cultural domains. This is the very reason why the case studies included in this thesis are so varied, and span across several domains. The method defended here should be equally fit to be applied to any cultural phenomenon of one's choosing.

6. Thesis Outline

In addition to supporting this general framework, the chapters are each intended first and foremost as a contribution to the study of their own topic. They try to answer questions on, respectively, experimental implementations of the concept of cumulative culture, Zipf's law of abbreviation in graphic communication, drivers of complexity in writing, spatial composition in human portraits, and the influence of physical constraints on rhythms in music making.

Chapter 1 is a critical and constructive review of the use of cultural transmission experiments, in particular their use for the study of cumulative cultural evolution. This chapter identifies a couple of mismatches between theoretical definitions of cumulative culture and the way cumulative culture is studied within cultural transmission experiments. It argues for a higher degree of coherence between theoretical concepts and empirical operationalisations in cultural evolution research. It provides grounding in the methods already developed in the study of cultural evolution, cultural transmission experiments in particular.

Chapters 2 and 3 focus on the evolution of visual complexity in two contrasting cases: heraldry and writing systems.

Chapter 2 focuses on relationship between heraldic motifs' complexity and their frequency of use. It was first conceived as a test of Zipf's law of abbreviation. Zipf's law of abbreviation binds micro- and macro-level phenomena: production costs at each occurrence (either directly or through their relationship to informativeness) lead to a specific large-scale distribution. By contrast with other known occurrences of Zipf's law of abbreviation (ZLA), our results on heraldic motifs suggest that (1) ZLA doesn't occur if the relationship between production costs and receiving events is not constant, as, for instance, when the cost has to be paid each time an information is to be transmitted, and that, (2), a 'reverse' ZLA, in which more complex motifs are more frequent, can occur in real-life settings under specific evolutionary pressures (signalling function, low production costs, iconicity presenting an alternative way of decreasing the cost of processing signals).
This second chapter goes hand in hand with another paper on heraldry, in which Olivier Morin and I suggested that a shuffling model and real-world data's departure from it can be used as a diagnostic tool for high-fidelity copying (Morin & Miton, 2018). It also informs current debates on the necessary conditions for Zipf's law of abbreviation to emerge, by documenting one -rare- instance in which longer signals (more complex images) were also more frequent.

Chapter 3 focuses on scripts & their characters' visual complexity. Hypotheses investigated in this chapter have different origins: the relationship between a writing system's inventory size and its graphemes' complexity has already been explored (Chang, Chen, & Perfetti, 2018; Changizi & Shimojo, 2005), and we replicate it. Drawing on (Morin, 2018) this third chapter gives a complete story of how some dimensions appear already at the attractor, while others come from repeated transmission.

The chapter is to be understood in relationship to one other paper (Kelly, Winters, Miton, & Morin, submitted), which offers an insight into the evolution of one emergent script in particular, the Vai script from Liberia. Chapter 3 draws on similar assumptions, mainly that visual complexity decreases over use and transmission episodes, but uses evidence from a dataset of 133 writing systems.

Chapters 4 and 5 both study factors of attraction. They differ in (1) which domain they focus on (depictions versus rhythmical sequences), (2) which methods and type of data they use (historical/cultural versus experimental), and (3) which type of factor of attraction they focus on (cognitive versus ecological).

Chapter 4 tests for the existence of forward bias in portraits and thus illustrates a cognitive factor of attraction in how to represent agents. We start with a well-evidenced tendency for cognitive processes to influence both production and aesthetical perception of agents' pictorial representations based on their spatial composition: people have been shown to prefer to represent agents with more space in front of them than behind them. We predict that some cultural items (here human profile-oriented portraits) will reflect this factor. In other words, we expect a continuity between cognitive biases at the level of the individual to impact large-scale distributions. Additionally, we also test for the existence of historical dynamics and interaction with norms.

Chapter 5 presents an iterated learning (transmission chain) experiment in which participants are asked to reproduce as faithfully as they can a rhythmic sequence using drum pads. It is a between-subjects design in which we varied the movements participants had to do while reproducing the rhythmical sequences they heard. This experiment was designed to test the role of ecological, non-cognitive factors of attraction. The affordances at the micro-level (i.e., for each participant) on the motor aspect lead to the emergence of stable and predictable patterns. In this chapter, experimental methods are used to create different conditions starting at different distances of predicted attractors.

- All empirical studies, i.e., chapters 2 to 5 were pre-registered on OpenScienceFramework (osf.io) – their associated data and R scripts used for analyses are also all online, at their respective URLs.

Chapter 1:
Cumulative culture in the laboratory: Methodological and theoretical challenges

1. Introduction

First introduced in Bartlett's classic study of memory (Bartlett, 1932), cultural transmission experiments, or CTEs, are now being used in cognitive science, social psychology, behavioural biology, and cultural evolution. Whereas psychological experiments on learning typically deal with individuals solving a task on their own, transmission experiments allow participants to learn from one another. Doing so, they make it possible to "capture the repeated occurrences of social learning involved in cultural change, as opposed to one-off cases of individual learning" (Caldwell, Cornish, & Kandler, 2016, p.2), making them powerful tools to study culture under controlled conditions. Ten years ago, Mesoudi & Whiten (Mesoudi & Whiten, 2008, p. 200) observed that "perhaps due to the sparseness of past experimental studies and the lack of any guiding theoretical framework, these questions and methods have not been addressed in a systematic fashion, and answers to each must be said to be sketchy at best". Since, these valuable experiments have become increasingly popular due to their intensive use in the study of cultural evolution and now ground a productive and exciting new experimental field (cf. ESM-1, Appendix A). Several recent reviews have summarized the various methods used in CTEs, the research topics typically investigated, and the findings of these experiments (Caldwell, Atkinson, & Renner, 2016; Caldwell & Millen, 2008b; M. Kempe & Mesoudi, 2014b; Mesoudi, 2016c; Mesoudi & Whiten, 2008; Whiten, Caldwell, & Mesoudi, 2016). In so doing, these reviews have contributed to making CTEs a scientific success.

In their ecological (i.e. real world) settings, cultural phenomena are often large-scale population-level phenomena and span over several biological generations. In contrast, laboratory experiments involve much smaller artificial groups over much shorter time periods. An intrinsic challenge faced by CTEs consists in dealing with this asymmetry: How can these experiments retain the relevant features of actual cultural populations so as to serve as proper models? In this review, we address these challenges in the context of the study of cumulative cultural evolution, i.e., the gradual improvement of traditions over time.

In section 1, we review the use of CTEs in the study of cumulative culture and detail the specific methods employed to do so. In section 2, we identify two issues raised by the implementation of CTEs and which challenge the interpretation of their results. We suggest solutions to these issues so as to get the most out of these experiments. In section 3, we argue that CTEs often fail to take full advantage of the information they collect. We show how the use of design spaces would make it possible to exploit this latent information, thus expanding the range of hypotheses about cumulative cultural evolution that CTEs could be used to test.

2. Cultural transmission experiments and cumulative culture

Cumulative cultural evolution refers to the process by which traditions are gradually modified and, for technological traditions in particular, improved upon over time. The repeated modification and transmission of culture is commonly described as a form of descent with modification characterized by a 'ratchet effect' (e.g., Tennie, Call, & Tomasello, 2009; Tomasello, 1999; Tomasello, Kruger, & Ratner, 1993). The ratchet effect is a mechanical metaphor that stresses the role of social transmission in "locking in" novel modifications of socially transmitted traits in a population's cultural repertoire. As modified traits are transmitted, further modifications can in turn be made and transmitted. With time, it is expected that populations will build up increasingly complex cultural traits that no single individual could have invented on its own. A capacity for cumulative culture would be adaptive by allowing populations to collectively reach solutions that are beyond the problem-solving capacities of the individual (Muthukrishna & Henrich, 2016) (or beyond the species' existing cognitive repertoire, i.e., their zone of latent solutions (Tennie et al., 2009)), but also by allowing the distribution of efforts, risks, and time costs of invention by trial and error over several generations (Boyd & Richerson, 1996; Lewis G. Dean, Vale, Laland, Flynn, & Kendal, 2014).

CTEs have been used to investigate three main questions regarding cumulative culture. The first concerns the identification of the key adaptations leading to the onset of cumulative culture in the human lineage, in contrast to other non-human species that may have cultural traditions but of the non-cumulative kind. CTEs have thus been used to identify social learning processes, such as imitation, emulation, and/or teaching, that can lead to the cumulative increase of performance in functional tasks (e.g., Caldwell & Millen, 2009; Wasielewski, 2014; Zwirner & Thornton, 2015). The second issue concerns the impact of demography on the cumulative process. It has been investigated by means of CTEs by varying the number of participants and of models (e.g., Caldwell & Millen, 2010b; Derex, Beugin, Godelle, & Raymond, 2013; M. Kempe & Mesoudi, 2014a; Muthukrishna, Shulman, Vasilescu, & Henrich, 2013) or by varying the density of their interactions (or 'connectivity'; e.g., Derex & Boyd, 2015, 2016). The third issue concerns the role of inductive biases in the emergence of new complex traits without design, for instance, in the study of linguistic structures (Kirby, Cornish, & Smith, 2008; Kirby, Tamariz, Cornish, & Smith, 2015).

When dealing with human adults as participants, CTEs of cumulative culture typically employ one of three methods: linear chains (also called diffusion chains), replacement, and closed-group (or constant-group) methods (M. Kempe & Mesoudi, 2014b). In linear transmission chains, a first participant is presented with a stimulus and must later recall it. The output of this first participant serves as the input stimulus for a second participant, who has to recall it in turn. This is repeated several times until it reaches the last participant in the chain. In the replacement method, groups of participants solve some task, either once or repeatedly, either individually or collectively. At each time step, new participants replace some of the previous ones and learn from the group how to solve the task. Finally, in the closed-group method a group of participants aims at solving some task. In between two trials, the participants can learn from one another, the specific means of doing so being dependent on the experimental design.

Most of the tasks used in these experiments consist in solving problems according to some performance criteria, for instance assembling a jigsaw puzzle (Kempe & Mesoudi, 2014a) or solving anagrams (Baum, Richerson, Efferson, & Paciotti, 2004). Many of these problem-solving tasks require the production of artefacts, either real ones – such as paper planes, spaghetti towers, rice baskets, weight-bearing devices, or stone flakes (Caldwell & Eve, 2014; Caldwell & Millen, 2008a, 2009, 2010b, 2010a; Morgan et al., 2015; Wasielewski, 2014; Zwirner & Thornton, 2015)–, or virtual ones – such as virtual totems, fishnets, and arrowheads (Derex, Beugin, et al., 2013; Derex & Boyd, 2015, 2016; Derex, Feron, Godelle, & Raymond, 2015; Derex, Godelle, & Raymond, 2013; Mesoudi, 2008, 2011a) –, or both (Muthukrishna et al., 2013). Other experiments consist of participants transmitting some information, with or without explicit instructions to transmit it as faithfully as possible (Beppu & Griffiths, 2009; Caldwell & Smith, 2012; Kirby et al., 2008, 2015; Martin et al., 2014; Tan & Fay, 2011). These tasks can usually be solved in multiple ways, with some solutions being more rewarding or effective than others.

CTEs using children or non-human animals as participants have employed linear chains, replacement, or seeded group (or seeded open diffusion). Seeded groups consist in training an individual to complete a task (e.g., how to use some apparatus) and then allowing the individual to freely engage with it. Other participants observe the individual and are in turn also free to engage with the task. Experiments with children and non-humans use a variety of tasks such as opening an artificial fruit (Dean, Kendal, Schapiro, Thierry, & Laland, 2012; Flynn, 2008; Vale, Davis, Lambeth, Schapiro, & Whiten, 2017), building or combining some tools in order to solve a foraging problem (Davis, Vale, Schapiro, Lambeth, & Whiten, 2016; McGuigan et al., 2017; Tennie, Walter, Gampe, Carpenter, & Tomasello, 2014), reproducing visual patterns (Claidière et al., 2018; Claidière, Smith, Kirby, & Fagot, 2014; Kempe, Gauvrit, & Forsyth, 2015), or finding a way back home (Sasaki & Biro, 2017).

3. Miniaturizing culture: Learning time and Task Complexity

According to theoretical accounts of cumulative culture (Boyd & Richerson, 1996; Caldwell, Atkinson, et al., 2016; Caldwell & Millen, 2008a; Dean et al., 2014; Mesoudi & Whiten, 2008; Richerson & Boyd, 2005; Tennie et al., 2009), a diagnostic criterion for a cultural process to be properly cumulative is that it leads human cultures to "accumulate changes over many generations, resulting in culturally transmitted behaviours that no single human individual could invent on his own." (Boyd & Richerson, 1996, p. 80). Exactly how the diagnostic criterion is to be understood remains largely open to interpretation. Under one understanding, satisfying the criterion would mean using a task that is too complex for a single individual to solve on its own, requiring a collective of individuals, such as a tradition, to reach the solution (Muthukrishna & Henrich, 2016; Tennie et al., 2009). Alternatively, the criterion can be interpreted to mean that a task is too complex for an individual to solve during her *limited lifetime* but that it could be solved by a tradition as it distributes the effort and time required to solve the complex task over the lifetimes of multiple individuals (Mesoudi, 2011c). Under both interpretations, aiming to satisfy the diagnostic criterion would "effectively eliminate the possibility of experimental research as participants would be incapable of completing the task." (Zwirner & Thornton, 2015, p. 7). Such experiments would additionally need to span over multiple lifetimes if any cumulative effect is to be detected. These two aspects of cumulative culture – complex ecological tasks and large timespans – pose serious challenges for any experimental work on cumulative culture so defined.

The strategy used to circumvent these difficulties consists in 'miniaturizing' generational change to fit laboratory conditions and to use tasks that can be solved by a small number of participants in a limited amount of time. The turnover of participants in the laboratory traditions, or microsocieties, is then taken to model actual generational change in virtue of linking the participants by episodes of cultural transmission (Mesoudi & Whiten, 2008). For transmission chains and replacement methods, modelling generational change is achieved by the replacement of a participant by another, with each individual representing a different generation. In closed-group experiments, generations are instantiated by the successive rounds of a given task that a same group of participants solves[1]. An experimental analogue of the diagnostic criterion could then be used to detect a genuinely cumulative process in the laboratory: showing that laboratory traditions can reach higher performance levels in solving some task than individuals doing so on their own then offers, it is suggested, a scaled-down cultural cumulative process (Caldwell et al., 2016).

In order to test whether a cumulative effect has been obtained, CTEs need to systematically compare how, for a same task, individuals fare on their own in solving the task with the solutions achieved by laboratory traditions[2]. Few CTEs effectively make this comparison. Those that do include as a control a "non-social condition" where single participants, with a limited time budget, are asked to solve a task and improve their performance on their own over repeated trials (Derex & Boyd, 2015; Derex et al., 2015; Derex, Godelle, et al., 2013; Mesoudi, 2008, 2011a; Tennie et al., 2014; Zwirner & Thornton, 2015). The accumulation of improvements during a run of the non-social condition is interpreted as an effect of individual learning, with the increase in performance observed in these conditions indicating that participants are getting more skillful in solving the task. In contrast, in "social conditions," several participants are involved in solving the same task, with the solution produced by one participant made available to the next participant(s). When the performance achieved in the social condition surpasses the performance achieved in the non-social condition, the difference in performance is seen as evidence of a cumulative cultural process.

---

[1] Although such successive rounds are sometimes understood as laboratory generations (e.g., Mesoudi & Whiten, 2008), in some CTEs they are instead understood as episodes of social learning during the lifetime of individuals (e.g., Derex, Godelle, & Raymond, 2014; Mesoudi, 2008).

[2] Most experiments with children and non-humans use seeded groups when testing for cumulative culture.

In those cases, the performance obtained within the (seeded) group is compared to the performance obtained in individual controls – i.e., individuals who can manipulate the apparatus or engage in the task without the possibility to observe other conspecifics. However, in seeded groups, the population is held constant and participants only take part to the task when they want to. It is thus unclear if these can be used to test for cumulative culture as there is no laboratory generations to implement the diagnostic criterion. This makes experiments using seeded groups even more susceptible to demonstrating different results between individual and social conditions not because of a cumulative effect, but for other reasons such as, e.g., social facilitation (see below).

Using small groups of participants over short periods of time is necessary to effectively run a CTE. However, the discrepancies between the spatiotemporal scale of the processes leading to real cultural change and those in action in the laboratory microsocieties can lead to several mismatches between the underlying causes of the differences in performance. We now turn to two types of such mismatches that may undermine the evidential value of the diagnostic criterion as it is implemented in CTEs. We then suggest means to overcome these problems.

(a) Learning time

Introducing some novel terminology will prove useful at this point. We refer to the sum of all the learning and trial time invested by an individual participant in a non-social condition as the total learning time of an individual. We refer to the sum of the time spent learning and performing a task by all the participants of a chain or group during one run of an experiment as the total additive learning time of a tradition. See ESM-2 (Appendix A) for a synthetic comparison of the two kinds of total learning times in the CTE literature and how these time budgets are calculated.

In experiments with both non-social and social conditions and with equal total learning time (i.e., of an individual and of a tradition), traditions in the social conditions tend to produce better performance results than the individual participants in the non-social ones (e.g., Derex, Godelle, & Raymond, 2014; Mesoudi, 2008). These differences in performance may suggest that the laboratory traditions produce traits that are out of reach to single individuals, and thus that a cumulative effect has occurred. We argue, however, that the control conditions currently used in CTEs are insufficient to guarantee that the observed differences in performance are the results of a genuine cumulative process.

The differences between the performance of participants in the non-social and social conditions may, in particular, be the result of laboratory traditions reaching performance levels that are well within the reach of a single participant working alone, but doing so faster than participants working alone. Consider first that the total additive learning time of an experimental tradition is at most a few hours (see ESM 2-2, Appendix A). It is thus a possibility that what the difference in performance measures is not so much a difference in capacity between a collective and an individual as one in improvement speed, e.g., it could be that participants in a social condition have their individual skill improvement facilitated by their learning in a social context

(Bond & Titus, 1983). If this were the case, then the observed asymmetry would not be due to a cumulative process producing traits too complex for an individual to invent alone, as demanded by the diagnostic criterion. Instead, by fixing the time allowed for participants in non-social conditions to improve their performance, the asymmetry would be the result of ending the experiment prematurely, i.e., were they given more time, the participants in non-social conditions could have reached the same solution as those in social conditions.

There is a straightforward way to ascertain whether participants in a social condition increase their performance not just faster but also beyond the reach of participants in a non-social condition and thus to determine whether a genuine cumulative process is implemented in a CTE. To begin with, all CTEs should systematically include a baseline non-social condition in which a single participant solves the same task over multiple trials. The majority of CTEs testing for cumulative culture lack such non-social control (only about a third of them have such control conditions, see ESM 2). The non-social condition should minimally involve as many individual trials as there are generations in the social condition, with equal learning times for each individual trials and social generations. In other words, the total learning time of an individual should be equal to the total additive learning time of the tradition to which it is compared – a measure so far only implemented in (Zwirner & Thornton, 2015). This would allow a consistent comparison of the results in social and non-social conditions; it could confirm that, across tasks, traditions perform better than individuals working alone. The total additive learning time of traditions should also be of the same duration across different social conditions, e.g., when comparing the effect of different transmission mechanisms (see ESM-2, Appendix A). These suggested controls participate to a more stringent criterion for the detection of cumulative culture in experimental settings. By doing so, they allow distinguishing between confounding factors and thus eliminate some risks of erroneously concluding to a cumulative effect when there is none (type 1 errors). Additionally, by equalling the time budgets between the relevant conditions, we can compare rates of learning and how they impact the emergence of cumulative culture.

(b) Task complexity

Miniaturizing cultural phenomena to fit experimental settings has further consequences for the validity of the results obtained through CTEs. CTEs demand that the experimental tasks presented to the participant be *rapidly* solvable – relatively to the time available to individual learners and to traditions in ecological conditions – if only because CTEs require a much reduced timescale for their implementation in contrast to real intergenerational cultural processes. In order to deal with this time constraint, CTEs typically include tasks that are easier and more straightforward to solve relatively to the real, complex ecological problems solved by human cultures. However, scaling down the complexity of a task risks trivialising the results of any CTE. Indeed, cumulative culture is a process that is supposed to allow collectives of individuals to solve complex problems, problems that no single individual could have solved on their own during their lifetime. Instead, most tasks used in CTEs are simple enough that they do not require the collective effort of a group of individuals to be solved (e.g., solving a jigsaw puzzle, building spaghetti towers), thus casting doubt on their ecological validity (Derex, Godelle, et al., 2013).

Coping with the constraints of the reduced spatiotemporal scale of laboratory experiments by using simpler problems can be both misconceived and misleading. Simpler problems and their associated solutions are *not* miniaturized, scaled-down versions of complex problems and solutions because complexity is not a scalable property of a system (e.g. building a nanorobot means making a smaller, but not a simpler version of a complex, human-sized robot). Simplicity is not complexity at a smaller spatiotemporal scale. Consequently, dealing with simple tasks in CTEs can lead to results that are in fact not representative of cumulative culture at all.

For instance, in most CTEs used to investigate cumulative culture, we observe a nearly systematic improvement of the traditions at each laboratory generation (e.g., Caldwell & Millen, 2010b). In other words, in a laboratory setting, most individual (adult) participants can and often do improve upon the tradition they inherit from the other participants. This general result among CTEs leads to a tension with both theory and fieldwork observations.

From a theoretical point of view, it is often assumed that inventions contributing to cumulative culture are uncommon, either because they are hard to achieve or because they are costly to acquire (Boyd & Richerson, 1996). "It is the selective transmission of *lucky errors* and *occasional experiments* that drives much of the evolution of adaptive technology, skills, beliefs, and practices" (Henrich, 2004, p. 202; emphasis added). Hence the alleged importance of high-fidelity transmission mechanisms to preserve these rare and precious innovations whenever they appear (securing the so-called "ratchet effect"; (Henrich, 2010; Tomasello, 1999). However, taking the results of CTEs at face value, it seems that even if cultural transmission had low fidelity, the individual capacity to successfully improve upon traditions is so common and systematic that it could easily compensate for the repeated loss of most improvements.

These results also clash with fieldwork observations, especially in regard to technological traditions. Participants in CTEs are rarely, if ever, experts in the task they are asked to solve. CTEs typically demand relatively simple tasks that neophytes can readily improve upon and do so in a very limited amount of time. In contrast, in naturally occurring cultures, it seems that most innovations originate in some rare creative individuals, or lead users (von Hippel, 1988), often experts in their field (Henrich, 2004; Jones, 2010; Lehman, 1953; Simonton, 1996). This expertise is itself acquired through repeated, deliberate practice, and over many years (Ericsson, 1996; Ericsson & Lehmann, 1996; Stout, 2002). Yet, CTE results seem to indicate that most participants are adept inventors, capable of innovating in matters of minutes. These results thus conflict with what we know about actual cumulative change, which suggests that the tasks used in CTEs are not a simplified version of cultural technical traditions but a kind of practice too different from these traditions, too devoid of ecological validity to allow confident extrapolation.

There are several possible ways to avoid reducing the complexity of the task without increasing the duration of the experiment beyond reason. A first possibility is to recruit participants that are already experts in the type of tasks used in the experiment, or that have already acquired some skills that would help them in succeeding at the experimental task (for experiments involving experts, see, e.g., (Keller, Knoblich, & Repp, 2007; Maguire et al., 2000; Sammler, Novembre, Koelsch, & Keller, 2013). To the best of our knowledge, no CTEs have reported using participants with previous experience in the type of task they were asked to solve (or evaluated their initial expertise in such tasks), in contrast with fieldwork experiments and observations dealing with the acquisition of complex skills (Gandon, Roux, & Coyle, 2014; Stout, 2005). A second possibility would be to train participants until they reach a given level of proficiency in the task used in the CTE (or a similar task recruiting the same set of skills) before they take part in the actual experiment. The level of proficiency could be set as reaching some performance score or reaching a degree of improvement of performance between runs of the training task that becomes small enough. Several experiments on language evolution have opted for this latter solution and include a training phase in their design (Kirby et al., 2008, 2015). Both solutions would ensure that observed increases in performance would not be due to an initial learning phase where the participants acquire initial skills but rather to the emergence of cumulative effects.

4. Using design spaces to study variation

In the previous section we have examined some issues arising from the necessary miniaturization of cultural phenomena and have suggested solutions and improvements in addressing them. We now turn to the measurement of variation in laboratory traditions. Just as the conservation of innovations is a key condition for cumulative processes, so is the production of the variation to be retained and passed on from one generation to the next. Two aspects of CTEs, namely the collapsing of each cultural generation into single individual participants and the use of unidimensional metrics of cumulativeness, drastically curtail the variation that may be observed, which raises specific issues in the interpretation of the experimental evidence. We suggest means to make information about this variation available and exploit it. By examining in more details the role of variation in the production of cultural traditions, CTEs might provide, we suggest, better insights about cumulative culture than they currently do.

(a) Accessing information about variation in CTEs

Most CTEs do pay some attention to the relationship between cumulative culture and cultural variation. These experiments, however, either aggregate data about variation *between traditions* at each generation or *between generations* of a same tradition, but never measure variation *within* generations (i.e., between participants of a same generation of a same tradition). For many CTEs, specifically those using linear transmission chains and replacement paradigms, the microsocieties used often reduce a whole cultural generation, which, in real life, typically involve many interacting individuals, to a single participant[3]. This precludes interactions between within-generation variability and the cumulative process, interactions that are likely to be important to understand the cumulative process. Does, for instance, cumulative culture depend on highly homogeneous populations, with little variance in their cultural traits, or does it depend on a population with a larger, richer cultural repertoire, one that would promote the recombination of existing solutions into better performing ones (e.g., Charbonneau, 2016; Mesoudi & O'Brien, 2008)? How do novel or improved traits diffuse throughout a population (a process known as innovation (Rogers, 2003)), which individuals are more receptive to innovations (e.g., venture,

[3] In contrast, closed group experiments do, by including several participants in a same generation, maintain some kind of intra-generational variation, which is measured but rarely interpreted. However, here the modelling of generational change is compromised as each generation is populated by the very same participants throughout the whole experimental tradition. This has the effect of reducing the impact of interindividual differences over the cumulative process.

early, or laggard adopters (Rogers, 2003)), and what kind of individuals tend to improve upon traditions (e.g., by comparing the innovative capacities of experts and non-experts)? Without intragenerational variability, we cannot address these questions.

We thus suggest that, for more informative experiments with greater ecological validity, CTEs should as much as possible have several participants at each generation instead of representing generations each by a single participant. Actually, multi-participant generations have already been used, and with great success, especially in research concerned with the evolutionary impact of the number of models one learns from and on the effect of group size on the cumulative process (Eriksson & Coultas, 2012; Kempe & Mesoudi, 2014a; Muthukrishna et al., 2013), although again with little if any detailed investigation of within-generation variation.

A second problematic aspect of CTEs is that cumulative change is typically measured strictly in terms of performance or complexity. Indeed, in many CTEs dealing with the creation of real (Caldwell & Eve, 2014; Caldwell & Millen, 2008a, 2009, 2010b, 2010a; Muthukrishna et al., 2013; Wasielewski, 2014; Zwirner & Thornton, 2015) or virtual artefacts (Derex, Beugin, et al., 2013; Derex & Boyd, 2015, 2016; Derex et al., 2015, 2014; Derex, Godelle, et al., 2013; Mesoudi, 2008, 2011a; Muthukrishna et al., 2013), the effects of cumulative culture are usually reported as some simple quantitative score – either a performance score or a measure of complexity –, where traditions with higher scores are understood to have undergone more cumulation. Cashing the effects of the cumulative process into a unidimensional score passes over relevant information about the relation between variation and the cumulative process. How do the properties of each specific solution (e.g., the size and shape of some artefact) vary from one tradition to another? This is generally left unreported. Some experiments have used similarity ratings, but since the metrics were based on the intuitions of naïve coders, they throw little light on the variability of the solutions devised by participants (e.g., Caldwell & Millen, 2008a).

Measures of performance and complexity are of course important. Nevertheless, measuring the effects of the cumulative process strictly in terms of a unidimensional score doesn't help us understand why and how differences in performance and complexity emerge in the first place and the extent to which different traditions have specific evolutionary trajectories. For this reason, we believe it would be beneficial to use multidimensional metrics to study cumulative culture. We propose to do so by using what is commonly known as 'design spaces'.

(b) Using design spaces

A useful strategy to examine how the solutions found by participants vary beyond differences in performance or complexity scores is to employ design spaces. The use of design spaces is not new. They have already been exploited in evolutionary modelling (e.g., Acerbi, Tennie, & Mesoudi, 2016; Acerbi, Tennie, & Nunn, 2011; Charbonneau, 2015) and in the study of specific traditions, especially in archaeology (e.g., O'Brien et al., 2016). The novel development we suggest is to exploit design spaces in the specific context of *experimental* investigations of cultural evolution.

In the context of CTEs, a design space is the set of all potential solutions to an experimental task that participants may produce. We can build a design space by identifying specific dimensions along which these solutions may vary. For instance, when dealing with spaghetti towers, one can measure not only the height of the structures (which typically serves as the measure of performance (e.g., Caldwell & Millen, 2008a; Reindl, Apperly, Beck, & Tennie, 2017) but also the number and mean angles of edges, number of vertices, number of levels, amount of spaghettis and plasticine used, the shape of each level (e.g., triangular, quadrangular, etc.), general symmetry of the structure, and so on. By measuring these different features, one can then plot each solution onto the multiple dimensions of the design space, together with the specific generation at which the solution was produced. When represented visually, specific measures are used as spatial coordinates. It is then possible to visualize how a population of solutions evolves in time by plotting the generation at which they were produced. The evolutionary trajectories of traditions and of the participants in non-social conditions can then be compared and differences in exploratory behaviours assessed.

Using design spaces enables a richer analysis of the data obtained through CTEs at minimal cost in three ways. First, it avoids losing information about the variation in the participants' solutions and allows to compare variability both within and between generations. Second, it offers finer-grained descriptions of the evolution of traditions than measures of performance and complexity. Third, design spaces can also be used to compare which aspects of the solutions are faithfully retained and which ones are modified (and how they are modified).

Moreover, design spaces can also be used to study aspects of cumulative culture that otherwise couldn't be studied through unidimensional measures of performance and complexity. We now introduce two such questions and discuss how design spaces can be profitably put to use in addressing them.

What sorts of evolutionary dynamics characterize cumulative culture?

When measuring cumulative change on a single dimension, e.g., in terms of performance, the only evolutionary patterns exhibited by experimental traditions are changes over time (increases, stagnations, or decreases) on this single dimension. However, this fails to capture the diversity of possible evolutionary trajectories that can lead to a same outcome (see ESM-3, Appendix A). When measuring changes on the multi-dimensional properties of solutions through design spaces, on the other hand, a broader range of evolutionary dynamics can be brought to light (Caldwell, Cornish, et al., 2016). Unlike unidimensional measures, a multidimensional design space makes it possible to distinguish and to properly describe traditions starting from different less effective solutions and converging on the same improved solution and traditions starting from the same less effective solutions but diverging towards two different but equally effective solutions. Examining whether traditions converge or diverge can help us study the conditions under which cumulative processes are more or less path dependent and which sorts of constraints act on the potential evolvability of cultural traditions (Charbonneau, 2015, 2018a; O'Brien, Buchanan, & Eren, 2018). See ESM-3 (Appendix A) for additional examples of alternative evolutionary trajectories.

Traditions and individuals may explore design spaces in different ways. A participant learning on her own tries to improve over her previous performances. Her own biases and idiosyncrasies are likely to lead her to produce new solutions that are nevertheless closely similar to her own previous efforts, effectively leading to small, incremental changes in the design space. In contrast, in a social learning context, participants have to improve solutions devised by other participants, solutions that they might not have thought of on their own. Consequently, each participant in a social condition may bring their unique contribution to the process and generate novel solutions that are further away in design space. For instance, recombinant learning can combine the innovations of others with one's own, thus effectively leading to larger jumps into design space (Charbonneau, 2015, 2016). Differences in the distances traversed by individuals and traditions could allow traditions to explore larger areas of design spaces, find novel solutions with greater ease, and even effectively jump from one local performance optimum to another (Acerbi et al., 2016, 2011; Charbonneau, 2015; Mesoudi, 2008, 2011a). Moreover, there are already some experimental results suggesting that single participants simply do not follow the same trajectories as traditions with multiple individuals. For instance, Claidière et al. (2014) have found, in a transmission chain with baboons, a higher performance and transmission fidelity in traditions with multiple participants than in individuals presented with their own earlier outputs (but see Ravignani, Thompson, Grossi, Delgado, & Kirby, 2017).

Do traditions using different types of social learning explore the design space differently?

Most CTEs testing for the effect of different types of social learning (e.g., emulation and imitation) generally only consider differences in the overall improvement in performance, viewed as an evolutionary signature of the learning processes (e.g., Caldwell & Millen, 2009; Derex, Godelle, et al., 2013; Morgan et al., 2015; Reindl et al., 2017; Wasielewski, 2014). By using design spaces, on the other hand, one could examine whether traditions using different learning mechanisms evolve along the same pathways or follow different trajectories.

Some empirical work has been dedicated to identify population-level patterns (or signatures) specific to different evolutionary processes and learning mechanisms in human populations and the archaeological record (Hamilton & Buchanan, 2009; Herzog, Bentley, & Hahn, 2004; Kandler, Wilder, & Fortunato, 2017). The use of design spaces would contribute to this line of work by experimentally identifying more such signatures and adding experimental evidence in support or against already used ones. For instance, in some social learning conditions, participants may be very conservative, both exploring very few novel solutions for a task and faithfully copying the models of the previous generation. In such scenarios, a cumulative effect would be marked mostly by the faithful retention of past innovations, leading a population to slowly move through the space (gradual change) and minimizing the spread of the population over the design space. In contrast, a cumulative effect could be due to participants selectively using some information from the previous generation (e.g., understanding the physical principles behind the working of some artefact) but innovating over it with little concern to produce similar solutions. In such cases, the cumulative effect would be mainly due to individual innovations, with the population exploring larger regions of design space and at a quicker pace, leading the population to spread more easily over the design space.

5. Conclusion

Culture is a large-scale phenomenon, taking place in populations over multiple generations. In contrast, cultural transmission experiments (CTEs) have narrowly restricted spatiotemporal scales. We identified two issues induced by fitting cultural phenomena in experimental settings and offered potential solutions for each. The first issue concerns the balancing of learning times between conditions. A second issue deals with the complexity of the experimental tasks and its relation to skill acquisition. We then argued that significant information about the evolutionary behaviours of traditions are often compressed into unidimensional scores but could be expanded and better exploited through the use of design spaces.

Experimentally studying cultural transmission is both a difficult and a promising endeavour. In this review, we suggested ways in which this already successful experimental field could be further improved.

Visual complexity in graphic communication systems

The two next chapters have three common aspects: (1) they focus on the visual complexity of elements in graphic communication systems, (2) have hypotheses grounded in trade-offs from language evolution, and (3) they use large cultural-historical datasets, exploited with computational tools.

1. Complexity

Complexity is, as evoked in chapter 1, one of the main criteria for cumulative culture. But complexity is also a term notoriously vague and hard to define. One illustration of this fact can be found in Lloyd's (2001) list of 40 definitions (table 1). His list is further organized under three main questions: how hard it is to describe, how hard it is to create, and what is its degree of organization.

How hard is it to describe? Typically described in bits
Information
Entropy
Algorithmic Complexity / Algorithmic Information Content
Minimum Description Length
Fisher Information
Renyi Entropy
Code Length (prefix-free, Huffman, Shennon-Fano, error-correcting, Hamming)
Chernoff Information
Dimension
Fractal Dimension
Lempel-Ziv Complexity
How hard is it to create? Typically measured in time, energy, dollars, etc.

Computational Complexity **Time Computational Complexity** **Space Computational Complexity** **Information-Based Complexity** **Logical depth** **Thermodynamic depth** **Cost** **Crypticity**	
What is its degree of organization?	
Effective Complexity: difficulty of describing organisational structure	**Mutual Information:** amount of information shared between parts of a system
Metric Entropy **Fractal Dimension** **Excess Entropy** **Stochastic Complexity** **Sophistication** **Effective Measure Complexity** **True Measure Complexity** **Topological epsilon-machine size** **Conditional Information** **Conditional Algorithmic Information Content** **Schema Length** **Ideal Complexity** **Hierarchical Complexity** **Tree subgraph diversity** **Homogeneous Complexity** **Grammatical Complexity**	Algorithmic Mutual Information Channel Capacity Correlation Stored Information Organization

Table 1. Lloyd's 40 possible definitions of (or ways to operationalize and quantify) complexity (Lloyd, 2001).

One major issue with complexity is that it can be defined in two diametrically opposed ways. It can be defined either in terms of regularities or in terms of randomness, i.e., departures from regularity and predictability. This dichotomy further applies to measures of complexity. These two ways of defining complexity make it difficult to both define complexity and find proper metrics to evaluate it. For instance, when applied to historical events, « We can measure the complexity of a time series in terms of both its regular and random components. […] This leads to two contrasting views of complexity, one emphasizing the random (Kolmogorov complexity) and the other, the regular (effective complexity). » (Krakauer, 2011, p. 89)

Whereas there might not be any easy way to decide on one definition of complexity at the conceptual level, it may be much easier to choose a relevant implementation of complexity. Two questions are relevant and productive when considering complexity in relationship to cultural types: (1) who or what processes tokens of a given cultural type (population dependency); and (2) by means of which perceptual modalities or psychological mechanisms are these token processed (domain-dependency). Complexity becomes easier to operationalize once both those aspects are taken into account – which is what the next two chapters do.

In such a perspective, the complexity of any given content depends on what or who processes (or creates) it. As a crude example, computers were able to win chess games against the best human chess players starting in the 1980s (Deep Blue defeated the world chess champion Garry Kasparov in 1997– yet, at that time, robots were also unable to even move one piece on a checkboard.

In the present cases, humans are the population culture has to run on, and human visual perception is well-documented enough to provide us with reliable proxies for estimating human-specific complexity (see Donderi, 2006 for a review of the historical development of measures of visual complexity). It is possible to quantify complexity in some specific human cultural productions, i.e., graphic communication systems (heraldry, writing systems). The measures used in the next two chapters have been linked to a variety of behavioural correlates, as, for instance, how easy they are to identify, even in noisy visual environments (Pelli, Burns, Farell, & Moore-Page, 2006).

Figure 1. Those two pictures would have, in most metrics, the same measures of complexity, yet the left one is much easier to process for humans, based on its resemblance with upright human faces: the position of the three squares in the left picture (two above, one below) matches the relative positions of the eyes and the mouth in an upright human face. This is not the case of the picture on the right, although it is the horizontal symmetry of the left figure, and hence it is harder to process.

2. Determinants of visual complexity in graphic communication systems

Chapters 4 and 5 both focus on what determines how visually complex are some cultural types. They respectively focus on motifs that are used on coats of arms, and characters that are used in scripts. Higher complexity makes, ceteris paribus, contents (in both cases – i.e., motifs and characters) harder to process – but in both cases, visual complexity is best predicted by trade-offs between different factors or evolutionary pressures. In doing so, they follow dynamics that are quite common in language evolution, such as trade-off between simplicity (pressure to be easily learned and communicated - see Chater & Vitányi, 2003; Culbertson & Kirby, 2016) and informativeness (pressure to refer to a specific meaning, i.e., for high informative content, Kirby, Tamariz, Cornish, & Smith, 2015; Piantadosi, Tily, & Gibson, 2012); see also Kemp, Xu, & Regier (2018) and Winters, Kirby, & Smith, (2018) for consequences of this trade-off . This way to integrate different factors in predicting and testing what influences the complexity of a set of signals also matches the logic sketched in the introduction.

Chapter 4 aims at testing whether one type of relation between frequency and complexity holds for heraldic motifs – Zipf's law of Abbreviation. Zipf's law of Abbreviation is a form of optimal pairing between signals' length (words or vocalisations' lengths, or, here, motifs' complexity) and frequencies. It minimizes the overall cost of communication by using shorter signals for rare meanings. Complexity, especially before the invention of printing, makes motifs more costly both to produce and to process. Iconicity, on the other hand, is a way to minimize processing costs. At equal measures of complexity, iconic motifs are easier to process than their non-iconic counterparts. Explaining the relationship between frequencies and complexity of motifs thus requires taking into account the influence different pairings and properties of motifs have on the communicative and cognitive costs of the system.

Chapter 5 partially replicates previous studies regarding the importance of script size (or inventory length, the number of characters included in a script). The more characters in a script, the more complex the characters tend to be, in line with Chang, Chen, & Perfetti (2018). The type of script (i.e., the linguistic unit represented by characters) and controls for the influence of parent scripts are both important predictor of character complexity. Here too, the visual complexity of a cultural type (characters) are best explained by a combination of factors, rather than by only one bias.

3. Dealing with culture's messiness: the variation we want to consider, and variation we want to eliminate

Another major difficulty arising from using real-world cultural data (i.e., large archival datasets, curated with methods inspired from corpus linguistics and quantitative history) is that it can include a non-negligible amount of noise. This noise can originate, for instance, from variation in the original materials, including in the way it was preserved (for instance, differences in processes used to scan or photograph artefacts).

One main challenge thus resides in curating datasets that are standardised enough for testing predictions on it, and yet retain all the relevant variation on the relevant dimensions from the original materials. Methods of chapters 2 and 3 illustrate this trade-off and how to deal with it.

Obviously, the variation that we chose to eliminate depends on what our measures are sensitive to. Here, measures of complexity tend to be sensitive to variation in size and in line thickness.

Chapter 2, on heraldry, starts with already fairly standardized material (i.e., a comprehensive collection of more than 100 000 engravings compiled by one two-authors team), in which all coats of arms were represented at the same size and drawn in the same style, with very similar line thicknesses. Thus, most of the treatment applied to the data was oriented towards simply collecting it and optimizing its quality (scanning, slight resizing and editing, Potrace algorithm, Selinger, 2003).

In chapter 3, the different fonts used for the different writing systems induced a much heavier variation on both the size of the graphemes and the thickness of their lines, although we used the same initial font size for all systems. Standardization was achieved through adaptive resizing (i.e., decreasing the variation in size between systems, while maintaining variation within systems) and through the use of a thinning algorithm to derive each grapheme's approximate medial axis - a form of skeleton with consistent line thickness.

Chapter 2: When iconicity stands in the way of abbreviation:

1. Introduction

Zipf's Law of Abbreviation

George K. Zipf's name is linked to two phenomena: the power law distribution of word frequencies, and the correlation that he observed between word lengths and word frequencies — often referred to as the « Law of Abbreviation », or « Brevity ». It states that shorter words tend to be more frequent than longer ones (Zipf, 1949). Zipf's Law of Abbreviation (ZLA) has been documented in various communication systems, both human and non-human. In the animal kingdom, a negative relation between signal length and frequency of use has been found, for example, in dolphins (Ferrer-i-Cancho & Lusseau, 2009), formosan macaques (Semple, Hsu, & Agoramoorthy, 2010), bats (Luo et al., 2013) and, to some extent, common marmosets (Ferrer-i-Cancho & Hernández-Fernández, 2013). Among humans, several empirical studies have verified Zipf's Law of Abbreviation with both spoken and written communication systems. A ZLA obtains for all the spoken human languages for which it has been tested. A ZLA for phonological word length obtains in American English, Croatian, Greek, Indonesian, Russian, Spanish and Swedish (Ferrer-i-Cancho & Hernández-Fernández, 2013). Other studies, using number of phonemes as a proxy for word length, also found a ZLA in Dutch, English, German and Swedish (Piantadosi, Tily, & Gibson, 2011; Sigurd, Eeg-Olofsson, & Weijer, 2004). The same result holds when orthographic word length (for alphabetically written languages) is used as a proxy for word length, as evidenced by studies based on more than a dozen languages (Piantadosi et al., 2011; Sigurd et al., 2004; Strauss, Grzybek, & Altmann, 2006) and one based on 986 languages (Bentz & Ferrer-i-Cancho, 2016). Although no conclusive argument has proven Zipf's Law of Abbreviation to be universal, it is certainly rather ubiquitous.

Zipf's original account suggests that this law of abbreviation results from a trade-off between a pressure for efficiency (favouring shorter forms) and a pressure for communication accuracy (favouring redundancy and unique, longer, forms). In this account, an optimal solution is a form of variable-length coding (similar to Huffman coding, Huffman, 1952) which assigns shorter words to more frequent meanings, and longer words to less frequent meanings. This type of coding would thus optimize the production cost of communication. Since frequently employed words or vocalizations overwhelmingly tend to be less informative than more frequent ones (Piantadosi et al., 2011), Zipf's Law of Abbreviation makes communication more efficient, by calibrating the amount of signal information that a receiver needs to process (e.g., the length of a word), to the quantity of information contained in the signal (e.g., a word's predictability). In this account, a « Principle of Least Effort » (Zipf, 1949) is to be understood as the functional explanation underlying the negative relation between words' lengths and their frequencies.

Both processing and production costs may cause the Law of Abbreviation

The exact causes of Zipf's Law of Abbreviation remain unclear, due to a persistent ambiguity in the notion of communication efficiency. On the emitter's side, efficiency refers to the effort spent on producing a signal; on the receiver's side, it relates to the costs of processing a signal. Production costs and processing costs are tightly correlated: long words tend to be effortful both to produce and to process. Yet, as sociolinguists have argued, processing effort is unlikely to be perfectly aligned with production effort, for two reasons at least (Trudgill, 2011; Winters, Kirby, & Smith, 2018b; Wray & Grace, 2007). First, emitter and receiver may not be motivated to communicate to the same degree. In some situations (compare, for instance, a mumbled confession to a security warning communicated loudly and clearly to distracted passengers), speakers do not care as much about being understood as listeners do: speakers have an incentive in reducing their production effort at the expense of the hearer's processing effort. Second, there are situations where context provides information that does not need to be linguistically encoded with precision. Here again emitters may reduce their production effort, this time without a corresponding increase in processing cost on the receiver's side, since missing information can be inferred from contextual cues.

This opens the way for at least two distinct interpretations of Zipf's Law of Abbreviation, depending on what one considers to be driving it. In one version, frequent words are shortened to make them more efficient to process, in the other, shortening facilitates the processing of frequent words. Although both versions result in tightly overlapping predictions, they are not impossible to tease apart. Studies addressing this issue (Cohen Priva, 2017; Piantadosi et al., 2011) show that a word's information value (its likelihood of appearing given the verbal contexts where it occurs) is a better predictor of word length than is word frequency (which is strongly but not perfectly correlated with information value). These studies are consistent with an interpretation of ZLA where abbreviation is driven by processing costs, rather than production costs, since a word's information value affects the hearer's capacity to anticipate it, but not the costs of producing it. In most studies, however, the exact roles played by processing *versus* production costs in the Law of Abbreviation are not teased apart.

The Law of Abbreviation in graphic codes

This uncertainty on the exact roles played by processing versus production costs makes graphic symbols particularly relevant to the study of Zipf's Law of Abbreviation (Rovenchak, 2011; Rovenchak, Mačutek, & Riley, 2009; Shu, Chen, Anderson, Wu, & Xuan, 2003; Tamaoka & Kiyama, 2013; Zhang, Zhang, Xue, Liu, & Yu, 2007). Graphic symbols like written letters or emblems consist of visual marks inscribed on an enduring support (unlike the gestures of sign languages, Morin, Kelly, & Winters, 2019). The balance of processing and production costs is arguably quite different for graphic symbols as distinct from spoken words or gestures. Graphic symbols can be produced once and be seen many times, in contrast with spoken words, which need to be produced every time they are heard (exception being made for recent recording technologies of no relevance to language evolution). Techniques of mechanical reproduction, from seal impressions to printing, bring down production costs even further. Additionally, visual processing is intrinsically more efficient than phonological processing (Cohen, Horowitz, & Wolfe, 2009). Graphic symbols, contrary to auditory signals, do not require their recipients to process them on the fly and on line, which could limit the impact of an increase in processing costs.

Testing for ZLA in a corpus of graphic symbols requires finding some graphic equivalent for the length of vocal signals. Image complexity is similar to the length of vocalizations in one key respect: complex images are harder and more costly to produce and to process. Longer words (above 7 letters) require longer reaction times to be recognized out of context (Barca, Burani, & Arduino, 2002; New, Ferrand, Pallier, & Brysbaert, 2006). Similarly, more complex images take longer to be identified, and also occasion more mistakes (Byrne, 1993; Donderi & McFadden, 2005; Pelli, Burns, Farell, & Moore-Page, 2006; Zhang et al., 2007). This effect of complexity is robust to participants' familiarity or experience with the images (Byrne, 1993), and to levels of noise, overall contrast, or eccentricity in the visual field (Shu et al., 2003). More complex shapes, like longer vocal signals, both require higher cognitive costs to be processed than their simpler or shorter analogues. Following a Zipfian logic, any communication system, vocal or graphic, should minimize its aggregate costs by reserving long or complex forms for infrequent symbols. This predicts that visual complexity would be lower for more frequent graphic symbols, in the very same way that more frequent signals tend to be shorter in other communication systems (Rovenchak et al., 2009).

Case studies have documented such distributions in a particular type of graphic communication system: writing systems. Consider as an example the visual complexity of logographic Chinese characters for Mandarin. A proxy for complexity, in this case, is provided by the number of distinct strokes that a character contains: 一 (pinyin *yi*, one) has fewer strokes than 五 (pinyin *wu*, five)**,** thus it is less complex. Frequently used Chinese characters tend to be simpler, consistent with Zipf's Law of Abbreviation (Shu et al., 2003). Unlike alphabetically written words, the complexity of Chinese characters is uncorrelated with the length of the morpheme they represent (which is one syllable-long, with rare exceptions): the "law of abbreviation" observed for Chinese characters thus cannot be due to the length of the underlying vocalizations. The same argument can be made for Chinese characters as used within the Japanese writing system (*kanji*): here again a "law of abbreviation" is observed (Tamaoka & Kiyama, 2013). Finally, it is also observed for large-size syllabaries or alphabets (Rovenchak, 2011; Rovenchak et al., 2009). These writing systems (at least Japanese, Chinese and Vai) have made extensive use of printing, showing that ZLA may obtain for signals with relatively weak production costs.

European heraldry

An equivalent of Zipf's Law of Abbreviation for graphic symbols thus looks plausible on theoretical and empirical grounds. We turned to European heraldry to test it. The coats of arms (hereafter simply "arms") used by notable European families since the late Middle Ages provide us with a corpus of graphic symbols that is abundantly and accurately documented over several centuries. Arms were versatile symbols. They could come in all sorts of sizes and on any and all kinds of support, from painted banners to impressed seals, from hand-drawn armorials to wrought-iron door knockers. Their uses ranged from the ostentatious (e.g., in tournaments, on monuments) to the mundane (e.g. as marks of property) (Fox-Davies, 1900; Pastoureau, 2007). The most important sources are, for the medieval period, painted armorials and engraved seals, joined for later periods by printed armorials and ex-libris plates. Heraldic emblems were created by combining motifs from a standardized repertoire of motifs that shows great stability across time and space (Fox-Davies, 1900; Pastoureau, 2007). Heraldic arms, thus, are ideally suited to a computational treatment: the appearance of motifs on the coats of arms of individual families can be estimated with precision, as well as the occurrences of motif combinations (Morin & Miton, 2018). In this respect heraldic motifs resemble the written words of a well-documented script.

In addition to the abundance of high-quality data, our decision to study heraldry was justified by several notable analogies and disanalogies between heraldic emblems and linguistic symbols (written or spoken), which would make the obtention of a ZLA anything but trivial — a strong confirmation of this phenomenon's apparent universality.

Why heraldry may be Zipfian

Ubiquity of Zipf's Law of Abbreviation

Zipf's Law of Abbreviation's quasi-universality is the first reason we would expect it to apply to heraldic motifs: as developed above, Zipf's law of abbreviation can be found in a large variety of communication systems, both for oral and written signals. The basic mechanisms that cause ZLA in spoken words and graphemes appear to be present in heraldry: symbols were produced to encode information — in this case, to identify a coat of arms as belonging to a given family—, at a non-trivial cost to the producers. The information conveyed by heraldic emblems could be more or less ambiguous, and makers of arms strove to maximize the distinctiveness of the emblems they designed (Morin & Miton, 2018). Although spoken and written communication systems (including heraldry) differ on specific properties, pairing shorter or simpler signals with higher frequencies and longer or more complex signals with rarer frequencies is an optimal solution, both in terms of minimizing the production effort (Least Effort interpretation of ZLA), and in terms of maximizing the informativeness in relation to the processing cost for receivers. Visually complex emblems, like complex letters and longer spoken words, are more costly to produce and process than their simpler or shorter analogues.

The appeal of simple motifs

In addition to this crucial pressure for distinctiveness, heraldic emblems were also required to be æsthetically pleasing. The search for æsthetic appeal may push down the complexity of the most popular heraldic designs, due to the well-attested link between the ease of processing visual stimuli and their perceived beauty (Reber, Schwarz, & Winkielman, 2004; Reber, Winkielman, & Schwarz, 1998). Shapes are seen as more appealing when they are easier to process in a variety of experiments that manipulate parameters with known links to visual complexity, such as asymmetry or noisiness. Visual complexity directly decreases the ease of processing a visual stimulus (Pelli et al., 2006). This possible link between a symbol's success, its æsthetic appeal, and its visual simplicity, was a good reason to study heraldic emblems, since it is unlikely to obtain in other graphic codes, such as writing systems (where a letter's frequency is chiefly driven by the frequency of the morphemes or phonemes that it stands for).

Heraldry also included, from its origins onwards, motifs varying in complexity from relatively simple forms to relatively complex ones. Although some very simple motifs (e.g., a pale) were deemed 'honourable' and reserved to the oldest noble families (Fox-Davies, 1900, p.), such motifs were relatively freely adopted. Such an association of simple forms with prestige, if anything, should favour the success of simpler motifs and hinder the diffusion of more complex motifs, thus predicting to the emergence of a ZLA.

Why heraldry may not be Zipfian

Production costs

Heraldic motifs also differ from linguistic symbols in ways that suggest ZLA may not obtain in their case. First, there are reasons to believe that production costs were particularly low for heraldic emblems as compared to writing. They were frequently used for public display, where a symbol is produced once to be seen many times. Written characters, as used in personal correspondence or regular account keeping, must be inscribed repeatedly and rapidly. Production costs were also dramatically reduced by techniques of mechanical reproduction. These include printing, but also (and arguably, more importantly) seal impressions, thanks to which one heraldic emblem could be engraved once and impressed hundreds of times. Both techniques were used for written letters as well, but at least as far as seals were concerned heraldry depended upon mechanical reproduction to a greater extent. As a consequence, one could afford to produce heraldic emblems slowly and painstakingly, while written symbols in most contexts had to be drawn in fast and effortless ways. Another consequence was an increased division of labour: relatively fewer people were involved in the production of heraldic arms, compared to written symbols.

Iconicity

The third major difference between heraldry and writing lies in the fact that heraldry makes a sharp distinction between iconic and non-iconic motifs, and includes both types of images. Iconicity is defined as a salient perceptual or structural resemblance between sign and object (Peirce, 1974). Iconic (or *concrete*) symbols are assumed to be *visually obvious*: they successfully figure real-world plants, animals, persons, or objects, in ways that are immediately transparent for an unacquainted viewer (Y. Rogers, 1989). In contrast, abstract symbols represent information using graphical features that have no obvious relation to what they represent. Both types of motifs figured on arms that stood for lineages, the mapping between arms and lineages being arbitrary most of the time. In this sense, both could be called "iconic" in the technical, Peircean sense that there was no resemblance between a symbol (the arms) and its referent (the lineage) (Jappy, 2013). Yet the motifs that we call iconic differ from the non-iconic in that they directly depict a real-world object (a lion, a cup, a knight, etc.), independently of their heraldic meaning(s). A wealth of arguments supports the idea that abstract graphic symbols evolve from earlier (more) iconic depictions, including both semiotic experiments (Caldwell & Smith, 2012; Garrod, Fay, Lee, Oberlander, & MacLeod, 2007a) and observations on writing systems (Schmandt-Besserat, 2010). For instance, although many of them begun as iconic signs, the figurative meaning of most Chinese character keys is either lost or beyond the uninitiated's grasp (Xiao & Treiman, 2012). By contrast, the rules of heraldic composition differentiate two categories of motifs. The "charges" (e.g. a lion, an eagle, a castle, etc.) can be placed anywhere on a coat of arms and they are overwhelmingly iconic; the "ordinaries" (e.g. a bend, a chevron, etc.) are not, and their location is constrained in various ways (Rietstap, 1884). This distinction allows us to separate iconic from non-iconic motifs using the categories given by our sources. In the rest of the paper we simply refer to charges as "iconic motifs", and to ordinaries as "non-iconic motifs" (see Methods and S1 File in Appendix B for more detail).

Iconicity may prevent the emergence of a ZLA. In order to successfully represent their real-world referent, concrete symbols include details enhancing the similarity between the symbols (or drawings) and the objects they are representing: for instance, symbols for bears would depict fur and other characteristics of actual bears. Concrete symbols have been found to be more complex than abstract symbols over a range of studies (García, Badre, & Stasko, 1994 analysing data from Arend, Muthig, & Wandmacher, 1987; Yvonne Rogers, 1986; Rohr & Keppel, 1984; Stammers & Hoffman, 1991). Concrete symbols also enjoy performance advantages over abstract symbols (Green & Barnard, 1990; Yvonne Rogers & Oborne, 1987; Stammers & Hoffman, 1991) in spite of their greater complexity: they are easier to recognize. For these two reasons, finding a graphic equivalent of Zipf's Law of Abbreviation for heraldic motifs in addition to writing systems would be a strong indication of its universality.

Cultural diffusion and the law of abbreviation

One last reason to study heraldry has to do with possible links between cultural diffusion and Zipf's Law of Abbreviation. A written character's frequency of use largely depends on the frequency of use of the underlying morpheme or phoneme. Heraldic motifs, in contrast, do not generally show such a dependency of their frequency on the frequency of what they represent. The frequency of the word "hedgehog" has much to do with the frequency of hedgehogs in the environment, but there is no equally direct link between the frequency of the hedgehog motif in heraldry and the frequency of hedgehogs, or that of families named "Hedgehog". In other words, the frequency of heraldic motifs cannot generally be said to follow from semantic constraints, unlike the frequency of words. Rather, their frequency of occurrence reflects cultural diffusion, i.e., the selective borrowing of motifs resulting in their spread. The role cultural diffusion might play in ZLA is, at the moment, under-explored. Does frequency of use cause signals to become simpler, or on the contrary, do simple signals find more users? Both hypotheses are plausible. They are also not mutually exclusive. Use, especially in interactive contexts, tends to produce ZLA distributions of label lengths (Kanwal, Smith, Culbertson, & Kirby, 2017), while on the other hand, a signal's brevity can lead such a signal's frequency to increase, as the worldwide success of the word « OK » can attest (Metcalf, 2010).

Heraldry enables us to explore the impact of cultural diffusion in the long run, with two large corpora of European arms, one gathering arms dating from the late Middle Ages (c. 1200–1500) ("Clemmensen" (Clemmensen, 2017)), the other gathering arms from the early modern period (c. 1600–1850) ("Renesse" (Renesse, 1894; Rietstap, 1884)). This allows us to test how the cultural diffusion of motifs may impact a relation between motif complexity and frequency, but also how such a relation may evolve over time. In addition, the two corpus capture two different states of European heraldry, whose history can be seen as leading from relatively simple arms to more complex ones, partly (but not only, see below) because the simplest motifs were thought to be the preserve of the most ancient families, who chose their arms before others did (Renesse, 1894).

Goals of the study

The present study aims to (1) test for a ZLA in heraldic motifs over two corpora, one medieval, one early modern; (2) investigate whether—and to what extent—iconicity interacts with ZLA, and (3) explore the impact of cultural diffusion upon these dynamics.

The complexity of graphic symbols can be measured in various ways. Of the multiple definitions that have been proposed for what makes a shape simple or complex (see Donderi, 2006 for a review), we focus on two measures from two distinct research traditions. The first one, Descriptive Complexity (DC), is based on Algorithmic Information Theory and uses the length of code required to store an image in optimally compressed form as a proxy for the image's complexity. Here, it is obtained by compressing the picture files and using the size of the compressed file (in bytes). The second one, Perimetric Complexity (PC), starts instead from the image's physical features. Some such measurements consider, for instance, the number of angles or edges in images, or their ratio (Mavrides & Brown, 1969; H. Thomas, 1968). Here we consider, instead, the image's contour length compared to its inked surface (Arnoult, 1960; Pelli et al., 2006) (see Materials and methods). Both these proxies for image complexity correlate negatively with ease of processing and performance for an array of tasks (see Pelli et al., 2006 for PC, Donderi & McFadden, 2005 for DC). In order to have reliable estimates of heraldic motifs' complexity, we used a compendium of more than 100 000 illustrated arms (Rolland & Rolland, 1969), from which we extracted three images for each of the motifs present in one or both of our corpora (1). All images used in the present study are available on the OSF depository associated to the project (https://osf.io/ykp37/).

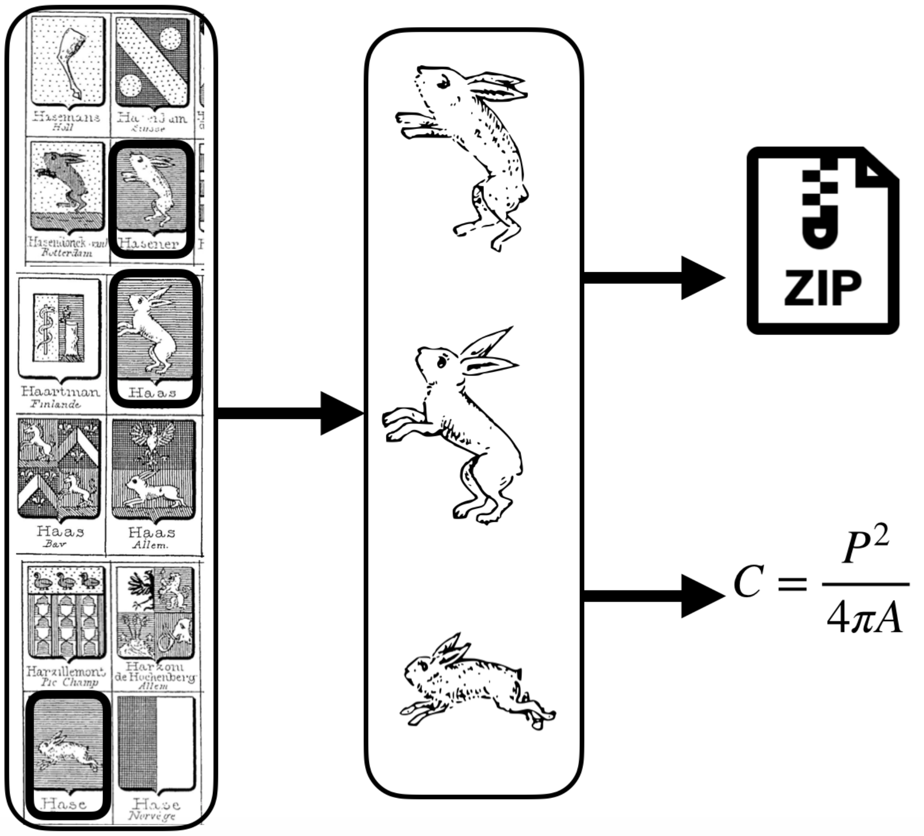


Figure 1. Visual complexity metrics derivation. Our complexity measures were obtained in the following way: (1) three arms were selected for each motif from our reference armorial (Rolland, 1909), (2) pictures of shields were edited to obtain a picture of the motif on its own, (3) the edited pictures went through the Potrace algorithm to improve their quality (through vectorization), (4) they were zipped and the zip file's size served as our measure of descriptive complexity, while their perimetric complexity was calculated using Mathematica. Finally, all the three measures from the three pictures obtained for each motif were averaged to get one reliable measure for each type of complexity for each motif in our inventory.

2. Materials and methods

In order to be able to test our predictions, we compiled (1) a list of motifs whose frequency and complexity we measure, (2) pictures to reliably measure the motifs' complexity, and (3) frequencies (i.e., number of occurrences) for such motifs, on two corpora corresponding to two different time points: the Clemmensen (c. 1200–1500) and Renesse (c. 1600–1850) corpora—see Table 1 for more details on our sources and how they relate to each other.

Pre-registrations

We kept a complete research diary on the Open Science Framework (https://osf.io/euck2/) where all analyses carried out were pre-registered and described. Pre-registration is an open research practice that consists in describing the research design and analysis plan as independently as possible from data collection (Munafò et al., 2017). The methods and analyses of this paper were pre-registered (recorded) in several waves.

Sources

Our primary materials were Renesse's *Dictionnaire des figures héraldiques* (Renesse, 1894), the *Armoiries des familles contenues dans l'Armorial Général de J.B. Rietstap,* by Victor and Henri Rolland, and Steen Clemmensen's Armorial (Clemmensen, 2017, armorial.dk). Renesse (Renesse, 1894) provides a motif-by-motif index of over 100 000 arms, indexing Rietstap's *Armorial Général* (Rietstap, 1884)*,* while the Rollands' compendium of arms (Rolland & Rolland, 1969) provides illustrations for over two thirds of those. Renesse provided a classification of motifs, which was used for both corpora, and frequency data for the Renesse corpus. Rolland provided the pictures of motifs we needed for our visual complexity measures (for both corpora). Finally, Clemmensen's armorial provided us with frequency data for the Clemmensen corpus.

Source	Relation to other sources	Format	Information extracted
Rietstap	indexed by Renesse, illustrated by Rolland	Armorial (list of branch's names and descriptions of their arms)	None directly
Renesse	indexes and classifies Rietstap's armorial	Dictionary (list of branches organized by which motifs they bear)	Frequencies Motifs' inventory (classification)
Rolland	illustrates Rietstap	Compendium (tables of illustrations)	Pictures of motifs
Clemmensen	None	Armorial (list of branch's names and descriptions of their arms)	Frequencies

Table 1. Our sources, how they relate to each other, their format, and which information was used from each one. While Renesse provides frequency data for the early modern period, Clemmensen provides it for the late Middle Ages.

Inventory constitution

A list of motifs corresponding to Renesse's classification was built taking Renesse's own subdivisions of his material as guide. Two aspects of motifs that were relevant for the author were not taken into account for classification: the orientation of a motif (i.e. whether the same motif is presented facing the left side or the right side of the arms), and the number of times that it is repeated. In other respects, we stuck as close as possible to Renesse's own descriptions. All further details and steps of sample constitution are reported in S1 File. Information on frequency and information on complexity were collected independently—i.e., the researcher and research assistants who collected the frequency data did not observe the visual complexity data, and vice versa.

Our classification of motifs between iconic and non-iconic motifs was directly built on Renesse's inventory. Following a long-established taxonomy, his inventory makes a sharp distinction between certain categories: "charges", which are any image that can be placed anywhere on the arms, and "ordinaries", which includes both "pièces", whose placement is constrained by rules, and "partitions", which are divisions of the arms. Ordinaries are abstract, geometric shapes that do not represent a natural object in any detail (e.g., saltires, bends, lozenges). The subset of motifs they represent is referred to as non-iconic. By contrast, charges are essentially figurative motifs, representing mainly animals, plants and various artefacts, and the subset they represent are referred to as iconic.

Visual complexity

All the complexity measures were taken as the average of three arms (i.e., three image files), selected among a standardized collection of thousands of drawings (Rolland, 1909; Rolland & Rolland, 1969)—see S1 File for details on arms selection and image files preparation. Using three pictures for each motif allowed us to have robust estimates of the motifs' complexity that would not depend on the specific picture chosen for each motif while still allowing to include a large number of motifs. We used two measures of complexity: perimetric complexity and descriptive (also known as algorithmic) complexity. Both complexity measures have previously been used in experimental investigations of cultural evolution (Tamariz & Kirby, 2015).

Descriptive complexity measures are obtained using the Potrace algorithm (Selinger, 2003) on the .pnm files, and then compressing the obtained .eps file. The proxy for descriptive complexity is then the size in bytes of the compressed file: it offers an estimation of the length of the shortest computer program that (losslessly) produces the image. This measure of descriptive complexity is identical to the one used by Tamariz & Kirby (2015) under the label *algorithmic complexity*. It is to be conceived of as an upper bound of a picture's complexity, as (1) it adds header information—which was kept minimal using the same folder for all pictures, and standardized file names of the same length, (2) it only searches for a small set of simple patterns and patterns in particular block lengths, and (3) it is not a mode of compression optimized for images per se.

We measured perimetric complexity (Pelli et al., 2006; Watson, 2012), defined as a ratio of inked surface to the perimeter of this inked surface. It is obtained, using Watson's implementation (Watson, 2012), by taking the squared length of the inside and outside perimeters of a motif P, divided by the foreground area A and by 4π, i.e.: $PC = \frac{P^2}{4\pi A}$. The measure was implemented in Wolfram (Mathematica), and applied after the pictures were processed using the Potrace algorithm.

As stated in the pre-registration documents, and in order to avoid motifs whose complexity measures would be unreliable (because of excessive variation in their depiction), we set a threshold over which motifs' occurrences were too variable to be comparable, such that motifs for which our set of three pictures had a standard deviations higher than this threshold were excluded from subsequent analyses. This threshold was pre-registered, and applied to both measures of complexity. It is defined as two standard deviations above the mean of standard deviations (calculated for each motif on the basis of three pictures, see Equation 1)

Equation (1) $t = M\big(SD(d)\big) + 2 * SD(SD(d))$

with d being the distribution of complexity scores in our dataset of motifs.

Applying this exclusion criterion did not change our results: the results obtained without applying the exclusion criteria are available in S1 File, and are very similar to the ones reported in the main text.

Frequency measures

A motif's frequency refers to the number of arms bearing the given motif, among all arms bearing only one motif in each of our corpora. Thus, we only consider motifs occurring alone, i.e., we counted the number of arms bearing the motifs of interest and nothing else. This allows us to have (1) exhaustive counts for both corpora, which (2) are associated with representative visual complexity measures: our visual complexity measures are taken on motifs occurring alone on arms, and so are our frequency measures. We thus avoid biasing our frequency or complexity measures by having either of them include arms in which motifs appear in combination with other motifs. We used two corpora, one made from medieval armorials and covering mostly the period 1200 to 1500, based on the work of Steen Clemmensen (Clemmensen, 2017) (here called "Clemmensen"), and another constituted by us from J.B. Rietstap's armorial (Rietstap, 1884) as indexed by T. de Renesse's dictionary (Renesse, 1894, here called "Renesse"). That second corpus covers a longer period, until 1880, although most of the arms that are dated occur between 1600 and 1850. Both corpora concern themselves chiefly with the arms of families and individuals, with little to no coverage of civic heraldry, and cover a wide range of European territories. We do not know to what extent the two corpora overlap: some arms are likely to be present in both. In other respects the corpora differ widely, and do not provide the same metadata. Although they do not classify heraldic motifs in identical ways, and have different ways of counting arms and families, our inventory of motifs was applicable to both. Details on how frequency measures were obtained for each dataset are available in S1 File.

Statistical analyses

None of the measures we analysed (for both Clemmensen and Renesse datasets, and all variables, i.e., frequencies and both measures of complexity) were normally distributed (all ps < .01 on Shapiro-Wilk tests). Hence, all statistical tests presented here are non-parametric (Kendall rank correlation tests, because of the presence of ties, which leads to inexact p-values in Spearman's rank correlation test). All analyses were run in R (R Core Team, 2018).

3. Results

Correlation between measures of complexity

Based on 447 motifs, our two measures of visual complexity, descriptive and perimetric, were highly correlated, $r_\tau = .69$, $p < .001$, 95% CI [0.657, 0.725]. See Figure 2 for illustrations of heraldic motifs of different visual complexity.

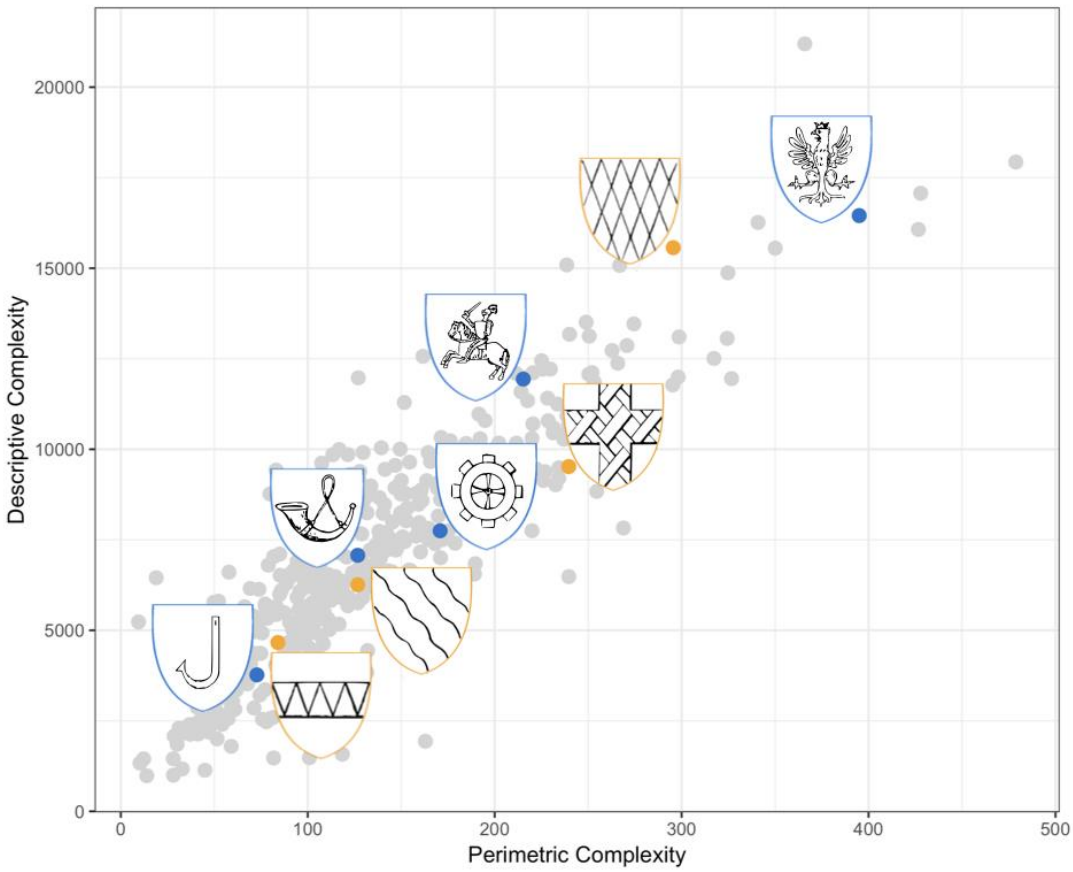


Figure 2. Relation between perimetric (x-axis) and descriptive (y-axis) complexity, with examples of iconic (blue frame) and non-iconic (yellow frame) motifs (total n = 447 motifs).

Iconic motifs are more complex than non-iconic motifs

Previous experimental studies have interpreted decreases in complexity as indicative of a loss of iconicity (Garrod et al., 2007a). Our study confirms that iconicity indeed tended to be associated with higher complexity: in our sample, iconic motifs (*N* = 295) had higher complexity than non-iconic motifs (N = 152), both perimetric (U = 32160, $p < .01$) and descriptive (U = 36110, $p < .01$), see Figure 3.

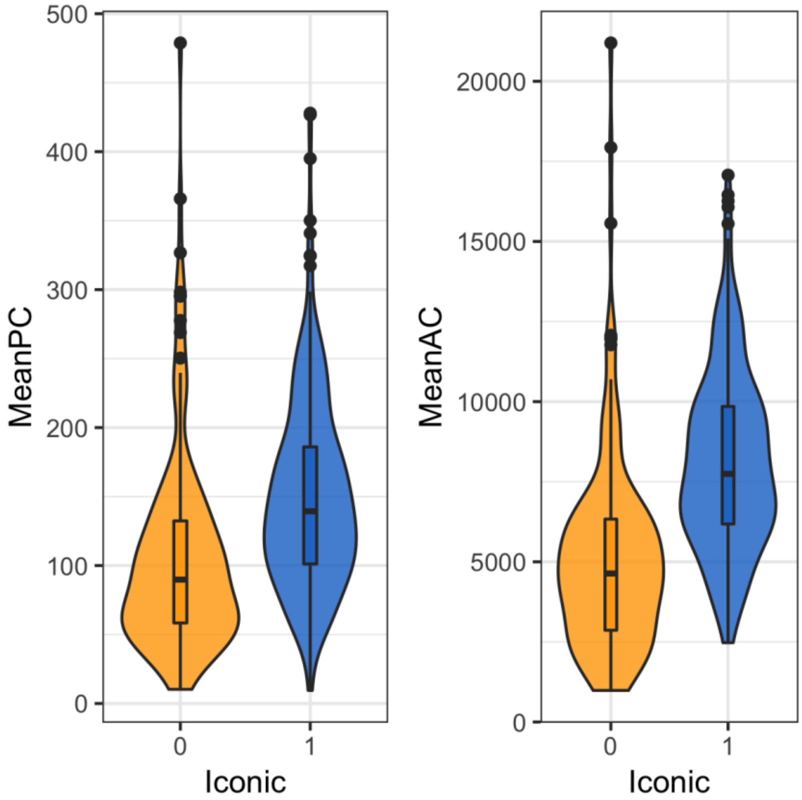


Figure 3. Distribution of iconic (blue) and non-iconic (yellow) motifs' complexity scores, perimetric (left) and descriptive (right).

No overall Zipf's law of abbreviation

For the Clemmensen corpus (296 motifs, total N = 8124 arms), we failed to conclusively observe a Zipfian correlation (see Figure 4). On the one hand, more frequent motifs also tended to be less complex when measured by descriptive complexity ($r_\tau = -.09$, $p = .018$, 95% CI [-0.177, -0.011]). On the other hand, there was no significant correlation between perimetric complexity and frequency ($r_\tau = -.02$, $p = .633$, 95% CI [-0.103, 0.065]).

Over the Renesse corpus (447 motifs, total N = 16991 arms), contrary to what ZLA predicts, more frequent motifs were also visually more complex. More precisely, there was a weak correlation between frequency and descriptive complexity (as evidenced by a Kendall rank correlation between frequency and descriptive complexity $r_\tau = .08$, $p = .008$, 95% CI [0.02, 0.149]), and a stronger relationship between frequency and perimetric complexity $r_\tau = .12$, $p < .001$, 95% CI [0.058, 0.186].

No relation between complexity and frequency for non-iconic motifs

On the Clemmensen corpus, non-iconic motifs (114 motifs, N = 4795 arms) tended to be more frequent when they were less complex, similarly to ZLA's predictions, but only when measured by descriptive complexity ($r_\tau = -.14$, $p = .035$, 95% CI [-0.27, -0.001]). This correlation was not observed when measured with perimetric complexity ($r_\tau = -.08$, $p = .233$, 95% CI [-0.211, 0.058]). For the Renesse corpus (152 motifs, N = 6529 arms), there was no correlation between frequency and perimetric complexity ($r_\tau = -.03$, $p = .534$, 95% CI [-0.147, 0.078]), and only a trend for simpler motif to be more frequent, when using descriptive complexity ($r_\tau = -.10$, $p = .07$, 95% CI [-0.211, 0.011]).

Emergence of a 'reverse' ZLA for iconic motifs

Within the Clemmensen corpus, iconic motifs (182 motifs, N = 3329 arms) tended to be more frequent when they were more complex when measured by perimetric complexity ($r_\tau = .12$, $p < .05$, 95% CI [0.021, 0.226]), but not when measured by descriptive complexity ($r_\tau = .08$, $p = .14$, 95% CI [-0.028, 0.18]). By contrast, the Renesse corpus showed clearer results: iconic motifs (295 motifs, N = 10462 arms) showed a positive correlation between complexity and frequency (for both perimetric, $r_\tau = .22$, $p < .001$, 95% CI [0.143, 0.291], and descriptive complexity, $r_\tau = .18$, $p < .001$, 95% CI [0.108, 0.253]).

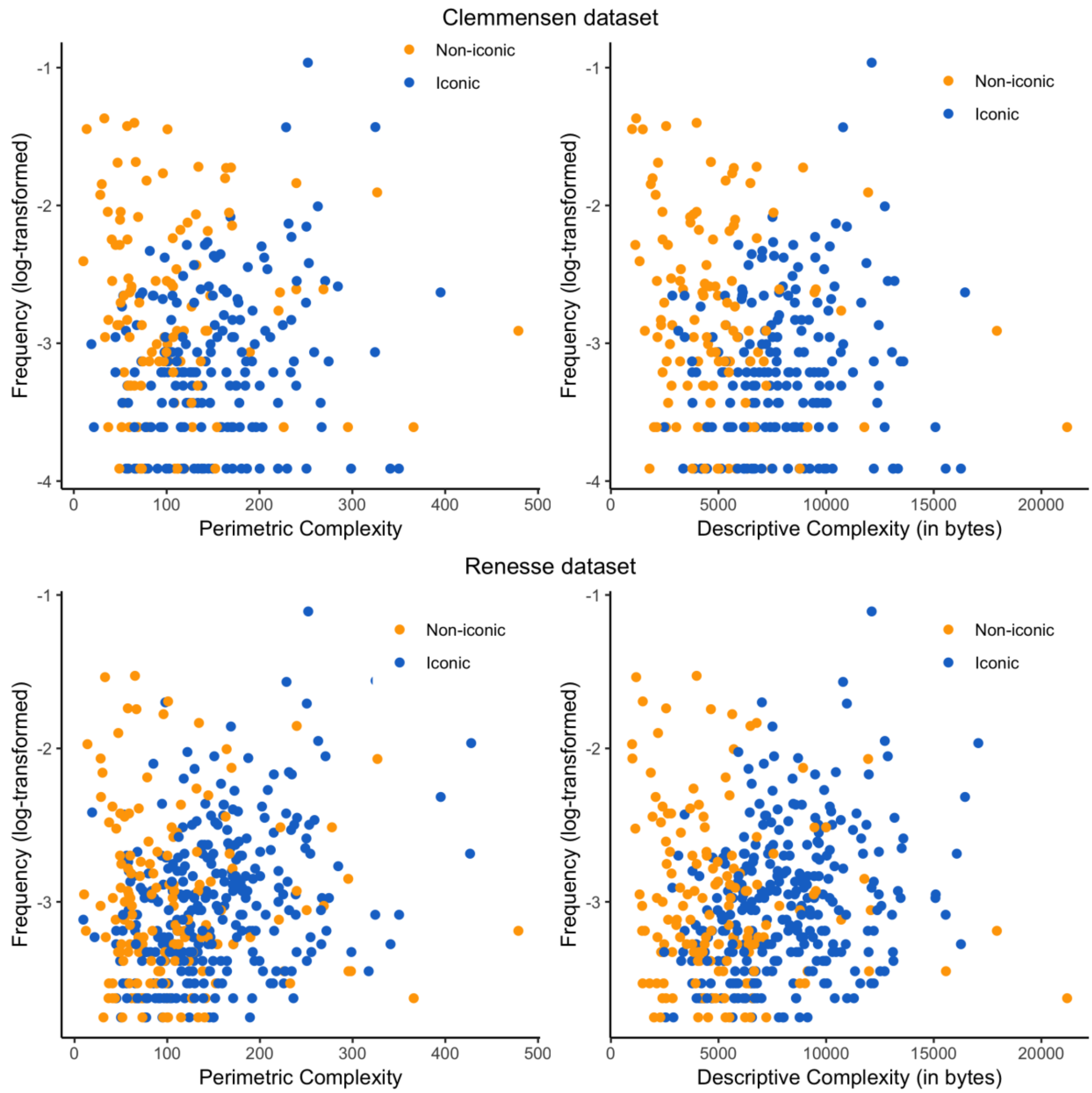


Figure 4. Frequency (log-transformed) as a function of perimetric complexity (left) and descriptive complexity (right), for both corpora (Clemmensen above, Renesse below). Colour indicates iconicity—i.e., iconic motifs are in blue and non-iconic motifs are in orange.

Iconicity and complexity both predict increases in frequency

The proportion of iconic to non-iconic motifs significantly increased between our medieval corpus (Clemmensen) and our early modern corpus (Renesse), from 0.41 to 0.62 (binomial test: $p < .001$). This change in favour of iconic motifs was driven by iconic motifs' frequencies increasing more than that of non-iconic motifs: when comparing the change in frequency for iconic and non-iconic motifs, iconic motifs (Mdn = 0.00035) increased significantly more than non-iconic motifs (Mdn = -0.00057, $U = 16004$, $p < .001$).

The more visually complex a motif was, the more likely its frequency was to increase between the two corpora, the medieval and the early modern, both when using perimetric complexity ($r_\tau = .16$, $p < .001$, 95% CI [0.076, 0.242]), and descriptive complexity ($r_\tau = .22$, $p < .001$, 95% CI [0.144, 0.297]) based on the 296 motifs present in both corpora.

The effect of visual complexity on changes in frequency differed for iconic as compared to non-iconic motifs (see Figure 5). On the subset including only iconic motifs ($n = 182$ motifs), there was a correlation between changes in frequency and perimetric complexity ($r_\tau = .11$, $p = .034$, 95% CI [0.005, 0.206]), but not between changes in frequency and descriptive complexity ($r_\tau = .08$, $p = .108$, 95% CI [-0.011, 0.172]). On the other hand, there was no effect of visual complexity on changes in frequency for non-iconic motifs ($n = 114$ motifs): there was no correlation between changes in frequency and complexity, neither for perimetric complexity ($r_\tau = .03$, $p = .590$, 95% CI [-0.105, 0.173]), nor for descriptive complexity ($r_\tau = .09$, $p = .161$, 95% CI [-0.049, 0.227]) on this subset.

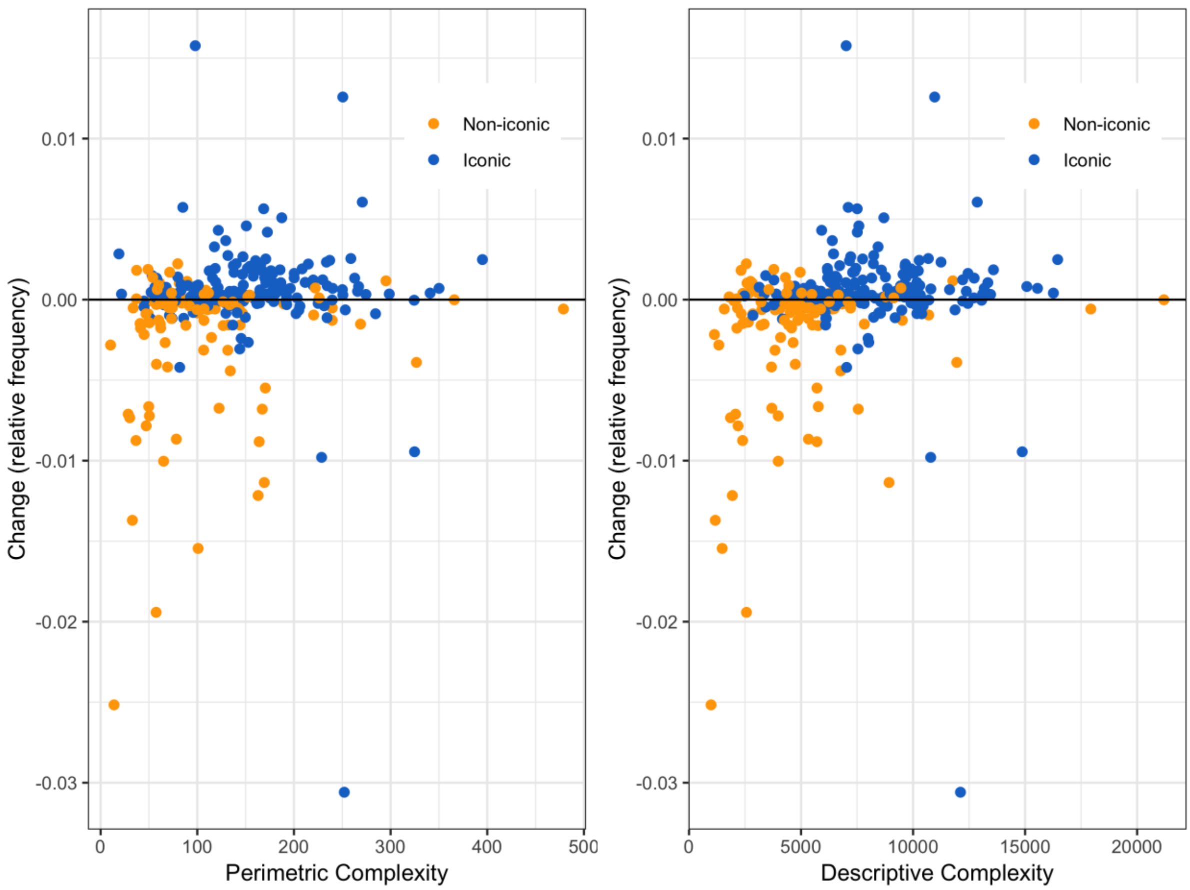


Figure 5. Changes in frequency between our medieval corpus (Clemmensen) and our early modern corpus (Renesse) as a function of perimetric (left) or descriptive complexity (right). Iconic motifs are represented in blue, and non-iconic motifs in orange. The horizontal line (y = 0) indicates no change in relative frequency between Clemmensen and Renesse. Points above the line represent motifs that increased in frequency in our early modern corpus compared to our medieval corpus.

4. Discussion

Our study has three main results. First, we failed to observe a consistent Zipfian Law of Abbreviation in our two corpora. Frequent motifs were simpler in our late medieval corpus, but only weakly so. Motif complexity and frequency showed a robust correlation in our early modern corpus, but it went in the direction opposite to our prediction: frequent motifs were more complex, not less. Second, iconic and non-iconic motifs did not exhibit the same relationship between motif complexity and frequency. More complex iconic motifs were more frequent than less complex ones in both corpora, showing a reverse Zipfian effect, especially for our early modern corpus. By contrast, non-iconic complex motifs did not show this positive effect of complexity on frequency. Third, the present study documented the successful cultural diffusion of iconic motifs over abstract ones: the frequency of complex iconic motifs increased more than the frequency of simpler or non-iconic motifs between our late Middle Ages and early modern corpora. Zipf's Law of Abbreviation, although frequently described as powerful and ubiquitous, fails to obtain for heraldic motifs.

Some previous empirical studies also failed to find Zipfian correlations. In animal behaviour, there have been at least two cases in which communicative behaviour failed to fulfil a ZLA: common ravens (Ferrer-i-Cancho & Hernández-Fernández, 2013, analysing data from Conner, 1985) and golden-backed uakari vocalizations (Bezerra, Souto, Radford, & Jones, 2011). These remain rare examples, confined to non-human animals. Our results differ from these in that they do not merely suggest adding heraldic motifs to the list of cases in which we fail to observe a ZLA: they also show that the law can be reversed, at least for graphic symbols.

We now discuss these results through two major questions. Why does heraldry differ from other systems of symbols? Why do our two corpora differ to such an extent that complexity and iconicity are linked to frequency in the later one, but not the earlier? We consider and discuss four mechanisms that could be relevant to both questions. The first two, processing costs and production costs, plausibly influence ZLA for all types of signals, not just heraldry. The second concerns two factors peculiar to emblems, as distinct from spoken or written language: diffusion dynamics and iconicity.

Processing and production costs

Processing costs might differ between written symbols, that exhibit ZLA, and heraldic emblems, that do not, since most written messages must be read at a faster pace than is required for emblems. This might weaken the pressure to simplify frequent symbols and explain why ZLA does not consistently obtain with heraldic symbols. However, this cannot explain the differences we observe between the Clemmensen and Renesse corpus.

Production costs. As argued in the introduction, the costs of producing heraldic symbols arguably differ for heraldic emblems compared to spoken language, and (to a lesser extent) to written symbols. Being destined for public display in durable formats, unlike private letters or spoken words, coats of arms were a type of symbol one could afford spending much effort on, in contrast with spoken, or even written, words, that have to be frequently produced anew. This decrease in production costs was compounded by the importance of heraldic seals, and, for the

Renesse but not the Clemmensen corpus, by the advent of printing. Printing in itself is not sufficient to cancel ZLA, as ZLA is found in Chinese or Japanese written characters, two systems that have used printing for centuries (Tamaoka & Kiyama, 2013; Zhang et al., 2007). However, it could be a facilitating factor. Processing costs thus provide a plausible, if incomplete explanation for the absence of a consistent ZLA in our two heraldic corpora, and for its reversal in the Renesse corpus. If right, this interpretation would imply that the principle of least effort at work in ZLA is (in this case at least) chiefly driven by production costs, not processing costs, in contrast with popular accounts of ZLA (Cohen Priva, 2017; Piantadosi et al., 2011). Note that production costs do not, by themselves, explain why the reversal of ZLA should be specific to iconic motifs.

Diffusion dynamics and iconicity

This leads us to two distinct differences between heraldry and other graphic communication systems: the inclusion of iconic and non-iconic motifs, and the fact that the frequency of heraldic motifs depends mainly on their cultural diffusion (whereas the frequency of words is linked to how frequent their referents get mentioned in speech, whilst the frequency of letters depends on that of the phonemes or morphemes that they encode). Iconic motifs and their cultural diffusion were shown to play an important role in reversing ZLA for both our corpora. Iconic motifs tend to be of higher complexity than non-iconic motifs, both in this study and in others (e.g., García et al., 1994) as they are depicted with more detail, enhancing their resemblance with the element they are depicting. This, in turn, suggests that what causes our reverse ZLA may have to do with pressures favouring the diffusion of iconic motifs.

Why do our two corpora differ to such an extent that complexity and iconicity are positively correlated to frequency in the later one, but not in the earlier? Asking this is tantamount to asking what caused iconic motifs to spread to a much larger extent than abstract ones. One reason could be a quirk of early heraldic history: the simplest abstract motifs (e.g. one bend, a pale) were reputed to be "honourable" and reserved for the oldest nobility (Fox-Davies, 1900; Rietstap, 1884): later-comers would need to make do with more variegated designs. True as this may be, this does not explain our results. The range of simple abstract shapes allowed by the rules of heraldry was in fact much broader than the set of "honourable" motifs, yet many possible designs did not find any adopter; and even "honourable" motifs were copied dozens of times (Morin & Miton, 2018). We cannot assume either that complex, iconic motifs were chosen because simpler, non-iconic motifs were no longer available: some of them, like the lion, were popular from the very start. The variety and evocative strength of iconic motifs made them supplant abstract patterns in other graphic codes, for instance the seal marks of ancient Mesopotamia (Wengrow, 2008). A more likely cause for the spread of iconic designs is thus, in our view, the printing press and the family of mechanical reproduction techniques that revolved around it, including etchings and lithography. Printing made it much easier to reproduce complex pictures reliably on multiple supports, and to diffuse complex, standardized motifs across distances. It marked a decisive break between the visual culture of the late Middle Ages and early modern periods (as witnessed, for instance, by the massive popularity of engravings). The introduction of mechanical reproduction techniques, by decreasing the production costs of complex iconic motifs would have driven the 'reverse' Zipfian effect we observed in our results, with more frequent motifs also being more complex.

Our results on heraldic motifs can be reconciled with the ubiquity of Zipf's law of abbreviation, not only in vocal communication, but also in writing systems. The law obtains, not only in the overwhelming majority of vocal communication systems, but also in those writing systems where a character's frequency is entirely decoupled from the length of the morpheme that it stands for. Yet these writing systems are devoid of iconicity. We argued that iconicity, aided by cultural diffusion and a change in production costs, stands in the way of abbreviation in the case of heraldic motifs. Lacking (or, at any rate, losing) iconicity may be a precondition for Zipf's Law of Abbreviation to emerge in a graphic tradition.

Chapter 3 – Graphic complexity in writing systems

1. Introduction

Humankind is now reaching an all-time high literacy rate all over the world (https://ourworldindata.org/literacy). Writing is a graphic code, i.e., a system of standardized pairings between symbols and meanings in which symbols take the form of images that can endure (Morin, Kelly, & Winters, 2018). It is a *visual* communication system which takes "the form of visible marks on the surface of a relatively permanent object" (Treiman & Kessler, 2015). Importantly, writing functions by encoding a natural language (Morin, Kelly & Winters, 2018).

Writing usually implies characters organized into sets, here called scripts. Characters are defined as the basic symbols that are used to write or print a language. A script is defined as "a set of graphic characters used for the written form of one or more languages" (ISO 15924 Registration Authority, Mountain View, CA, 2013). Scripts do not exactly map with spoken languages: for instance, the Latin script is used to write a diversity of European languages. In other words, scripts do not determine what writing encodes, but they determine what writing looks like.

Written languages contrast with spoken languages. Writing is a relatively recent invention: it can be traced back to a few invention events that occurred no more than a few thousand years ago, while spoken language has a much longer evolutionary history. Writing requires explicit and deliberate effort in learning and transmitting it to a much larger degree than speaking. The visual character of writing implies that written characters have to fit constraints of the human visual system (Dehaene, 2010; Dehaene & Cohen, 2007). At least two characteristics of writing systems reveal their adaptation to the human visual system. The characters of scripts use both anisotropy and symmetry (Morin, 2018), and tend to mimic natural scene statistics (Changizi et al., 2006; Testolin, Stoianov, & Zorzi, 2017). Both these characteristics effectively reduce the cost of their processing by the human visual system.

One other aspect of characters that would load on their fit with the human visual system can be found in their visual complexity, which influences performance (see Donderi, 2006 for a review and historic overview of visual complexity studies). Lower visual complexity correlates with easier learning, processing and use (Pelli, Burns, Farell, & Moore-Page, 2006). The present study has two main goals. It aims at testing different hypotheses regarding (1) what drives character complexity and (2) evolutionary patterns in character complexity.

What determines the graphic complexity of a Script's characters: type, size, phylogeny?

Two main drivers of characters' complexity have been hypothesized: the size of a script's inventory, and its type. The size of a script's inventory is the number of characters included in the script. Script's type is a way to categorize scripts based on which mapping principle they use, i.e., which linguistic unit is encoded by their graphemes (graphic units). To this list, we add phylogeny: the character complexity of one script can also bear the influence of which script it itself descends from. In other words, all other things being equal, a script descending from a complex script could have more complex characters than a script descending from a simpler script.

Changizi & Shimojo (2005) have put forward the idea that character complexity is mostly invariant, situated around three strokes per character, and so, independently of both the size of the script and its type. They suggested that scripts always maintain the same number of basic strokes per character, with bigger scripts having a greater variety of basic strokes but no need for more strokes per letter.

Size: Do scripts with larger inventories have more complex symbols?

Studies dwelling on the relation between complexity and inventory size have yielded conflicting results: Changizi & Shimojo (2005) found that characters' complexity was of three strokes on average, independently of the script considered, while Chang, Plaut, & Perfetti (2016) found that character complexity increased with the number of characters included in a script – and assumed that this relation was mediated by the type of script. The script's type is, in that case, thought to influence the complexity of characters through the number of graphemes required by the mapping between characters and linguistic unit. According to Chang et al.,

> « grapheme complexities covary with mapping principles between orthographic and linguistic units (Perfetti & Harris, 2013) in that more visually complex orthographies

tend to map onto higher level linguistic units. Indeed, **the need for complexity is driven by the size of the grapheme inventory, which in turn is driven by the size of linguistic units to which they map**: phonemes, syllables, syllabic morphemes, in increasing order. In perceptual judgments, there is little reason to think that mapping principles are relevant, especially for orthographies not known to a participant. »

Chang et al. (2016), our emphasis.

Both Changizi & Shimojo (2005) and Chang et al (2016) had substantial shortcomings, whether considering their datasets or the hypotheses they tested. (Changizi et al., 2006) limited their analyses to scripts that included less than 200 characters and thus excluded logosyllabaries. Their dataset was heterogeneous, and included 22 numeral systems alongside systems that encode spoken languages. This means that their dataset mixed systems of symbols that do encode spoken languages and linguistic units along with systems of symbols that do not. Numerical notation systems have their own classification and constraints that are not shared with written languages (Chrisomalis, 2004). Finally, the complexity of characters within scripts, in this study, was measured through manual coding of the number of strokes required to create it, making it difficult to replicate.

(Chang et al., 2016), while including some large scripts in their analyses, did not control for effects of phylogeny on their conclusions. It is worth noting that their unit of analysis if that of *written languages*, rather than scripts. Strictly speaking, scripts are just collections of images. They are not writing systems: a given script can be used to encode various languages, in various ways. Scripts do not exactly map with spoken languages: for instance, the Latin script is used to write a diversity of European languages. But our typological classification (alphabets, abugidas, logosyllabaries, etc.), strictly speaking, is based on semiotic criteria. It applies to writing systems not to simple collections of images. The reason we allow ourselves to speak of "types of scripts" is because the vast majority of scripts we consider are used for only one writing system, with a clearly-defined type. The remaining scripts are used for several writing systems, the vast majority of which belong to only one type (e.g. the vast majority of writing systems based on the Latin script are alphabetical). We thus use the semantic shortcut of speaking about "script types".

Not accounting for common ancestry can be particularly problematic for cross-cultural data, as some of the characteristics observed in a population might be due to their common ancestry: it is necessary to account for the fact that some observations are not independent from one another (Mace et al., 1994). This problem of non-independence of observations due to common cultural ancestor is referred to as Galton's problem. Additionally, rather than *testing* for the impact of the type of script on the characters' complexity, (Chang et al., 2018, 2016) *assumed* that there should be a difference in complexity based on the type of script, and used such differences as the chief assumption allowing them to evaluate different complexity metrics.

The present study aims to overcome the limit of both studies by including information on both types *and* phylogeny of scripts. Additionally, we do not exclude any particular type of writing systems, as long as they encode spoken languages. In other words, we only exclude numeration systems and include scripts in which characters are mapped on any *linguistic* unit, from alphabets to logosyllabaries. To the best of our knowledge, our dataset is the largest and most diverse dataset of scripts to date (see part 2.2.3 for more details on its composition).

Homogeneity: How much variance in characters' complexity is captured at the level of the script?

Several lines of evidence suggest that belonging to a given script should capture most of the variance in character complexity. First, inclusion in a given script captures many important sources of variance in character complexity that should not be expressed at the level of individual characters. This includes the material that the script is usually written on; the shape of the basic strokes making up the script; or general stylistic influences, among others. Second, we could expect that something like the principle of uniform information density (Jaeger 2010) that obtains for spoken language, may also obtain for written language, so that writers could be pushed to maintain a more or less constant complexity throughout the various letters that they write. Third, similarity between characters helps making features predictable, and thus makes reading easier (Treiman & Kessler, 2011). Homogeneity between characters within a script also facilitates the statistical learning required to learn how to read (Treiman & Kessler, 2011). Characters could thus be relatively homogeneous in complexity within each script. For those three reasons, belonging to a particular script should be the most important factor affecting character complexity, when compared to others the factors that are relevant to the complexity of individual characters-i.e., type or family (category based on phylogenetic and geographic grounds) (***Homogeneity*** hypothesis).

The Cultural Evolution of Writing: Do scripts' characters become less complex over time?

Cultural evolution and language evolution allow us to understand which evolutionary pressures act on communication systems. In particular, we can expect writing to have evolved to be an efficient communication system. Most communication systems, writing included, are codes, i.e., sets of pairing meanings with signals. An efficient code can be achieved by pairing signals and meanings in order to minimize their cognitive and communicative costs. Signals carry cognitive costs inherent to their production (for sources) and to their reception (for receivers). In both spoken and written languages, more complex signals have higher cognitive costs than simpler signals. Communicative costs refer to how informative – i.e., how likely to be paired with the correct meaning- are the signals once used in context. Efficient communication involves minimizing cognitive costs and maximizing communicative gains (i.e, informative content), thus determining an 'optimal' frontier (for a review and implications on semantic systems, see Kemp et al., 2018). Minimizing the cognitive costs of signals is one aspect of communication systems' efficiency. Simpler signals have lower cognitive costs than more complex or longer ones. Having simpler characters for the same set of meanings would thus optimize the cost associated with communication (it is a form of compression).

Weak biases, i.e., probabilistic tendencies to transform in systematic directions a content as it is transmitted and remembered, can determine which forms are stable and prevalent in experimental populations (Kalish et al., 2007). Such weak biases are known to explain cultural phenomena, including linguistic universals (Thompson, Kirby, & Smith, 2016). If (even weak) biases in favour of simpler characters exist, they would predict a decrease in complexity over the evolutionary history of scripts. Two main lines of evidence converge in favour of expecting writing systems to minimize their visual complexity: (1) higher graphic complexity makes scripts less efficient (Pelli et al., 2006), and (2) graphic complexity is hard to maintain in use, and especially through transmission events (Tamariz & Kirby, 2015).

First, higher visual complexity makes visual stimuli, and thus characters scripts, harder to learn to recognize (van der Helm, 2014). Simpler symbols are easier to learn, distinguish and remember (Pelli et al., 2006). More complex pictures take longer to identify, and also occasion more mistakes (Byrne, 1993; Donderi & McFadden, 2005; Pelli et al., 2006; Zhang, Zhang, Xue, Liu, & Yu, 2007). This effect of complexity is robust to participants' familiarity or experience with the images (Byrne, 1993), and to levels of noise, overall contrast, or eccentricity in the visual field

(Shu et al., 2003). Graphic complexity also weighs on the working memory load, thus making visual search harder (Alvarez & Cavanagh, 2004). These results imply that a script can increase in functionality by decreasing its characters' complexity.

Complexity is not always a detrimental property for graphic communication systems, and having relatively complex characters can also be efficient for a script. Overly simple characters can be hard to distinguish. At least some degree of visual complexity is required in order to allow letters to be distinct from one another. In turn, this allows scripts to encode more information with the same number of characters. Additionally, redundancy is a useful feature in differentiating and recognizing letters (as assumed by Changizi & Shimojo, 2005) – but it is also a feature that would increase the visual complexity of characters. Taken together, those opposite pressures (in favour of simpler or more complex characters) predict that characters of a script should be found around an optimal level of complexity: neither too complex nor too simple, given the information content they are used to convey.

Second, complex drawings and scribbles are known to simplify in experimental settings. Drawings, in particular, have been among the first type of stimuli used in transmission chain experiments at the very end of the 19^{th} century (Balfour, 1893) and in the first half of the 20^{th} (Bartlett, 1932). Transmission chain experiments function as games of "telephone": one participant is given a stimulus that she has to reproduce. Her reproduction is then given to a second participant who has to reproduce it in turn, and so on until it reaches the last participant in the chain. Scribbles have been showed to decrease in complexity over experimental generations, especially when drawn from memory rather than directly copied (Tamariz & Kirby, 2015). Interaction games also suggest that written communication should show some form of compression. These experiments require one participant to guess a meaning among a set of possible options, based drawings or scribbles produced by another participant. Whenever the same participants were allowed to play several rounds in a row, the drawings they produced became simpler, more abstract and less iconic (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007b). Scripts, during their lifetimes are submitted to very similar constraints of being reproduced from memory, transmitted, and used in communicative interactions. We would thus expect them to become simpler over time.

Finally, it has been suggested that changes in writing systems over time are relatively directed and would move from relatively iconographic or figurative variants (think Egyptian hieroglyphs) to more abstract and simpler characters (Gelb, 1963).

Ideally, the existence of such evolutionary patterns would be tested by using the time elapsed since a given script emerged as a proxy for the exposure to such evolutionary pressures. This is not a possibility in the current case, because that would not account for other factors potentially impacting characters' complexity, such as the number of its users, its function and its context of use. We thus make two predictions that should still allow testing for the existence of those evolutionary patterns. The first of those predictions bears on idiosyncratic scripts, i.e., scripts created after 1800 by identifiable creators, with no main influence from other existing scripts. The second of those predictions focuses on branching out events, i.e., the fact that one script (or more, called descendant) differentiates from another (ancestor) script.

Invention: Are recently invented scripts more complex than more ancient scripts?

If functional pressures drive the evolution of scripts, we can expect that recently invented (i.e., invented in the last two centuries) scripts would thus have had less exposure to such pressures. In turn, this would predict that the characters of recently created scripts would be more complex than those of scripts that were exposed to evolutionary pressures for a longer amount of time (***Invention*** hypothesis).

Descendants: Do parent scripts have more complex characters than their offspring?

Similarly, branching out events occur whenever a script differentiates from its parent script: a large share of scripts were formed by branching out from other scripts. They did so either as independent offshoots of continuing scripts, or as continuations of extinct scripts. Such branching-out events provide the opportunity to increase a script's efficiency, by simplifying its characters. If branching-out events favoured an increase in efficiency, it would predict that the characters of the "parent" script would, on average, be more complex than their offspring's characters (***Descendants*** hypothesis).

Distribution shape hypothesis

Previous work made remarks, in passing, on the distributions of the complexity measures of various graphic symbols being positively (i.e., right-) skewed. McDougall, Curry, & de Bruijn (1999; 2000), for instance, noticed that their sample of iconic and non-iconic symbols was positively skewed. Such a positively skewed distribution was also observed for European heraldic motifs (Miton & Morin, 2019), and in the different versions of the Vai syllabary (Kelly et al., submitted). Although shy on causal explanations, the literature on words' lengths in different languages suggests that this type of patterns is widespread in communication systems. Words length distributions have commonly been interpreted as following a type of negative binomial distribution (R. D. Smith, 2012; Wimmer & Altmann, 1996), which are effectively right-skewed too.

We expect constraints on the production, recognition, and reproduction of graphic symbols to weigh on the complexity distribution of a given script. While very simple symbols are easy to produce and to recognize, they quickly become 'saturated', i.e., it becomes harder to keep them easy to differentiate while keeping them simple. We expect this, to result in a positively (right-) skewed distribution. A script with a positively skewed distribution thus has most of its characters at a relatively low complexity, compared to the full range of complexity scores that it covers. We thus predict that scripts' characters complexity would follow right-skewed distributions (***Distribution*** hypothesis).

2. Methods

Pre-registration and data accessibility

We kept a complete research diary on the Open Science Framework website (https://osf.io/9dnj3/?view_only=203375dd595c4d098ee07ead30470ffb) where all analyses carried out were pre-registered and described. Pre-registration[4] is an open research practice and consists in describing both the research design and analysis plan as independently as possible from data collection.

Inventory constitution

System-level inclusion

The inventory of scripts included in our study was compiled from the Unicode 10.0, updated according to the Unicode 11.0 (*The Unicode Standard, Version 11.0, (Mountain View, CA: The Unicode Consortium, 2018. ISBN 978-1-936213-19-1)*, n.d.), and enriched with proposals that were available at that time. Proposals are detailed plans to format a script for inclusion into the Unicode Standard. Proposals differ from the Unicode standard in that they have not yet been validated and are still in the process of being included. All scripts included in the Unicode Standard are associated to unique 4-letter identifiers in the ISO15924, (e.g., [Latn] for the Latin script), which we use here to avoid ambiguities. This study excluded the following: secondary scripts (defined in (Morin, 2018) as scripts used by a writing system that encodes another system (e.g. Stenographics such as Duployan shorthand [Dupl]), non-visual scripts (e.g. Braille [Brai]), scripts that do not directly encode a spoken language (e.g., Blissymbols [Blis]), and undeciphered scripts (e.g., Linear A [LinA]). Further exclusions occurred during data collection. Because our study required us to generate pictures of each character for each script, scripts for which we could not find a font (necessary to generate the pictures) were excluded (see Appendix C for an exhaustive list of exclusions). Finally, a symbol was considered as missing whenever we could not produce a picture for it (i.e., whenever the font for the script did not have it). Scripts with up to 5 missing symbols were included.

[4] Time-stamped registration is available here: https://osf.io/dh4wg/?view_only=a0fa74fdf11a4e0d9dfaff8faf81818c

Character-level inclusion

Drawing on (Morin, 2018), a character was included if it could stand alone as one sound or, in the case of logographic systems, as one word or phrase. We thus exclude the following: punctuation marks and ligatures, diacritic marks, number symbols, honorific marks, and currency marks. The exclusion of ligatures and diacritic marks implied that the SIZE variable (i.e., the number of characters included in a script) was to a small extent, underestimated for abugidas and abjads (and their average complexity overestimated, as those types of signs tend to be very simple), compared to syllabaries and alphabets.

Description of the final dataset

Our final dataset was large and diverse: it included 47 880 characters from 133 scripts. It included (see Figure 1): 5 East Asian scripts, 23 European scripts, 35 Indian scripts, 24 Middle East scripts, 23 Modern Inventions, 11 South East Asian Insular scripts and 12 Mainland South East scripts. By types, it included: 17 abjads, 56 abugidas, 44 alphabets, 1 featural system[5], 4 logosyllabaries, and 11 syllabaries.

[5] The one featural system included in the present study is Hangul [Hang]. In featural systems, the shape of characters correlates with language's features. In Hangul, the shape of characters was designed to resemble the shape taken by the mouth to produce the corresponding sounds.

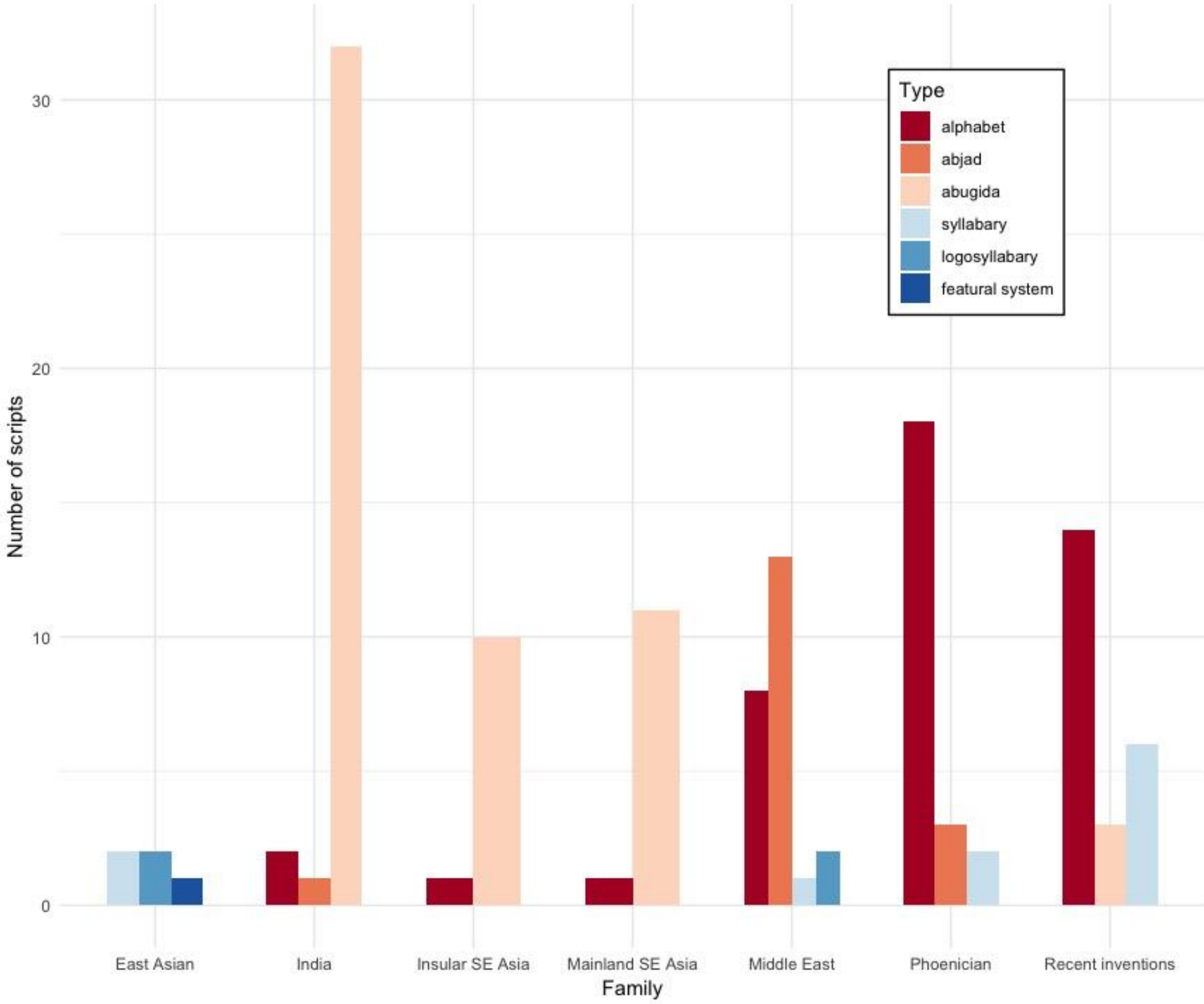


Figure 1. Composition of our dataset, by family and type of scripts.

Pictures Processing

Our analyses required a standardized collection of pictures, in which the amount of variation due to the use of different fonts would be minimized, while the variation due to actual characters shapes would be preserved. In particular, fonts vary on two properties that can affect the measures of characters' complexity: size and line thickness.

Generating pictures of characters

A picture of each character was generated using a Unicode range (code of four or five alphanumeric characters defining unique identifier attributed to each character) and a font. The bash script fixed the size of the picture at 500 by 500 pixels, and an initial font size for drawing the symbols at 60. Whenever a script presented characters that would be to too big to fully fit within the 500 by 500 pixels canvas, it was rerun at a smaller point size. In such cases, we decreased point size 5 by 5, until reaching a size at which all characters would fit inside the canvas. This was necessary for only four scripts ([Egyp], [Bali], [Mymr], [Gran] with respective final point sizes of 55, 45, 55 and 40).

Resizing

In order to standardize our pictures for size across scripts, we adaptively resized them. We first trimmed all the pictures. We then selected, for each script, the character with the largest picture (on either dimension, i.e., height or width). From this picture, we derived a ratio specifying how much it had to be resized for its largest dimension to fit a 490 by 490 pixels square (maintaining the aspect ratio and thus avoiding distortions). Finally, we used this ratio for resizing all pictures from the same script, and placed the resulting pictures back on a 500 by 500 pixels white canvas—see Figure 2. This procedure allowed us to minimize variation in size between different scripts, even when they used very different fonts, while maintaining the variation in size occurring *within* each script.

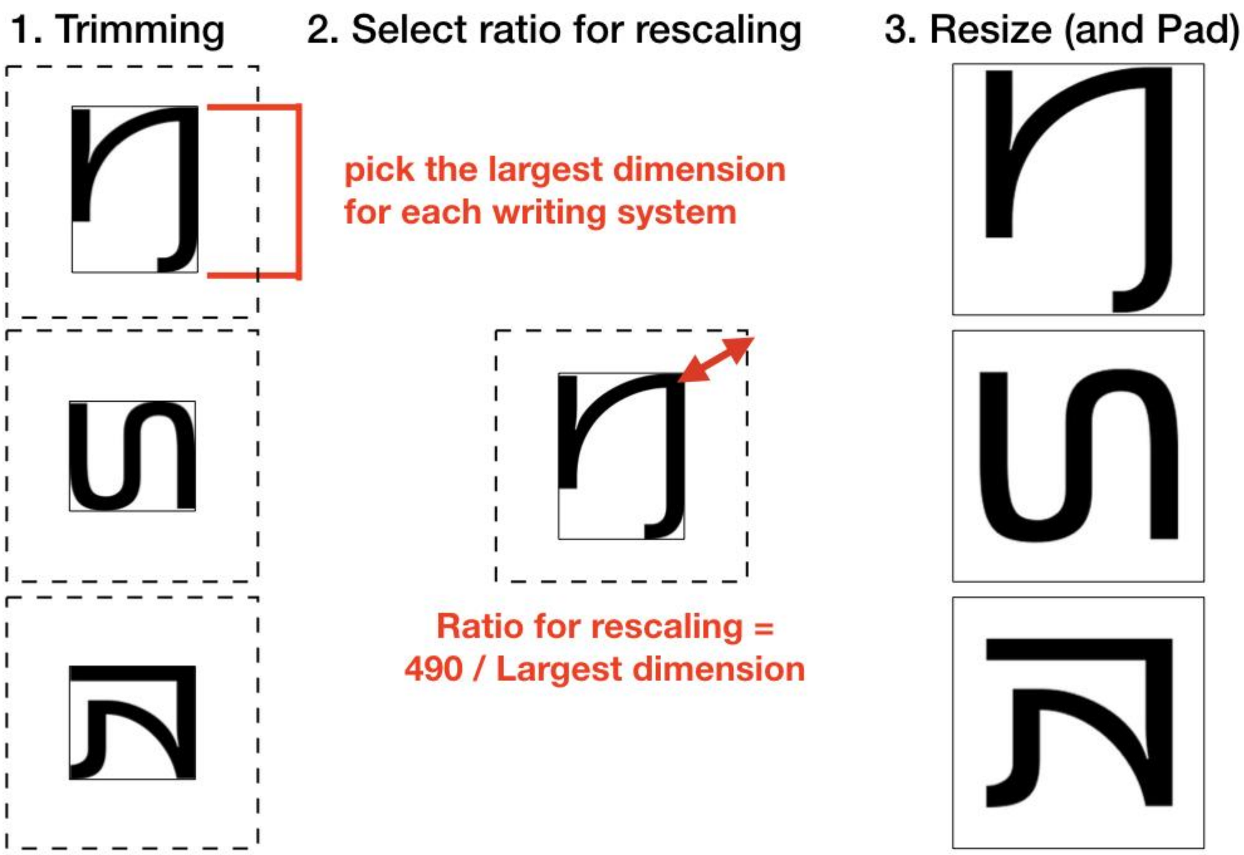


Figure 2. Procedure used to minimize the variation in characters' size between scripts.

Homogenize line thickness

In order to obtain a collection of characters that all had the same constant line thickness, we used a combination of functions in Mathematica (Wolfram language): first, thinning, then pruning, and finally, dilation. First, the Thinning function, (argument "Method" set on "MedialAxis") returned the approximate medial axis of the picture. Then, we applied a Pruning function (argument = 35) in order to get rid of some of the artefacts emerging with obtaining the approximate medial axis. This effectively removed the small segments that appear during the extraction of the approximate medial axis but were not part of the optimal (i.e., representative) skeleton of the character. Pruning branches whose length was inferior to 35 pixels yielded satisfactory results, as decided on the base of visual inspection of the pictures. Finally, the Dilation function (argument = 2), made the trait thicker, and more akin to usual characters. This procedure resulted in white characters on a black background (on which perimetric complexity measures were computed in Mathematica).

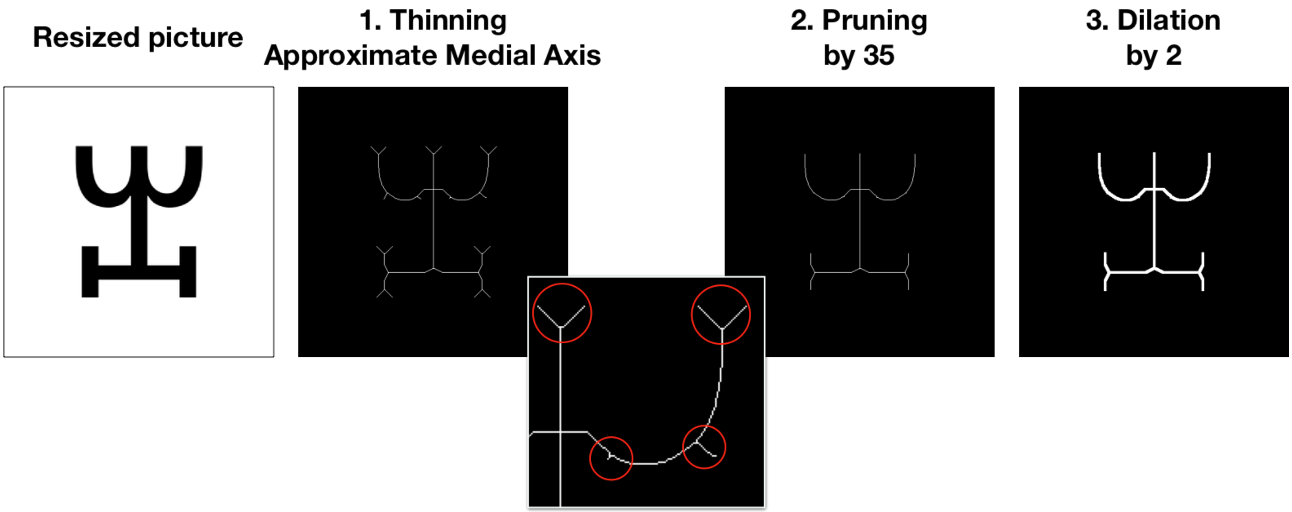


Figure 3. Procedure used to minimize the variation in line thickness, both within and between scripts. The red circles show examples of small problematic strokes that appeared from the thinning step.

Additional treatment for algorithmic complexity

Algorithmic complexity metrics were computed on pictures having black foreground (black character) over a white background. Each character's picture also went through the Potrace algorithm (Selinger, 2003) in order to get rid of any superfluous pixels and to get vectorised version before zip compression.

Measures of Visual Complexity

Following previous studies in cultural evolution (Kelly, Winters, Miton, & Morin, submitted; Miton & Morin, 2019; Tamariz & Kirby, 2015), two measures of visual complexity, here called "perimetric" and "algorithmic", were used.

Perimetric Complexity

Perimetric complexity is defined as ratio of inked surface to the perimeter of this inked surface (Attneave & Arnoult, 1956). It is obtained, using Watson's implementation (A. B. Watson, 2012), by taking the squared length of the inside and outside perimeters of a motif P, divided by the foreground area A and by 4π,:$C = \frac{P^2}{4\pi}$.

The measure was implemented in Wolfram (Mathematica), and correlates with human performance (Liu, Chen, Liu, & Fu, 2012; Pelli et al., 2006).

Algorithmic Complexity

Algorithmic complexity measures are obtained by compressing the .eps file (resulting from the Potrace algorithm). The proxy for algorithmic complexity is then the size in bytes of the compressed file: it offers an estimation of the length of the shortest computer program that produces the picture of the character without loss of information.

Phylogeny and Other Information on Scripts

The variables included in our analyses were the SIZE of each script (the number of characters that it includes), its FAMILY (defined based on information on both the geography and ancestor of each script) and its TYPE (e.g., alphabet, abugida, syllabary, etc.), and whether or not they were IDIOSYNCRATIC (i.e., created by identifiable creators in the last two centuries and with no overwhelming influence of any existing script). For analyses using character-level measures, an additional grouping variable SCRIPT refers to which script they are part of. Whenever applicable, ANCESTOR, i.e., which other script is considered an ancestor of the script, was also used.
We here present which sources we used to gather this information, and the coding of each of our variables.

Sources

We used the same sources as (Morin, 2018) to complete the information (characteristics and ancestor) for scripts that were included in this study but were not in (Morin, 2018), as well as to obtain our Type variable. One of the sources used in Morin (2018), The Ethnologue ("Ethnologue: Languages of the World" n.d.), could not be used in our study, due to its shift to a for-pay model. Whenever we needed to complement our information (e.g., when new scripts were included that were not present in the original study), we used all the other sources mentioned in

Morin (2018). All variables were coded by pooling together all available information from our sources. A majority rule was applied whenever our sources gave contradictory information.

Script size

Script SIZE was measured as the number of unique characters included in our sample for each script. When a letter or glyphs exists in several possible versions depending on its position (e.g. capital letters vs. minuscules in the modern Latin script), we count each version as one distinct character, following the Unicode Standard.

Scripts classification in families

Script FAMILY used the classification established by Morin (2018) on phylogenetic (i.e., ancestor) and geographic grounds. We slightly adapted definition of the East Asian family, due to differences in inclusion criteria between both studies. The seven families were the following:

- Middle Eastern family: direct descendants of the main scripts of the Middle East (these main scripts being Egyptian, Cuneiform, South Arabic and Aramaic).
- Phoenician family: all the direct and indirect descendants of the Phoenician alphabet, including Greek and its descendants.
- Indian Brahmic family: all the descendants of the Brahmic script in Modern India, Pakistan, Sri Lanka, Mongolia and Tibet.
- Mainland South-East Asian Brahmic family: all the direct and indirect descendants of the Brahmic script in mainland South-East Asia .
- Insular South-East Asian Brahmic family: all the direct and indirect descendants of the Brahmic script outside of mainland South-East Asia, in Indonesia and the Philippines.
- Recent inventions family: all the scripts created after 1800.
- East Asian family: Korean Hangul, Japanese Kanas and Chinese scripts that were not related to the Brahmi script (Han [Hani], Yi [Yiii], and Tangut [Tang]).

This FAMILY variable is not strictly phylogenetic: includes phylogenetic information under the form of ancestry (i.e., parent and offspring scripts), but also geographic, and resembles closely that offered by reference documents in the study of writing systems (Daniels & Bright, 1996). Phylogenetic information strictly speaking is captured instead by our ANCESTOR variable.

Like for all other variables, each script's last common ancestor was determined by pooling together information from all our sources. When sources were consistent with one another but differed in their specificity, the most specific source (citing the ancestor that was closest in time to its descendant) was chosen.

Types of scripts

Based on definitions from Daniels & Bright (1996, p. 4), we classified scripts according to the linguistic unit their graphemes map on and recoded the information from our sources as:
- Alphabet: "the characters denote consonants and vowels" – they are usually defined as "systems using the smallest possible phonemic subunit".
- Abjad ("consonantary"): "the characters denote consonants (only)" – in other words, such scripts leave readers to supply the appropriate vowel.
- Abugida: "each character denotes a consonant accompanied by a specific vowel, and the other vowels are denoted by a consistent modification of the consonant symbols" - They are also referred to as syllabic alphabets or alphasyllabaries in other sources.
- Syllabary: "the characters denote particular syllables, and there is no systematic graphic similarity between the characters for phonetically similar syllables".
- Logosyllabary: "the characters of a script denote individual words (or morphemes) as well as particular syllables".
- Featural: « the shapes of the characters correlate with distinctive features of the segments of the language ». The only such script in our sample is Hangul [Hang].

Idiosyncratic scripts

Scripts were considered idiosyncratic if they fulfilled the following criteria: (1) Precise information is known about their inventors (most often, their name), (2) There is no scholarly consensus that they derive their shape from the influence of one single identified ancestor. Most resemble no known script, others fuse many influences together so that no single dominant influence is discernible. They don't have any identified ancestor, though they may still take inspiration from known scripts in form and principles, (3) The script was invented (or scrapped altogether) after 1800. This definition excludes invented scripts such as Cherokee [Cher], which were invented *de novo* by an identifiable inventor, but nevertheless bear the dominant influence of one script (in Cherokee's case, the Latin script [Latn]).

3. Results

The two measures of complexity (perimetric and algorithmic) were reasonably correlated ($r = .84$).

Size hypothesis

Our hypothesis predicted that scripts that included more characters would have more complex characters. Our main result was that size had an impact on character complexity. In other words, our hypothesis was confirmed: the more characters in a script, the more complex the characters. However, this effect depended on whether large scripts (with inventory size > 200), including logosyllabaries, were included or not (see Figure 4). Our analyses also revealed that both FAMILY and TYPE of script were important predictors of character complexity.

The SIZE variable (number of characters in a script) was used as predictor in a nested regression analysis. The scripts were grouped into seven families based on the classifications most common in the literature, in order to account for shared cultural influences between distinct scripts. These families were used as the grouping variable in linear mixed models with random intercept, using the lmer function of the lme4 package for R (Bates, Maechler, Bolker, & Walker, 2014). A null model was built first, with a random intercept for FAMILY and for script TYPE; a second model introduced the script's SIZE (i.e., inventory's length) as a fixed effect.

On the full dataset (total $N = 47{,}880$ characters from 133 scripts), the best null model for characters' perimetric complexity included both TYPE and SCRIPT (which was nested by FAMILY) as random effects. A model adding SIZE as a fixed effect shows larger scripts to be more complex than simpler ones ($\beta = 0.12$, 95%CI [0.073, 0.175], $df = 21.222$, $t = 4.78$, $p < 0.001$). The two models were refitted using maximum likelihood for comparison purposes, showing that the test model was more informative (Akaike information criterion—AIC—of -8609.7 vs. -8597.2 for the previous model).

We replicated this result using algorithmic complexity instead of perimetric complexity. The best null model for characters' algorithmic complexity included both TYPE and SCRIPT (which was nested by FAMILY) as random effects, just like the best null model for perimetric complexity.

The test model showed larger scripts to be more complex than simpler ones (β = 0.04, 95%CI [0.0245, 0.0747], df = 24.79, t = 3.873, $p < 0.01$). The two models were refitted using maximum likelihood, revealing that including the script's size resulted in a more informative model (AIC— of -39144 vs. -39098 for the previous model, which did not include size as a predictor).

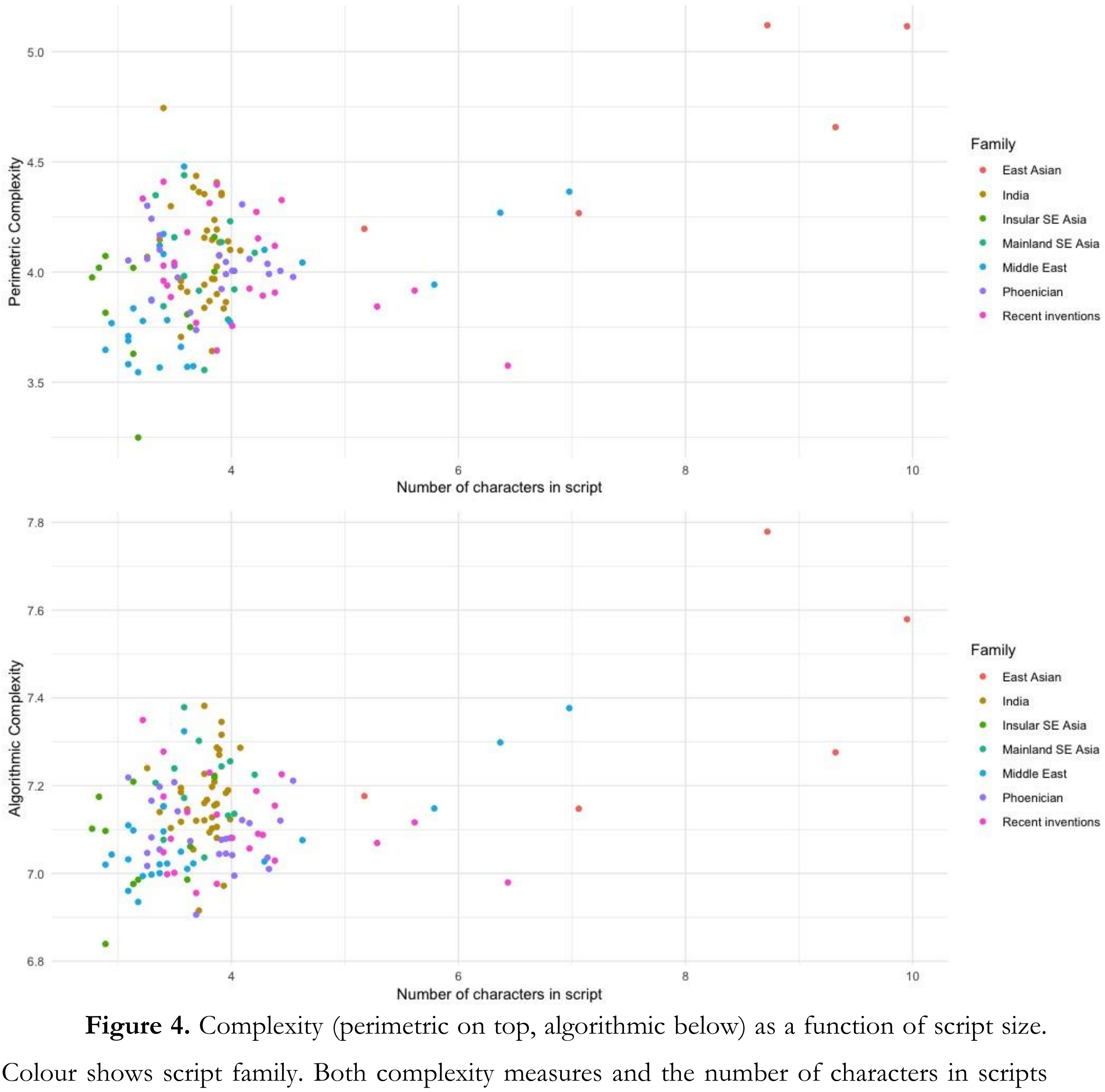


Figure 4. Complexity (perimetric on top, algorithmic below) as a function of script size. Colour shows script family. Both complexity measures and the number of characters in scripts were log-transformed.

On a subset including only scripts that had less than 200 characters (*N* = 5566*,* on 124 scripts), similarly to Changizi & Shimojo (2005), the effect of script SIZE seems to disappear. Most of the effect of the size of the system seem to depend on the inclusion of a few very large systems (mostly East Asian) which also tend to have very complex characters. We thus partially replicate Changizi & Shimojo, in the sense that characters' complexity do not seem to be impacted by SIZE, as long as we restrict our analyses to the scripts in the same range as their analyses.

Having removed these large scripts from the dataset, the best null model for characters' perimetric complexity (df = 2.78, t = 64.32, $p < 0.001$) included both TYPE and SCRIPT (which was nested by FAMILY) as random effects. A test model did not show larger scripts to be more complex than simpler ones (b = 0.06, 95%CI [-0.048, 0.168], df = 92.03, t = 1.086, p = 0.28). Adding SIZE did not make the model more informative (AIC of 3269.4 vs 3268.7). The best null model for characters' algorithmic complexity (df = 3.33, t = 285.3, $p < 0.001$) included both TYPE and SCRIPT (nested by FAMILY) as random effects. Adding SIZE did not make the model more informative (AIC of -4394.7, against -4394.6 for the null model). This model did not show larger scripts to be significantly more complex than simpler ones (b = 0.03, 95%CI [-0.012, 0.076], df = 100.58, t = 1.42, p = 0.159).

Homogeneity hypothesis

Our hypothesis predicted that the script to which a character belongs would predict over half of the variance in character complexity. Contrary to our predictions, TYPE captured more of the variance in character complexity than either SCRIPT or FAMILY. This was true for both perimetric and algorithmic complexity, see Figure 5.

The intraclass correlation (ICC) was calculated on raw values for perimetric complexity measures and on log-transformed values for algorithmic complexity (in order to avoid convergence issues), using the ICC1.lme function in the "psychometric" R package. An ICC for letter complexity nested by SCRIPT showed that 38.57% of the variance in perimetric complexity and 38.49% of the variance in algorithmic complexity is accounted for by their inclusion in a particular writing system. While this represent a relatively high percentage of the variance, it remains under the predicted value of 50%. By comparison, FAMILY accounts for 29.74% (algorithmic complexity) to 45% (perimetric complexity) of the variance, and type captures 68.26% of the variance in perimetric complexity and 55.43% of the variance in algorithmic complexity. Type was thus the variable that captured most of the variance in character complexity, contrary to our predictions.

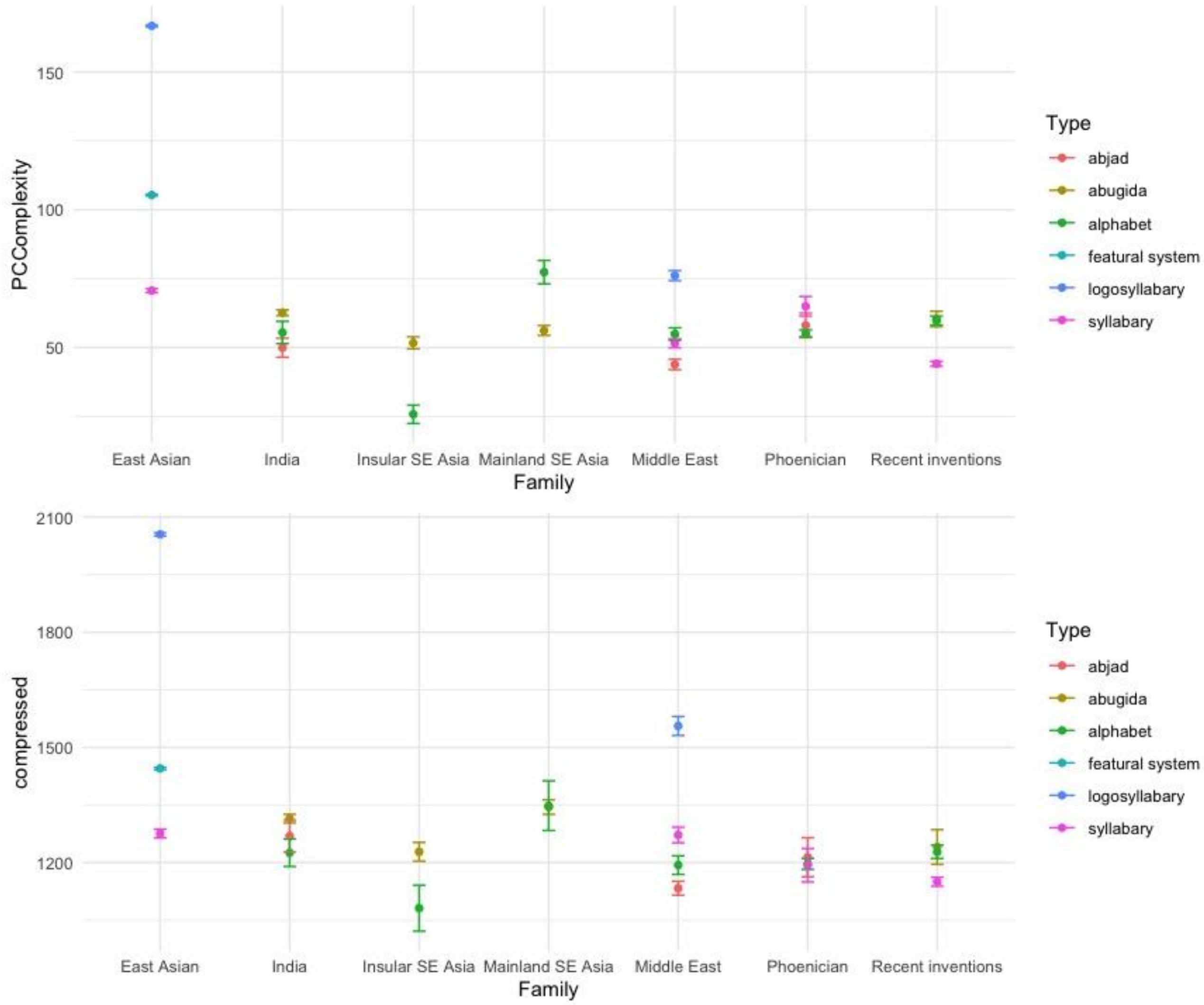


Figure 5. Complexity by family and type (error bars represent 95% confidence intervals): the top panel represents perimetric complexity, the bottom panel represents algorithmic complexity.

Invention hypothesis

Our hypothesis predicted that characters from idiosyncratic scripts would be more complex than characters from non-idiosyncratic scripts. Idiosyncratic scripts (i.e., created by identifiable creators after 1800, with no overall influence of another existing script) were, by definition, all classified as belonging to the Recent Inventions family, which includes all scripts created after 1800 (idiosyncratic and non-idiosyncratic). The null model for this hypothesis thus did not include FAMILY as a random effect. Contrary to our predictions, there was no significant effect of adding the IDIOSYNCRATIC variable to the model - the AIC increased, rather than decreased, when we added it, for both perimetric and algorithmic complexity measures.

The test model failed to show any effect of IDIOSYNCRATIC (β = 0.016, 95%CI [-0.128, 0.161], df = 126.08, t = 0.226, p = 0.822), when compared to the best null model for characters' perimetric complexity. Adding IDIOSYNCRATIC actually increased the AIC, indicating a less informative model (AIC—of -8595.2 vs. -8597.2 for the null model).

IDIOSYNCRATIC did not seem to have an effect on algorithmic complexity either (β = -0.003, 95%CI [-0.062, 0.056], df = 125.05, t = -0.108, p = 0.914), when compared with the best null model for algorithmic complexity. Adding IDIOSYNCRATIC resulted in a less informative model (it increased the AIC to -39131 versus -391333 for the null model). Idiosyncratic scripts were also neither more nor less complex when compared with other scripts from the Recent Inventions family that were *not* idiosyncratic (β = 0.68, 95%CI [-10.67, 12.79], df = 20.98, t = 0.11, p = 0.913 for perimetric complexity, β = 9.72, 95%CI [- 103.45, 121.56], df = 20.28, t = 0.172, p = 0.865, for algorithmic complexity, characters nested by SCRIPT for both), see Figure 6.

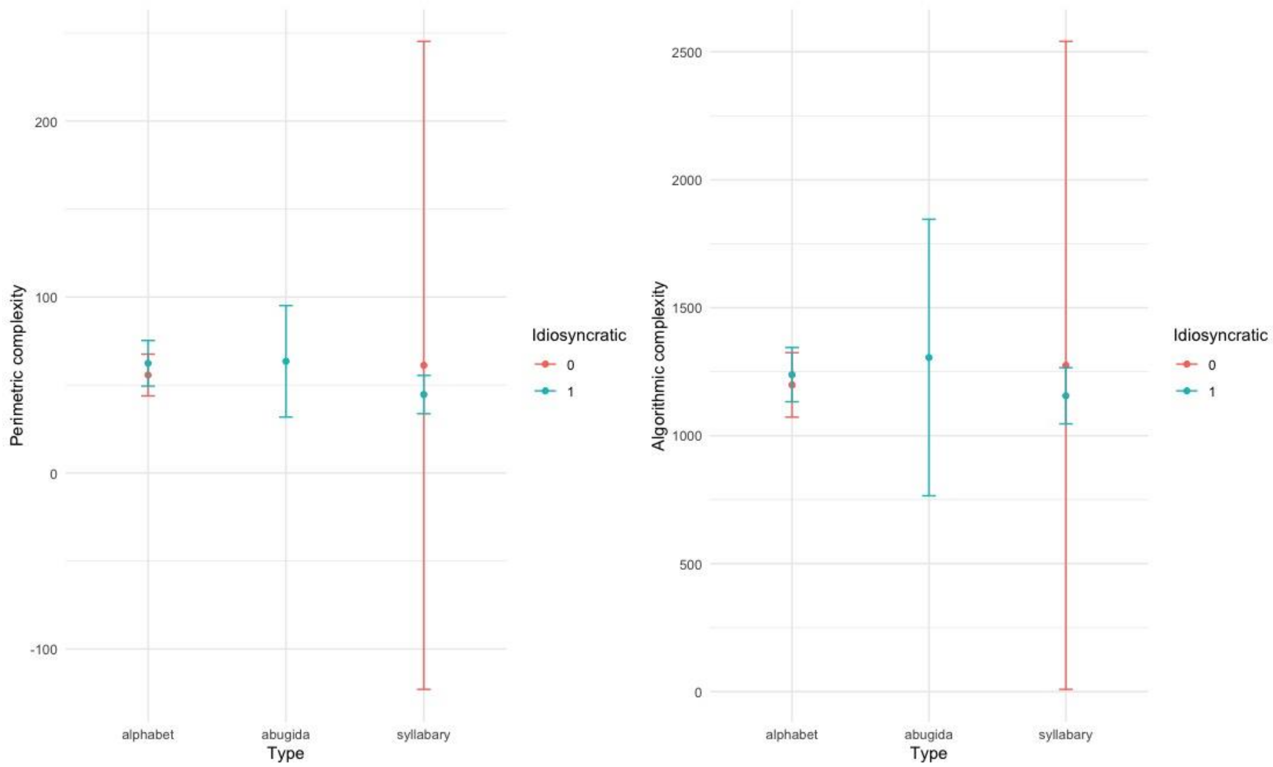


Figure 6. Complexity by type, comparing idiosyncratic (in blue and non-idiosyncratic (in pink)

scripts from the Recent Invention family for both perimetric (on the left) and algorithmic (on the right) complexity measures. Error bars represent 95% confidence intervals.

Descendants hypothesis

We hypothesized that, in case of branching-out events, a « parent » script's characters would be more complex than its offspring's characters. For each pair, the ancestor's average complexity (i.e., the mean complexity of its characters) was subtracted from the descendant's average complexity, as pre-registered. Our dataset includes information on 102 branching out events, from 29 different ancestor scripts. The most frequent parent script was Brahmi [Brah], with 25 offspring scripts. A parent script had, on average, 3.55 descendants ($SD = 4.98$).

When controlling for ancestor (i.e, including ANCESTOR as a random effect), algorithmic complexity did not seem subject to any systematic effect: no significant increase nor decrease in complexity occurred with branching-out events ($\beta = 12.87$, 95%CI [-34.98, 57.71], $df = 29.69$, $t = 0.563$, $p = 0.577$). Perimetric complexity tended to *increase* (not decrease) with branching out events, but this trend failed to reach significance ($\beta = 3.734$, 95%CI [-0.65, 7.35], $df = 21.44$, $t = 1.823$, $p = 0.082$), see Figure 7. These results suggest that the null hypothesis may be true (no tendency for descendants to diverge from ancestors changes in complexity in a particular direction).Nevertheless, the linear mixed effects model analyses so far presented do not test this directly.

A Bayesian one-sample t test was conducted to see whether the data supported the hypothesis that descendants do not, on average, decrease or increase their complexity compared to their ancestor. It found moderate support for the null for both perimetric ($BF = 4.02$) and algorithmic complexity ($BF = 5.06$) – see Figure 7. Differentials were averaged for each ancestor, rather than for each descendant-ancestor pair: this avoided giving more weight to ancestors with numerous descendants (such as the Brahmi script).

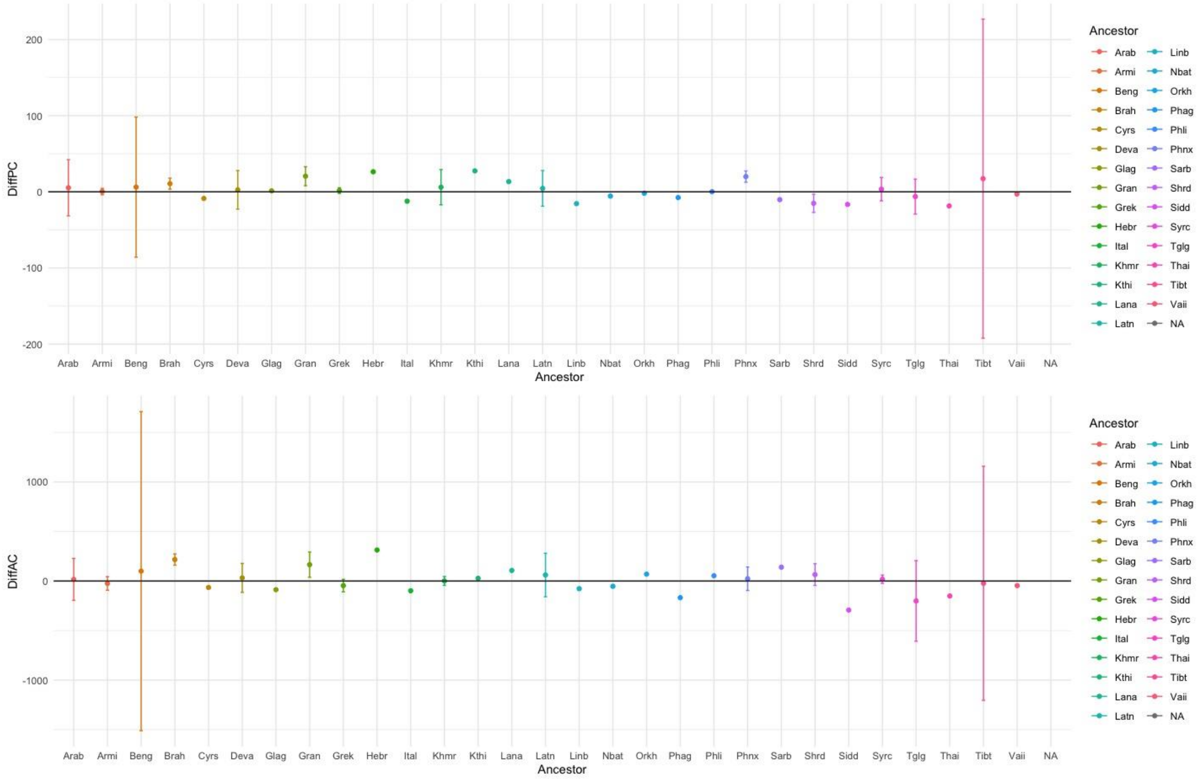


Figure 7. Difference between means of descendant scripts and ancestor scripts plotted for each documented script, by alphabetic order (ISO key), for perimetric complexity (top) and algorithmic complexity (bottom). Error bars represent 95% confidence intervals.

Distribution hypothesis

Our hypothesis predicted that the complexity of characters within a script would follow a right-skewed distribution. This was confirmed by a Bayesian one sample t-test (against the null hypothesis of skewness measures being equal to zero, BF = 44184058 for algorithmic complexity, BF = 19155462142 for perimetric complexity). This overall tendency for scripts' distributions of complexity to be positively skewed was relatively evenly distributed over both families and types of scripts, see Figure 8.

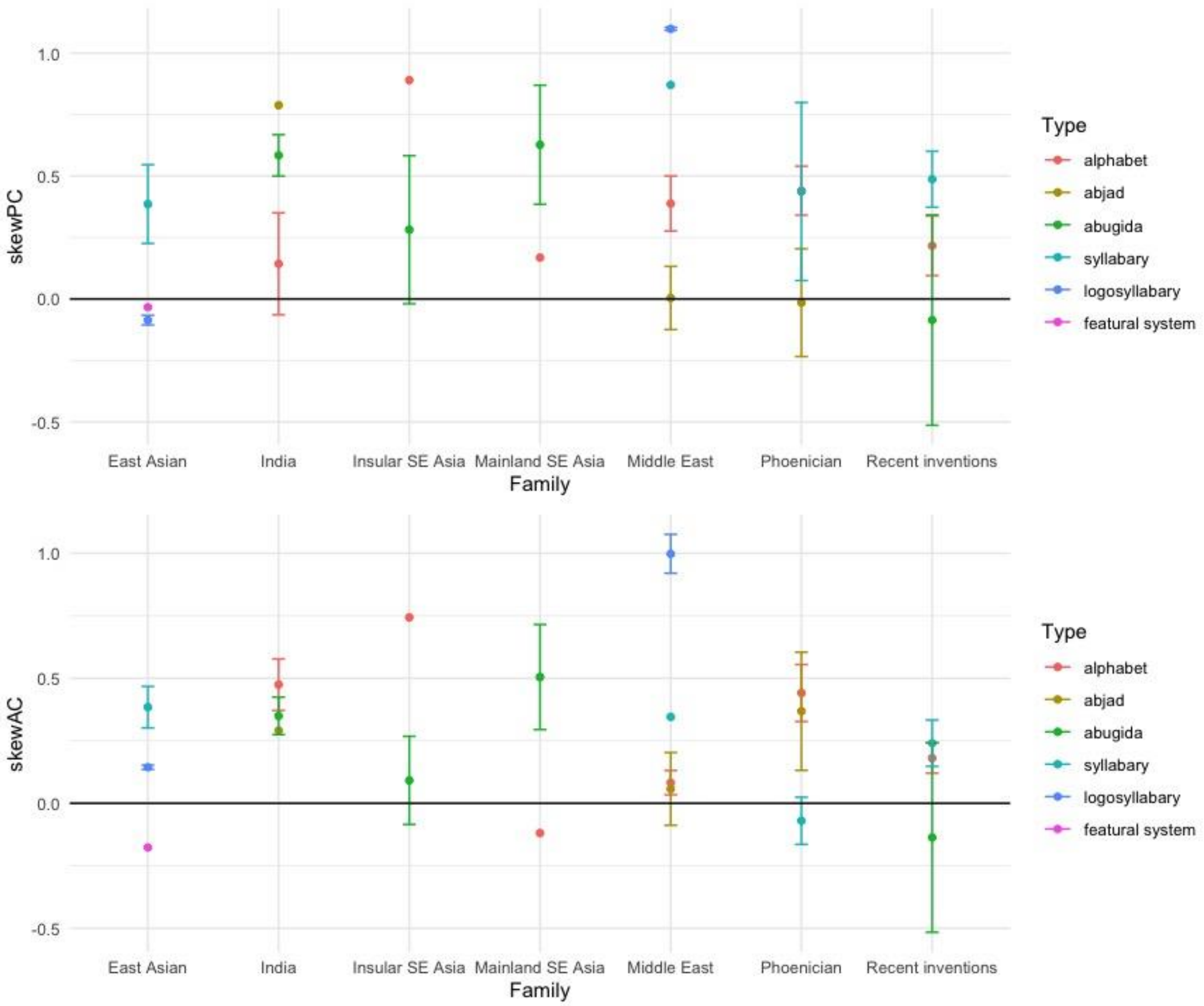


Figure 8. Skewness measure for all 133 scripts in our dataset, ordered by alphabetical order of the IsoKeyA. Bold line indicates a skewness of 0. The top panel shows skewness measures for perimetric complexity, and the bottom panel skewness for algorithmic complexity. Error bars represent the standard error to the mean.

4. Discussion

Importance of scripts' type on character's complexity

This study gives us insights on what constitutes the main determinant of character complexity: the number of characters included in a script did not robustly impact characters' complexity independently of script type: most of the variance in complexity was accounted for by the type of writing system that a script was mainly used for (e.g. alphabetic, syllabic, etc.).

Our study's inventory size estimates are based on (1) our inclusion criteria, i.e., what we consider to be a character able to stand on its own, and (2) how much of each script had already been encoded in the Unicode at the time we constituted our dataset. In particular, we excluded ligature marks, which might have influenced the estimates of scripts including such type of characters, mostly abjads. Because this would have overall lead to overestimating the average complexity of abjads' characters, which are nevertheless among the type of writing systems that have the least complex characters. In other words, this cannot explain our results, and would have worked against, rather than in favour, of our hypothesis. The contrast between our results and previous results suggest that previous studies on written characters' complexity might be quite heavily influenced by whether (large) logosyllabaries are included in the dataset.

Contrary to our predictions, the script that a character belongs to did not account for as much variance in character complexity as did the script's type. In previous studies, causality was usually assumed to flow from type of script to number of characters to complexity of the characters. This assumption is present in (Chang et al., 2018) who evaluated different measures of character complexity in relation to their capacity to distinguish between types of scripts. Our results suggest the relationship between type and characters' complexity might not be mediated by the size of the script, as previously assumed (Chang et al., 2016), but instead determines both script size and character complexity. This also contradicts Changizi and Shimojo's (2005) claim that the type of script did *not* influence character complexity. Although we replicate their result when we reproduce their decision to exclude some scripts, we reverse it when the full range of script type and sizes is taken into account.

No Decrease in characters' complexity

Overall, there is very little evidence of a decrease in complexity over the macro history of scripts. We put forward two hypotheses ("Invention" and "Descendants") derived from our assumption that scripts should manifest a decrease in character complexity. Neither were supported. Idiosyncratic (i.e., recently invented scripts) were *not* more complex than scripts that were exposed to evolutionary pressures for several centuries, sometimes more. Character complexity did not overall decrease when parent scripts branched out into descendant scripts. We discuss three possible interpretations of these results: (1) differences in use of scripts might cause a lot of noise, (2) the decrease in character complexity occurs early and rapidly in a script's "life", or finally, (3), scripts tend to appear at an already close-to-optimal level of complexity.

First, social and cultural factors varying from context to context could have impacted the complexity of scripts' characters. Variation in social and historical contexts implied that scripts were not necessarily used for the same purposes, or by the same populations. It is unclear how variation in function would have impacted complexity directly, but it implied different populations of users, ranging from trained scribes to nearly everyone in the population in cases of widespread literacy. Higher complexity could have been maintained more easily whenever scripts were used only by a specialized fraction of the population. Finally, in some cases, writing, and especially handwriting, is made to reflect social belonging, by being unnecessarily sophisticated (Thornton, 1996). Nevertheless, variation in complexity due to function or users can only be expected to have a local influence, i.e., to be circumscribed to the specific contexts and environments in which there is either a narrow function for writing and/or restrictions on who can join the community of users and how. Any impact such context-dependent and localized factors may have had on our results can be assumed to be itself localized and context-bound, thus unlikely to bias our results in any systematic way.

Second, another possibility would be that compression processes could not be captured in the data we gathered and analysed. This could be the case if the graphic simplification of characters occurred early and rapidly in the "lives" of scripts. This is also in line with the fact that in experimental settings, such effects are known to occur over very short timespans. We know from a more focused study on the Vai script (developed in Liberia during the 19th century), that at least *some* simplification of characters occurred during that syllabary's very first decades of existence.

Third, one last possibility is that scripts' characters tend to be created already compressed, i.e., at or close to an optimal level of complexity. In this scenario, characters when created are more or less as simple as they can be while remaining informative and easy to discriminate. If they are not, they become optimally complex shortly thereafter. The broad tendency for distributions of characters' complexity to be right-skewed might indicate some form of optimality *from the start,* as it can be interpreted as a bias in favour of producing mostly relatively simple forms. If most spontaneously produced characters were relatively simple, and only a few characters were more complex, this would result in right-skewed distributions. Those results resonate with Morin's (2018) results on cardinality in writing systems. Scripts exploited the sensitivity of the human visual perception to oblique and cardinal orientations: their characters over-represented those orientations, compared to what could be expected by chance. In this study, even ancient writing systems extensively used cardinal orientations, and offspring scripts did not use make a more extensive use of those orientations than their ancestors. Both of those results suggested that cultural evolution had relatively little importance in explaining this extensive use of cardinal orientations.

Limits and Future Directions

The current study has several limits. First, the images we used to analyse scripts were (as any representation of a script must be) idealisations. They abstract away a great deal of internal variation due to time, to space, to differences between writers, etc. There is also a need to study writing systems on their own, through their own chronological trajectory (see Kelly et al., *submitted* for an example of such a case study on the Vai syllabary of Liberia).

Second, the visual complexity of individual characters is only one of many possible ways to consider complexity in scripts. Future research could address other types of complexity and their evolution, such as set-level complexity (i.e., how compressible is the whole set of characters included in a script, including how many patterns or features are re-used by different characters in the same script, Bennett, Gacs, Ming Li, Vitanyi, & Zurek, 1998; Vitányi, Balbach, Cilibrasi, & Li, 2009), which might be able to capture constraints related to features extraction and discriminability between characters (Mueller & Weidemann, 2012). Third, scripts' characters' complexity also depends on the way characters are combined and occur in real world settings. The visual complexity of scripts *in use* differs from the complexity measured on their characters independently from one another. When used alongside other characters, higher complexity can actually sometimes provide with a processing advantage, for instance during visual search (Bernard & Chung, 2011; Chanceaux, Mathôt, & Grainger, 2014)– such effects have been interpreted as more complex characters having more different features that can be used to help the search. Additionally, although complexity is hard on naïve learners, its inconvenients tend to fade off with expertise (Wiley, Wilson, & Rapp, 2016).

In this study, we focused on the visual aspect of scripts' characters, and how they are perceived, i.e., how they are recognized as characters by a reader. Nevertheless, characters also have to be produced, which meant, for most of their history, hand-written. Although more complex characters would, overall, be more costly to produce than simpler ones, some shapes might be easier to produce than others. Changes related to motor production of the characters might also impact the graphic complexity of characters (see Parkes, 2008 on cursivisation). Motor-program based estimates of complexity (Lake, Salakhutdinov, & Tenenbaum, 2015) might offer computational options for such future research. Differences in techniques used for producing characters offer natural experiments in parsing out the influence of motor production on (e.g., appearance of printing).

By suggesting that type trumps other characteristics of scripts as a predictor of characters' complexity, our results have implications for future research on scripts and the evolution of writing. Scholars like Gelb (Gelb, 1963) have put forward positivist accounts of the evolution of writing, in which writing evolves from one type of writing system to the next, from large linguistic units (logosyllabary) to smaller linguistic units (syllabary, then alphabet). Although this hypothesis was mostly based on anecdotal evidence, our results suggest that changes in character complexity might occur as a consequence of changes in type of writing systems.

Cognitive and environmental factors of attraction

As sketched in the introduction, factors of attraction are to be understood as causal factors that shape cultural phenomena. The next two chapters differ in two main respects. They focus on different types of factors of attraction (cognitive or environmental) and differ in how they test for such factors (methods and type of data).

1. Contrast in types of causal factors: cognitive and ecological factors of attraction

Two kinds of factors of attraction are usually distinguished: cognitive factors on one side, and ecological or environmental factors on the other side. This distinction is based on the location at which such factors of attraction are taken to operate, inside the mind/brain in the case of cognitive (or psychological) factors, or outside of the mind for ecological factors. Still this distinction should not be seen as one between two exclusive categories. Inside or outside the mind is nowhere near a clear distinction. Some factors having an impact on cultural transmission are not easily located precisely inside or outside the mind. Psychological mechanisms are, in general, adapted to the environment in which they evolved – just as well as the environment's features are partly the outcome of a coevolution with humans. Conversely, the environment's relevant properties might result from human behaviour or intervention (especially for artefacts).

The cognitive / ecological distinction roughly charts out one axis among others on which causal factors vary. Mixed or ambiguous factors that do not clearly belong to one or the other category are also clearly to be expected.

Cognitive factors of attraction have focused a lot of Cultural Attraction's research program (Miton et al., 2015; Morin, 2013, 2018). It is also the type of factors of attraction that was explored in chapters 2 and 3: they both focused on one cognitive factor of attraction in visual communication system, i.e., their complexity. Chapter 4, focusing on spatial composition biases in artworks (human profile-oriented portraits) is also mostly cognitive.

Chapter 5 focuses on physical and motor constraints, which makes it a study of an *ecological* factor of attraction. Examples of ecological factors of attraction abound in all domains that have strong constraints based on the availability of raw materials, or of technical constraints from their format. For instance, consider again the China girls' example presented in the introduction: their existence (and persistence) was very much sustained by the risk for the materials (films) on which they occurred might vary in how they rendered colours from one reel to the next. This material variation of colours when developing films was the reason for the existence and spread of China girls (Yue, 2015). Other examples of ecological factors of attraction are provided by the way properties of raw materials influence the cultural practices that exploit them. For instance, the invention and spread of the spinning wheel changed which type of wool was the most priced, decreasing the appeal of wool with especially long threads, as mentioned in (Clark, 2016).

The next two chapters thus illustrate two types of factors of attraction: cognitive factors for chapter 4, and environmental factors, including physical constraints, for chapter 5.

2. How to test for factors of attraction: methods and types of data

The two next chapters also test for attraction on two different types of data, and in two different ways (briefly described in the introduction). Chapters 4 and 5 differ in which types of data they use: chapter 4 uses large-scale historical data, while chapter 5 uses experimental data.

In chapter 4, testing for the existence of an attractor is done by predicting and observing frequencies of different spatial compositions in human profile-oriented portraits over historical time periods. By contrast, chapter 5 presents a transmission chain experiment, and exploits the possibility to artificially create less attractive variants as seeds in these chains.

In chapter 4, we start with the existence of cognitive processes that impact production and aesthetic reception of different types of spatial composition in pictures. These cognitive processes may bias which spatial composition is produced and judged as aesthetically pleasing (in favour of centering and in favour of ex-centering a focus element in a frame). They constitute a hypothesized factor of attraction. Human profile-oriented portraits were chosen as a cultural type on which this factor of attraction should have an impact. This predicts a relative prevalence of portraits in which the sitter is ex-centered, with more space in front of the sitter than behind them. To test whether this was the case, we used a large-scale dataset, which was curated from two websites (WikiArt and ArtUK). In this case, a cognitive forward bias as a factor of attraction predicts that placing more free space in front of a sitter rather than behind them would be an attractor. We had a secondary hypothesis that historical norms would be another factor of attraction, modulating how strongly this forward bias would be expressed. We could thus use frequencies (number of occurrences), and how they clustered in the variation space (i.e., their proportion of free space placed in front of the sitter) to test for our hypothesis.

For chapter 5, the hypothesized factor of attraction resides in physical (motor) constraints. This predicts that rhythms produced by participants, especially novice ones, would reflect motor constraints. We use a transmission experiment to test that prediction: the initial sequence participants in the first generation are asked to reproduce is the same for all conditions–a short metronome sequence. Conditions differ in which sequence of movements participants had to do while reproducing the sequence they heard. These movements were either all of the same amplitude or mixed two different amplitudes. They further branched out in either all large movements or all small movements, and on either starting with a large or a small movement. We hypothesized that these different conditions (i.e., variation in the factor of attraction) would determine different attractors: isochronous rhythms for conditions with only one type of movements (with a different average tempo based on movements' amplitude), and non-isochronous rhythms for conditions mixing movements of different amplitudes. This is tested by tracking transformations. We expect transformation of the initial rhythmical sequence from a not-so attractive variant (i.e., it does not match the motor constraints) into more attractive variants (i.e., rhythmical sequences matching the motor constraints) through successive reproduction and transmission events.

These two pairings between large-scale cultural data and cognitive factor on one hand, and experimental data and ecological (physical) factor are by no mean implying that some types of factors are better tested in one way and other types in another way. Both types of data are, in principle, well-suited to test for both types of factors of attraction.

Chapter 4 –
A forward bias in human profile-oriented portraits

1. Introduction

Human cognition impacts the forms taken by culture throughout a variety of domains, including medical practices (Miton et al., 2015), or the shape taken by written characters (Changizi et al., 2006; Morin, 2018). Within the diversity of human practices, art might offer an especially suitable domain in which to expect cognitive factors to shape cultural productions. In particular, aesthetic appreciation of visual art depends on characteristics of the human visual system, which affect colour preference or spatial structure (see Palmer, Schloss, & Sammartino, 2013 for a review). It has also been argued that aesthetic feelings are related to how easy it is to process a specific content. The Processing Fluency theory (Nadal, Munar, Marty, & Cela-Conde, 2010; Reber et al., 2004, 1998) posits that ease in sensory information processing has a key role in eliciting aesthetic, hedonic feelings.

Human portraits have non-random characteristics, some of which might reflect the cognitive processes implied in their production and reception. First, there is a tendency to show more of the left than the right cheek, in paintings (McManus & Humphrey, 1973), and in photographs (Labar, 1973) This preference to depict a sitter's left cheek has been attributed to a desire to convey emotions (Lindell, 2013; Nicholls, Clode, Wood, & Wood, 1999). A more direct bias[6] known in human portraits relates to gaze-orientation (Morin, 2013). Direct gaze of the sitter tends to predominate when pictorial conventions allow it, as they started doing, for instance,

[6] We here reserve the term *bias* for the observation level, i.e., whether it is observed in the portraits that were produced, and so, independently of which process (cognitive or not) is responsible for it.

during the Renaissance**.** This bias can be explained by the cognitive disposition to treat gaze as a source of social information that is easier to exploit in direct than in averted gaze.

Spatial composition has long been a topic of interest in understanding aesthetical appreciation in visual arts (Puttfarken, 2000). It refers in particular to the positioning of depicted objects or agents in the frame and relative to one another. We are here particularly interested in what happens when representing only one human agent, in a non-ambiguous orientation.

Human cognition may drive spatial composition of artworks, such as portraits, in two opposite directions. On the one hand, not all of a canvas' space is equipotential: there are preferences in spatial composition that do not depend on the depicted object. There is, for instance, a preference for centering horizontally any object represented on its own on a rectangular canvas (Arnheim, 1983; Palmer & Guidi, 2011; Tyler, 2007). Other preferences depend on the type of object depicted. There is for instance, a preference for having more space in front of agents or objects that have a clear front-back orientation when they are represented in profile. Both tendencies in favour of specific spatial composition (for centering, and in for ex-centering) are robust through different experimental paradigms, and are known to weight in perception, production and aesthetic appreciation.

Centering the object of attention in the middle of a canvas is an attractive type of spatial composition. It has been one of the focus of theories of spatial composition in art. (Arnheim, 1965, 1983), drawing on his background in Gestalt theories of perception, suggested that rectangular frames come with a particular structural skeleton supporting spatial composition. This idea of structural skeleton (sometimes called net) received mixed support from different experimental paradigms : it was supported by fit-rating results (Palmer & Guidi, 2011), and partially supported by an analysis of pictures with high aesthetic ratings (Jahanian, Vishwanathan, & Allebach, 2015); on the other hand, McManus et al. (2011) did not find support for its existence. Horizontally centering an object in the frame increases explicit both judgments of balance (Leyssen, Linsen, Sammartino, & Palmer, 2012; McManus, Edmondson, & Rodger, 1985), and aesthetic preference (Leyssen et al., 2012; Mitsui & Noguchi, 2002; Palmer & Guidi, 2011; Sammartino & Palmer, 2012).

Spatial compositions in which the focal object or agent is centered are also easily and spontaneously *produced*. In experiments in which participants controlled the adjustments of pictorial elements in a frame, they mostly produced pictures in which the focus object was centered (P. Locher, Cornelis, Wagemans, & Stappers, 2001; Locher, Jan Stappers, & Overbeeke, 1998; Puffer, 1903). Aleem and colleagues (Aleem, Correa-Herran, & Grzywacz, 2017) also found a tendency for sitters to be centered in a set of Renaissance portraits. Finally, there is evidence that painters tend to place one of the sitter's eye so that it crosses the vertical center line (Tyler, 1998, 1998, 2007). However, this observation about production is not matched by a similar effect in aesthetical preferences: in a 2 alternative forced choice paradigm (2AFC): portraits with one eye perfectly centered were not preferred by naïve viewers over portraits which did not center one of their sitter's eye (McManus & Thomas, 2007). These results suggest that centering one of the sitter's eyes might be more present in *producing* portraits than it is in *appreciating* it.

On the other hand, as soon as objects have an intrinsic orientation, the preferences for their spatial position on a canvas change, in favour of more ex-centered compositions. Changes as small as using oval shapes rather than circular ones influence which spatial composition were rated as best fitting, driving it away from the center (Guidi & Palmer, 2015). Representing oriented objects and agents in particular shows a tendency to have more space in front of them than behind them. This tendency goes against the more general tendency of strictly centering a single focal objects. This ex-centering tendency has been evidenced in perception, production and aesthetic appreciation. Preferences for spatial compositions that do not center their focal object come in three forms: (1) orientational or directional effects related simply to shape ("inward bias"), (2) effects due to the representation of agents and their bodily organization ("anterior bias") and, in particular, effects due to expectations of movements and speed, referred to as a "motion bias" (McBeath, Morikawa, & Kaiser, 1992), (3) effects related to monitoring of the agent's gaze-direction ("forward bias").

More specifically: first, participants tend to put more space in front, rather than behind oriented objects, such as ovals or triangles (Guidi & Palmer, 2015). These oriented objects' also became less centered over successive episodes of reproduction (i.e., iterated learning, Langlois, 2018). Those types of composition are also judged as more aesthetically pleasing (Guidi & Palmer, 2015).

Second, directionality of agents and directionality of their movements are related and perceptual and cognitive systems throughout the animal kingdom rely on this relation (Apfelbach & Wester, 1977; Catania, 2009; Cooper, 1981; Hernik, Fearon, & Csibra, 2014; S. M. Smith, 1973). Most animals' body-plan is bilaterally symmetrical with an antero-posterior axis, and thus their movement potential tends to be along this main axis and towards the front, rather than toward the back. Cues of this anteroposterior axis are the basis for perceptual decision-making in humans. For instance, movement direction can be used to disambiguate the front from the back of an ambiguous figure (Bernstein & Cooper, 1997; Pavlova, Krägeloh-Mann, Birbaumer, & Sokolov, 2002), animal body's directionality (as well as directionality of non-animate yet directional objects such as letters) can be used to disambiguate an ambiguous movement direction (McBeath et al., 1992), and orientation of a moving elongated symmetrical objects biases perception and prediction of its motion trajectory (Morikawa, 1999). Directionality of a stationary novel agent may support expectations of its future action-direction from early on in human infancy (Hernik et al., 2014). These results suggest a close link between perception of directionality of agents' bodies and representing their observed and anticipated motions. They give substance to the hypothesis that, in visual arts, whenever profile-oriented agents are not perfectly centered, they should be depicted with more space in front of them than behind them.

Indeed, experimental participants tend to put more space in front, rather than behind oriented objects (e.g., teapots), vehicles, or agents in a drag-and-drop task (Palmer & Langlois, 2017). This anterior bias has also been observed in fully ecological settings - i.e., within cultural productions. It has been observed both in depiction of animals in three European sources (two animal painters, Stubbs and Bewick, and a medieval bestiary, in Bertamini, Bennett, & Bode, 2011), and in contemporary movies from 4 different movie directors (Bode, Bertamini, & Helmy, 2016).

Pictures in which directed objects such as vehicles and teapots, or agents, humans and animals, are depicted with more free space in front than behind them have also been judged aesthetically more pleasing than cases in which they have more space behind them when using a 2-AFC (Two Alternatives Forced Choice) task (Palmer, Gardner, & Wickens, 2008).

Third, another clear source of directional information easily accessible to the observer is agent's gaze. Observed shifts of gaze-direction bias covert visual attention (gaze cueing) in very young human infants and possibly new-borns (Farroni, et al. 2004). They elicit overt shift of observer's own gaze (gaze following) in infants as young as 5-6 months of age, raised in diverse cultural contexts (Senju & Csibra, 2005; Hernik & Broesch 2019; Gredebäck et al 2008, 2018). Gaze-cueing and -following may be best expressed in response to dynamic and communicative gaze-signals, especially in infants (Farroni, Johnson, Brockbank, & Simion, 2000; Gredebäck, Astor, & Fawcett, 2018; Hernik & Broesch, 2019; Senju & Csibra, 2008). Gaze following is also widespread across non-human animals (Itakura, 2004; Kano et al., 2018). The strong tendency of a gaze-cue to engage observer's attention in the direction indicated by the cue, suggests that, in visual arts, we should expect agents to be represented with more space in the direction in which they are looking, than in the opposite direction.

The three sources of bias may contradict one another – in particular, agents can be represented with a body-direction that does not match their gaze-direction. For this reason, we selected only profile-oriented portraits, with matching body- and gaze-direction. Consequently, we cannot disentangle the sources of bias.

For all those reasons, in the current study, we test for the existence of a forward bias in human profile-oriented portraits. When they are not exactly centered, we predict that there should be more space in front of the sitter than behind her (hypothesis 1).

Given that we know of biases favouring two opposite types of spatial composition, i.e., centering agents versus ex-centering agents represented in profile (and leaving more free space *in front*, rather than behind, the agent), we expect that the relative frequency of portraits exhibiting the bias would mostly depend on external factors, such as historical norms and favoured formats. We thus aim to test whether through Western portraits' traditions a Center bias has been in force and whether it has been progressively relaxed with the greater freedom of composition encouraged in Europe from the Renaissance onwards (Puttfarken, 2000), allowing for a stronger effect of the forward bias. In an historical perspective, this could have been helped, in particular, by a change in the type of supports (coins, medals, medallions, framed canvases) and techniques and styles (e.g., the decline of 15th -16th centuries Italian tempera) which might have led to purely centered profile portraits becoming less taken for granted. We test whether this forward bias becomes stronger and more frequent over time (**hypothesis 2**). This imply two predictions: **(a)** The more recent the portraits, the more frequently they should show the bias, and **(b)** the stronger should the bias be (i.e., the larger the difference between the free space in front and behind the agent).

We tested those two hypotheses (existence of a forward bias and historical emergence of a forward bias) on one large-scale dataset (N = 1831 portraits, from 582 unique painters), curated from two main sources ArtUK.org and WikiArt.org.

2. Methods

Data collection - sources

We defined profile portraits as portraits having only one eye visible. Additionally, to be included in our analyses, a portrait had to fulfil the following criteria: (1) it had to depict only one sitter, (2) its sitter had to neither be interacting with any particular object nor involved in performing any particular activity, and (3) the sitter's orientation must have been unambiguous.

Pipes, cigarettes and masks, because they are usually placed in front of the agent and thus might influence the sitter's placement, also constituted a reason for exclusion.

We also excluded artworks for which spatial composition could not be assessed, such as cropped details of artworks (rather than full frame), and portraits with frames that were not rectangular (e.g., oval frames). Artefacts other than rectangular paintings or drawings (e.g., plates, mixed materials including 3d parts, stamps, and canvases or supports with irregular shapes) were also excluded. Finally, all pictures that included mirrors or reflections of the sitter were also excluded, as they included two pictures of the sitter.

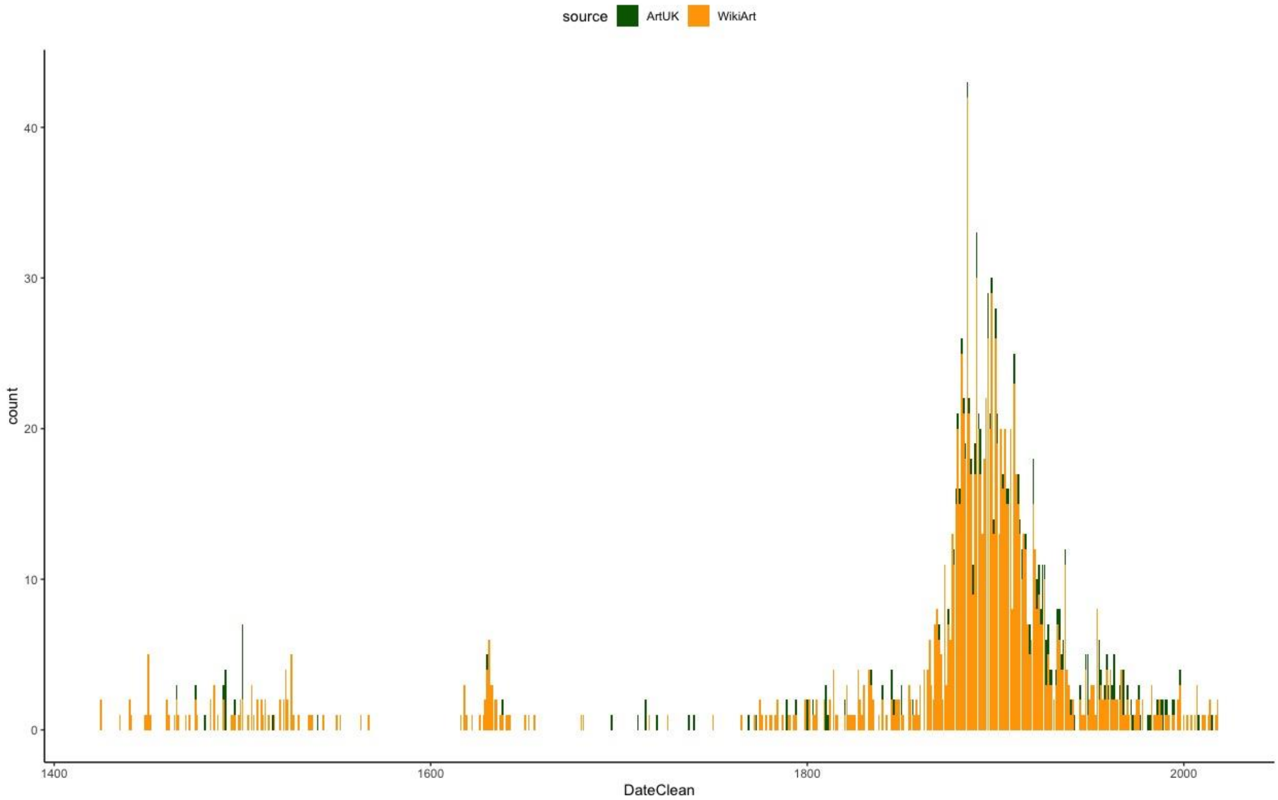


Figure 1. Composition of our dataset, including the number of portraits by source (WikiArt in orange, ArtUK in dark green). It represents the 1429 portraits (out of 1831) for which we were able to obtain a precise date.

The ArtUK website was consulted between August and early October 2018. As stated in the pre-registration, we first used the keyword *profile* to search the website. Search results were then filtered by a naive research assistant to keep only the portraits fulfilling our inclusion criteria (N = 221). The WikiArt website was consulted between early October 2018 and early February 2019. A research assistant, who was unaware of our research hypotheses, went through each painter's page within the portraits subsection, and selected paintings that fulfilled our inclusion criteria ($N = 1610$ portraits). There was no overlap between both datasets – given that both datasets had very similar results, they are presented here collapsed into a single dataset (see Appendix and report on OSF for results on each dataset separately).

Our full dataset included a total of 1831 portraits, and included 1095 Female portraits, 729 Male portraits and seven sitters of ambiguous gender. 1086 of the portraits were oriented to the left, and 745 to the right. 427 unique painters were represented, with 1 to 70 portraits by painter (M = 3.09, SD = 5.58). The earliest paintings were from 1425, and most recent ones from 2018, with most of the portraits dated late 19th and early 20th century.

Measures

Akin to Bertamini et al. (2011), relevant measures were defined as horizontal distances between the sitter and the margins of the frame, taken from the farthest extremity at one end (most often, chin or nose) to the farthest extremity at the other end (back of the head, hair) of the body of the sitter itself, as shown on Figure 2 (see Appendix for details about frames). Hats and other paraphernalia were not taken into account, as they are not part of the agent per se. Whenever portraits depicted sitters' bodies below the shoulders, another set of measure was taken, again from the farthest extremity at one end to the farthest extremity at the other end, but at the level of the body, rather than at the level of the face. We present results for both sets of measures.

The proportion of free space in front of the sitter, referred to as strength of the bias, is defined as (pixels in front) / (pixels in front + pixels behind) – i.e., as the proportion of the free space placed in front of the agent relative to all the free space available – see Figure 2. For a portrait to be considered as showing the bias, it had to have its proportion of free space in front of the sitter strictly larger than the proportion of free space behind the sitter.

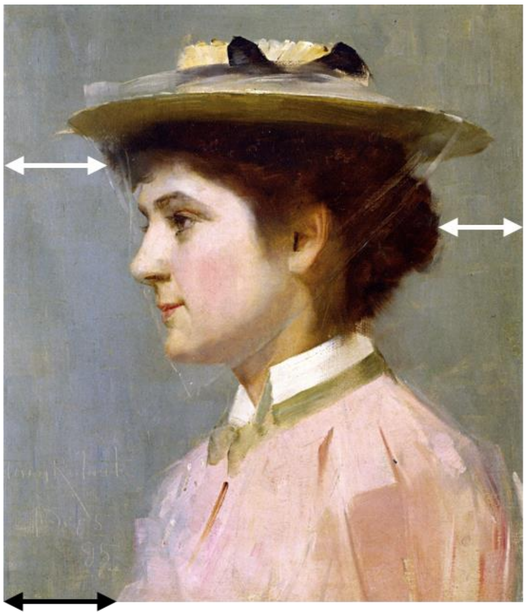


Figure 2. *Miss Isobel McDonald* (1895) by Tom Roberts. Here, 54.54% of the free space is placed in front of the sitter for the head measures (white arrows), and all (100%) of the free space measured at the level of the body is placed in front of her (black arrow, the sitter's garment touches the limit of the frame behind her, meaning that this measure equals zero). Measures excluded hats but included hair as part of the sitter's head.

All measures were taken by a research assistant who was naïve to the present study's hypotheses. Slightly over 30% of each dataset's portraits were recoded by a second coder (author HM). Intercoder reliability was high, as confirmed by two-way mixed intra-class correlations assessing absolute agreement (icc function, irr package), ICC = 0.965 (95% CI [0.962, 0.968]), on 2416 measures taken over 604 portraits.

Pre-registration and data accessibility

The research design and data analyses were decided before data collection and pre-registered on June 28th, 2018 on Open Science Framework. The initial registration[7], full research records (including supplementary analyses) and raw data can be found here: https://osf.io/6dsnx/?view_only=9835c19958754d5ab0672e7e54a9edae

3. Results

Hypothesis 1: Prevalence of a forward Bias

When using head measures, 1395 out of 1831 of the paintings showed the forward bias, which is significantly different from the 50% chance-level as tested by a Fischer exact test (OR = 11.73, 95%CI [9.31, 14.84], $p < .001$). On average 62.32% of the free space was located in front of the sitter's head, which was higher than expected by a one-sample Wilcoxon test (V = 1378700, $p < .001$, $r = 0.575$). Similarly, when using body measures, 926 out of 1619 of the paintings showed the bias, which is significant as tested by a Fischer exact test against the chance level of 50% (OR = 3.58, 95%CI [2.85,4.51], $p < .001$). On average 60.55% of the free space was located in front of the sitter's body, which was higher than the 50% expected by chance, by a Wilcoxon test (V = 669140, $p < .001$, $r = 0.262$).

This forward bias was present in both left- and right- facing portraits. For head measures, the bias was significantly more marked in right-facing portraits (Med = 61.25) than in left-facing portraits (Med = 59.31), as confirmed by a Mann-Whitney U test, U =434150, Z = 2.81, $p = 0.004$. For body measures, left-facing portraits (Med = 64.06) had not significantly more of their free space localised in front of their sitters than right-facing portraits (Med = 62.70), as indicated by a Mann-Whitney U test, U = 247340, Z = 0.776, $p = 0.438$.

---

[7] Time-stamped registration: https://osf.io/2n5k8/?view_only=df8ad420e3884161b8383f04cc45b8ba

Both datasets showed that painters had a tendency to put more free space in front, rather than behind the sitter, and this whether measures were taken from the body or from the head of the sitter. The forward bias was slightly more marked for right-facing than for left-facing portraits (see Figure 3).

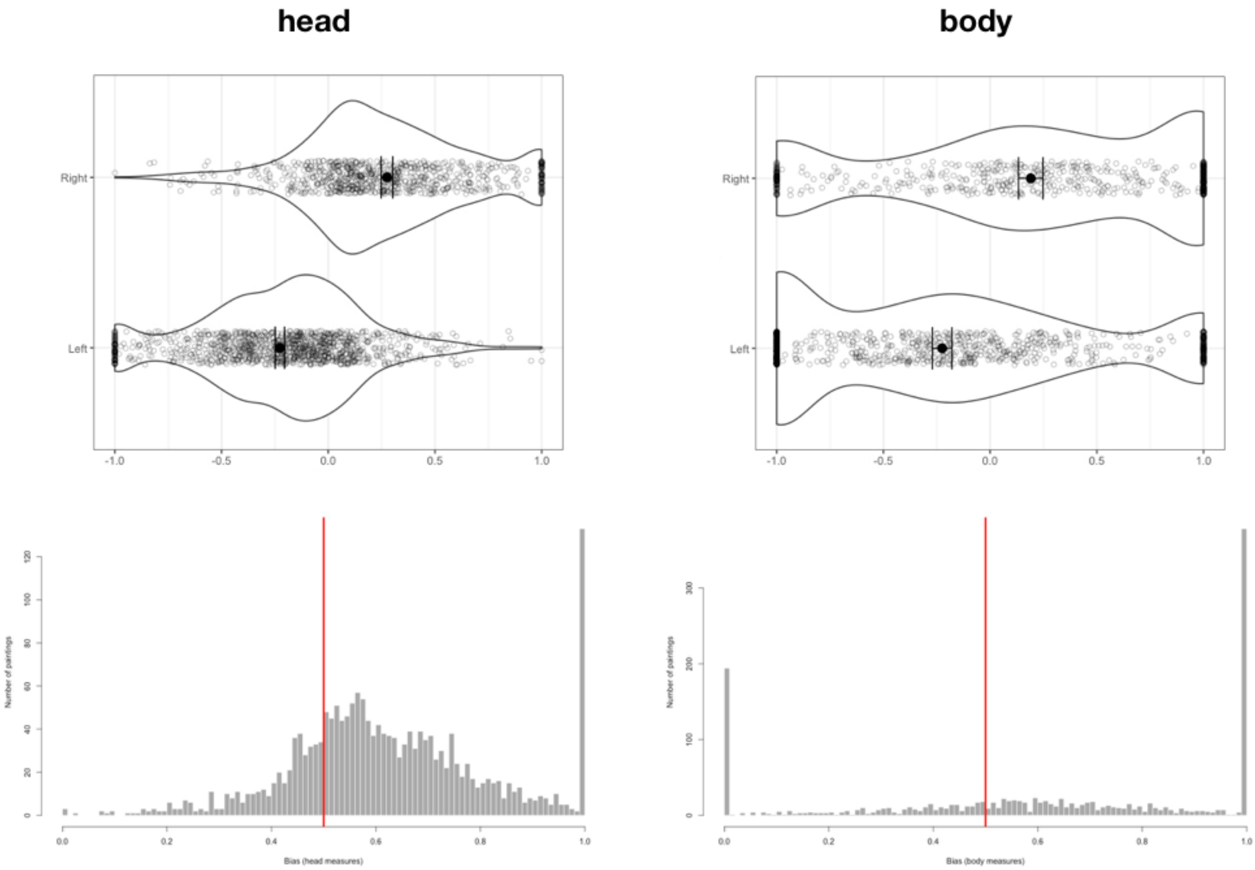


Figure 3. Prevalence of the forward bias in profile-oriented portraits. Above: The x axis represents the difference between the proportion of free space on the left and on the right of the sitter (Right - Left) (y axis = orientation, left or right facing). The values on the x axis range from -1, meaning that all the free space was located on the sitter's left, , to 1 meaning that all the free space was located on the sitter's right (0 means equal amounts of free space on both sides of the sitter). The plain round dot represents the mean, the error bar the 95%CI, the outer shape the distribution (density), and the grey small circles are individual data points. Below: Number of paintings by their ratio of the free space in front of the sitter to free space behind the sitter. On the X axis, 0 means that all the free space is situated behind the sitter, 0.5 that there is as much space in front and behind, and 1 that all the free space is in front of the sitter. Head measures are

represented on the left (and were available for all the 1831 portraits in our dataset), and body measures on the right (available for 1619 portraits).

Hypothesis 2: Historical emergence of the Forward bias

Out of 1831 portraits, 1429 had a precise date. Dated and non-dated paintings did not significantly differ in how their proportion of space in front versus behind the sitter they were, neither for head measures (Mann-Whitney U test, $U = 281490$, $Z = -0.474$, $p = 0.635$) nor for body measures ($U = 171100$, $Z = 0.08$, $p = 0.929$).

The distribution of dates in our datasets was strongly negatively (left-) skewed, and did not follow a normal distribution (Shapiro-Wilk: $W = 0.65p < .001$, skewness = -2.34). In order to have reliable results on regressions, the date variable was mirrored and log-transformed and finally mirrored back. This treatment ensured that the date variable's skewness remained under 1, while keeping the association between the date variable and the dependent variable.

Hypothesis 2a: Increase over time in the prevalence of the portraits conforming to the forward bias

We ran a binary logistic regression with portrait's date as independent variable and showing (coded as 1) or not showing (coded as 0) the forward bias to determine whether date impacted how likely a portrait was to exhibit the bias. This model did show that more recent paintings were more likely to show a forward bias when measured from the sitters' head (OR = 3.15, 95%CI [1.96, 5.12], $p < .001$ – Wald $\chi^2(2) = 369.4$, $p < .001$ for the whole logistic regression), but not when measured from her body (OR = 1.27, 95%CI [0.85, 1.93], $p = 0.248$ - Wald $\chi^2(2) = 26.7$, $p < .001$ for the regression overall).

Hypothesis 2b: Increase in the amplitude of the forward bias

Overall ex-centering, i.e., the asymmetry between the spaces on both sides of the sitter (either in the direction of a forward bias or opposite to it), increased over time in both our datasets. There was a positive correlation between date and overall excentricity: the more recent the portrait, the less centered its sitter, on both our datasets and types of measures $r_\tau = .13$, $p < .001$, 95%CI [0.095, 0.161] for head, $r_\tau = .08$, $p < .001$, 95% CI [0.046, 0.118] for body measures). This was also confirmed by a simple linear regression with log-transformed date as an independent variable and how excentered were the sitters as a dependent variable, ($F(1, 1427) = 51.16$, $p < .001$, $R^2 = .03$): more recent portraits had more excentered sitter ($b = 0.09$, 95%CI [0.07, 0.12], $t(1427) = 7.15$, $p < .001$), when using measures from head. Similar results were obtained for a simple linear regression on body measures ($F(1, 1264) = 22.44$, $p < .001$, $R^2 = 02$): more recent portraits excentered their sitters more ($b = 0.06$, 95%CI [0.03, 0.08], $t(1264) = 4.74$, $p <.001$).

The more recent a portrait was, the more pronounced was the forward bias (see Figure 4). When considering only portraits exhibiting a forward bias (i.e., have more free space in front than behind a sitter), the more recent the portrait, the stronger the forward bias, $r_\tau = .14$, $p < .001$, 95% CI [0.104, 0.184] for head measures; $r_\tau = .10$, $p < .001$, 95%CI [0.046, 0.144] for body measures).

A linear regression ($F(1, 1097) = 20.03$, $p < .001$, $R^2 = .02$), with (log-transformed) date as an independent variable and strength of the bias (measured as the ratio of space in front of the sitter to the space behind the sitter) as the dependent variable, suggested that the more recent the portrait, the stronger its expression of the bias $b =0.07$, 95%CI [0.04, 0.10], $t(1097) = 4.48$, $p < .001$. For body measures, a simple linear regression ($F(1, 721) = 0.66$, $p = .417$, $R2 < .01$), with date (log-transformed) as an independent variable and strength of the bias (measured as the ratio of space in front of the sitter to the space behind the sitter) as the dependent variable failed to show a significant effect of date on how marked the forward bias was.

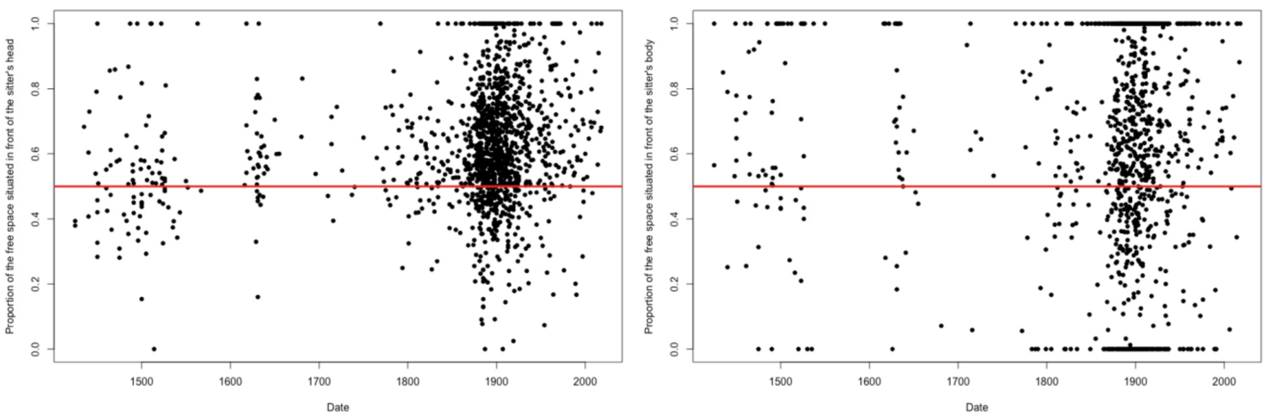


Figure 4. Measures of the forward bias (i.e., proportion of the free space that has been placed in front of the sitter)**, by date.** Head measures are presented on the left, and body measures on

the right). The red line at y = 0.5 indicates the ratio at which there is an equal amount of space in front and behind the sitter (above the line are all portraits showing a forward bias).

4. Discussion

We found evidence that the forward bias is at play in the spatial composition of human profile-oriented portraits. In two large and diverse datasets (total of 1831 paintings by 582 unique identified painters), painters tended to put more free space in front of, rather than behind, the sitters they depicted. Additionally, in both datasets, the sitters became more ex-centered over time, thus suggesting that this bias was modulated by historical trends and norms.

The widespread presence of a forward bias in both datasets was robust. This is particularly remarkable since it is goes against another known bias that favour centering sitters. The present study, using large-scale databases, is ideally suited for investigating production within ecological conditions. Future research could address whether this forward bias also impacts the processing and aesthetical appreciation of those portraits. It cannot be automatically assumed that biases in production are caused by biases in reception, even when both are supported by experimental studies. For instance, while there is evidence supporting a bias in favour of centering one of the sitter's eye in painters' *production* of portraits (Tyler, 1998), this couldn't be replicated in an aesthetic judgment task (McManus & Thomas, 2007).

As detailed in the introduction, several cognitive processes might participate in producing the bias we observe here. They range from preferences for specific pairings between spatial compositions and shapes to the role of an agent's direction of gaze in social cognition. The bias we observe could be due to the low-level processing of human shapes as oriented / directed stimuli, or to the processing of humans as agents with typical direction of movement, or finally, as an effect of gaze-monitoring systems. Future research should investigate which of these mechanisms drive the effect we observed.

Our dataset shows an increase in the number of portraits showing a forward bias and amplitude of this bias- portraits picture their sitters with an increasing proportion of their free space in front of their sitters. The fact that ex-centering in general increased can be interpreted as an effect of the relaxation of a norm of centering, which is itself a aspect of the general relaxation of norms of composition in the history of Western painting since the Renaissance (Puttfarken, 2000). More fine-grained possible historical factors regarding changes in the conventions, techniques, formats, and the social role of portrait painting should also be investigated. The increase in frequency and in amplitude of the forward bias calls for an explanation in terms of psychological factors, given that they were given greater sway with the relaxation of pictorial norms. Our dataset is not an exhaustive corpus, thus its evidential value depends of the degree to which the dataset is representative, a degree which we assume is positive but which we are not in a position to evaluate precisely.

Finally, this case study illustrates how cultural contexts may counteract, favour, or otherwise modulate the expression of cognitive processes in cultural productions.

Chapter 5 – The role of physical constraints in rhythms

1. Introduction

Skateboard decks' shapes have changed dramatically over the last 50 years or so: their evolution can be explained mostly by the skateboard decks adapting to new uses and performances (Prentiss, Skelton, Eldredge, & Quinn, 2011). Overall, it is a familiar idea that tools improve over time by becoming better adapted to perform a specific function or action. This increase in tool efficiency—and cultural practices in general—is claimed to occur through variation in production coupled with selective retention of the most successful variants, which is taken as characteristic of cumulative culture (Mesoudi & Thornton, 2018).

There are, however, clear examples where causation operates in reverse, from available material to new uses or end-products. Countless examples of such directionality can be found in the history of cooking, where causality flows from the material to the production. The availability of different raw materials – as well as heating sources or cooking instruments – can push cooking practices towards different types of dishes (see Wilson, 2012 for an overview). Before the advent of the large-scale circulation of humans and goods, medicine was also a cultural domain heavily influenced by local factors that determined both the pathologies humans had to face and their available pharmacopeia (Anyinam, 1995; E. Thomas et al., 2009).

This type of causal dependency from the available 'raw' material in the environment to performance or final cultural product, is understood within cultural attraction theory's ontology as an instance of ecological factors of attraction. These are defined as "those factors in the shared local environment that play a role in people's mental processes and in their interactions, and which are thus relevant to cultural dynamics. […] They include the biological and physical environment external to the organism (food and materials) and also behaviours and artefacts, including public representations such as speech, writings, and ritual performances, through which people interact with one another" (Scott-Phillips et al., 2018). Although there has been a growing corpus of research on cognitive factors of attraction (including, for instance, the direction of gaze in portraits, (Morin, 2013); or the wide-spread practice of bloodletting, (Miton et al., 2015), ecological factors of attraction still haven't been empirically investigated as such.

Universals have been the focus of a variety of major works and initiatives in studying musical production. Exploring both universality and diversity of musical productions was already at the heart of Lomax's *Cantometrics* projects (Lomax & Berkowitz, 1972), continued by more recent initiatives such as *CantoCore* (Savage, Merritt, Rzeszutek, & Brown, 2012). Focusing on universals in music has been a fruitful research avenue, including investigations of statistical universals in human music (Brown & Jordania, 2013; Savage, Brown, Sakai, & Currie, 2015), and of the human ability to identify the functions of songs even from cultures other than that of the participants (Mehr, Singh, York, Glowacki, & Krasnow, 2018). Such approaches have built on a more general scientific interest for the human species-wide capacity for music and its evolutionary basis (Brown, 2000; Cross & Morley, n.d.; Merker, Morley, & Zuidema, 2015; Ravignani, Bowling, & Fitch, 2014).

Previous experiments with artificial languages (Kirby et al., 2008, 2015) or musical sequences (Ravignani, Delgado, & Kirby, 2016) have used transmission chains (in combination or in addition to communication games) in order to demonstrate how structure can emerge from random experimental inputs. This cumulative emergence of structure through transmission can be understood, as a *subtractive* (Reindl & Tennie, 2018) ratchet effect, on the basis of the assumption that complexity gets reduced over generations with no decrease in performance (i.e., efficiency increases). By contrast, we start with a musical sequence that is as simple as possible, expecting transmission episodes to lead to stable, but also more complex, rhythms, resulting from biases created by kinematic patterns in a drumming task –which would demonstrate an *additive* ratchet effect.

Our aim in this experiment is not, however, to test for cumulative culture per se or claim that it is what we observe (Miton & Charbonneau, 2018). We do not, for instance, test whether participants' production will be out of reach of individuals performing repeatedly on their own. We similarly do not maintain intra-generational variation by having multiple ancestors at each step. Rather, this study is to be thought of as proof of concept for the role of physical constraints in creating stable cultural production. As such, it is closer *in logic* to previous experiments using diffusion chains to show how some priors in the participants' minds determine which content is stable through transmission (e.g., Kalish et al., 2007), with the notable qualification that we expect such biases to come from the physical arrangement and participants' movements rather than from a psychological prior. Motor constraints have been shown to influence melodic aspects of songs (predominance of arch-shaped and descending melodic contours in musical phrases, tendency for phrase-final notes to be relatively long, bias toward small pitch movements between adjacent notes in a melody) in both human and bird songs (Savage, Tierney, & Patel, 2017; Tierney, Russo, &

Patel, 2011), yet little is known about the impact of motor constraints on rhythmic aspects of music, which is the focus of this study.

We aim to test whether transmission chains can lead to different stable 'cultural items' (rhythmical sequences) from the same 'seed' (i.e., input given to the first generation of participants). If physical arrangements and motor constraints constitute a factor of attraction in this music-making task, then changes in physical arrangements across chains that start from the same seed should result in a pattern of divergence. The possibility that physical properties of tasks, and how they interact with human cognition, can shape specific characteristics of cultural items, although acknowledged in theoretical works, hasn't been explored empirically in cultural transmission studies. One recent experiment (Ravignani et al., 2016) focused on how transmission can magnify weak individual learning biases and thus produce outputs with universal features (such as an isochronous underlying beat, hierarchical organization of beats of unequal strength, and grouping of beats in groups of 2 or 3) from random, computer-generated, sequences.

Here, we test whether physical properties of a given task can predict which shape a cultural item will take through transmission chains. We test whether physical affordances and motor constraints can act as a factor of attraction, by predicting characteristics of participants' productions from parameters relating to the ease with which the movements required by the sequence can be produced.

This study departs from previous studies in two major aspects: (1) it is not aiming for a form of simplification (from random to structured output), and (2), it is not aiming for a convergence between chains (i.e., convergence on a 'universal-like' property of the production, e.g., integer-ratio intervals, (Jacoby & McDermott, 2017)). By contrast, this study predicts both an increase in complexity in some conditions (from a uniform beat to one including a structure mixing two different Inter Onset Intervals - henceforth IOIs) and a divergence, as the different conditions predict the emergence of different rhythmical sequences. We expect transmission, through our experimental design, to superimpose a quality (structure) to the initial content, rather than getting rid of artificial variability (coming from the random character of the seed in previous experiments).

2. Methods

Participants

A total of 120 participants (38 Male, 80 Female, 2 NA, mean age = 26, SD = 4.5) participated in this experiment. All participants were right-handed, and had no musical experience (they had neither learned to play an instrument, nor taken music lessons). All participants gave their informed consent and received gift vouchers as compensation. This research was approved by the United Ethical Review Committee for Research in Psychology (EPKEB) on the behalf of the Central European University, (ethics approval number 2018-18).

Sample size & sample size rationale

We collected five transmission chains of six participants (generations) per condition. Each participant took part in 24 trials. These sample sizes were decided and registered *a priori* on the basis of previous experiments and pre-registered [8], and are available here: https://osf.io/8p4mz/?view_only=ec41d8ed252d4cbf963aac4abc172a7b. Participants were randomly assigned to the four different conditions.

Stimuli

Seed and transmitted sequences

For the first generation, the seed - i.e., the first generation's input – was a regularly spaced beat, isochronous (metronome-like) sequence, of 13-beats. For all subsequent generations, the

[8] Time-stamped registration: https://osf.io/w5pmj/?view_only=686fb54a17354aaba44a88ddd2dec102

input consisted of whatever sequence was produced by the previous participant in the chain. All drum pads produce the same MIDI tone (pitch, timbre, duration).

The initial seed interval of evenly spaced beats (1:1:1) has two main advantages. For one, it represents a case of "extreme" simplicity: the sequence is as simple as possible to describe, as its description includes only one ITI, and the number of repetitions. Secondly, it is a rather infrequent interval in most music our participants might be familiar with, thus avoiding biased priors or showing strong cultural variation (Sadakata, Desain, & Honing, 2006; Sadakata, Ohgushi, & Desain, 2004).

The seed is played at a tempo of 120 BPM. This tempo was chosen based on the results from a pilot experiment which indicated that, with this tempo, it is possible to reproduce the given pattern, but the reproduction was not trivially easy with our task setup. Moreover, 120 BPM (beats per minute) or an IOI of 500 ms, is known to be a 'preferred' tempo for humans, being easy to process (Moelants, 2002) and used in a variety of tasks, including serial interval production (Collyer, Broadbent, & Church, 1994).

Physical setup

Participants were given headphones, one single drumstick, and a set of three independent Millenium drum pads connected to a MacBook pro laptop via a trigger box (ddrum DDTi) that sent midi notes. The three drum pads were evenly spaced. We can refer to Fitts' law (Fitts, 1954) to have an approximation of how difficult the movements were for the different distances, but overall, our large movement had to cover twice the distance covered by small movements. All three drum pads used produced the same sound (same pitch) – they were all set on a percussion sound with a sharp onset (MIDI Note 60 from the standard Mac OS sound bank).

Procedure

Participants were asked to recreate a pattern of sounds by tapping different drum pads in a given order. The experimental design used a linear transmission chain method, i.e., with output from a participant serving as input for the next participant, akin to the telephone game. Participants were explicitly asked to reproduce the audio they heard *as faithfully as possible*, with these instructions:

“In the following experiment you will be asked to reproduce a musical pattern by hitting drum pads with a stick. For about 10 seconds you will hear clicks. Once the clicks stop, please try to reproduce as faithfully as possible what you have heard (volume and rhythm). You will be asked to tap on all three drum pads, in a given direction. Please stop tapping once you’ve heard a different, cymbal sound.”

They heard a sequence of 13 taps, which was either a metronome (for the first participant in the chain), or a sequence produced by the previous participant in a chain (for participants in generations 2 to 6), which they then had to reproduce using the drumstick and the drum pads in front of them. Depending on the condition they were assigned to, they either had to produce only movements of the same amplitude (only small or only large movements), or a mix of large and small movements, in a different order in each condition (see Figure 2). They had to listen and reproduce 24 trials one after another.

Each sequence was recorded and given to the next participant in the chain. Participants were unaware they would be listening to stimuli produced by a previous learner. All sequences produced by one participant were transmitted to the next participant in the chain, with no change and in their entirety. The order in which sequences were presented to participants was randomized at each transmission step. We recorded all taps produced by the participant, with their timestamps (from which we get Inter Onset Intervals - IOIs) and velocity (ranging between 0 and 127).

After the behavioural task, participants completed a short questionnaire with the following questions: (1) How difficult was the task? (Answered on a scale from 0, very easy, to 7, very difficult), and (2) Do you have any musical experience? (Classes, played an instrument?) If yes, please specify.

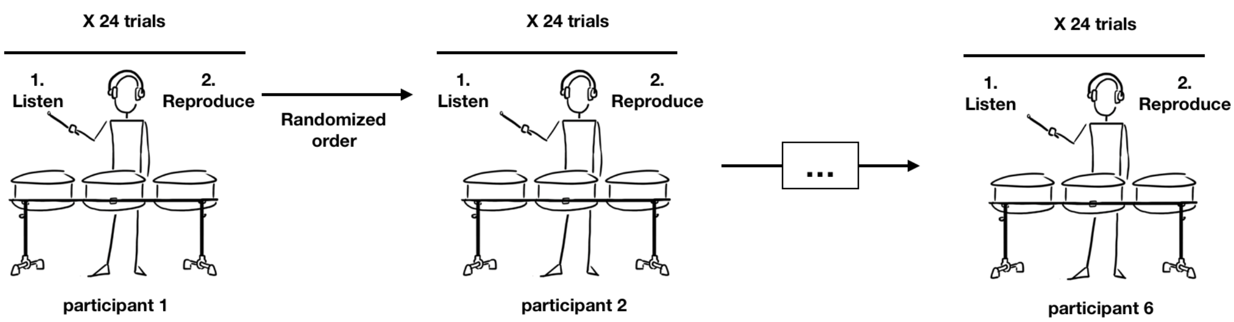


Figure 1. A schematic description of the experimental procedure. The participants first heard a sequence of 13 taps. This sequence was either a metronome-like sequence (for participants in the

first generation) or a sequence produced by the previous participant in the chain (for participants in generations 2 to 6).

The physical apparatus and the initial sequence were the same in all four conditions. The study had a between-subjects design such that each participant took part in only one of the four conditions. As displayed in Figure 2, our physical setup made kinematic patterns vary along two dimensions: (1) whether all movements in the sequence are the same or not, and (2) at which point in the sequence the movement requiring a larger amplitude occurs (all movements, none of the movements, as the first or as the last of three movements).

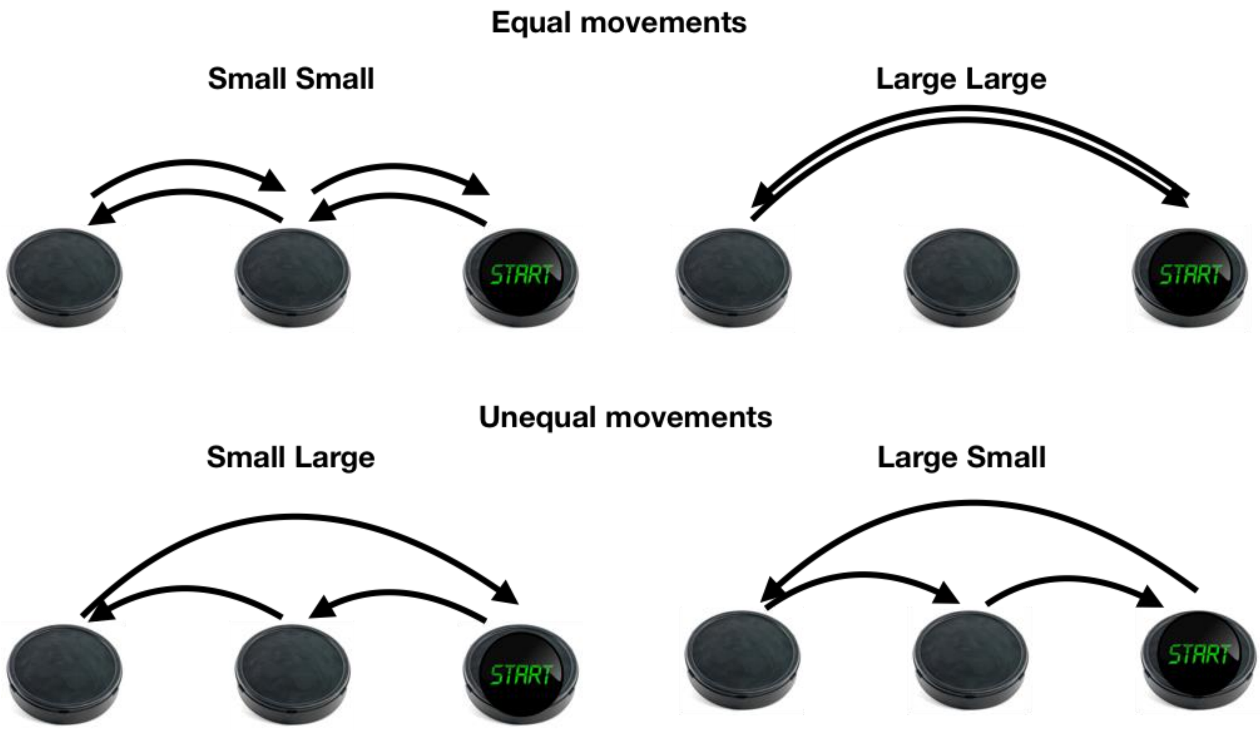


Figure 2. Depictions of our four conditions. All conditions started by tapping on the right pad. The upper row depicts the two conditions with all movements being of equal lengths, and the lower row depicts the two conditions that include a mix of large and small movements.

3. Hypotheses

In this experiment, we test whether different motor constraints cause participants to produce different rhythms. We manipulate the order in which the drums have to be hit, and as a consequence, the amplitude of the movements required to produce taps. As the movement required to produce a given tap over a drumming tasks becomes larger, it becomes harder to produce short IOIs, and easier to produce longer ones (Fitts, 1954). We thus predict that the difference in kinematics between the different conditions will lead to qualitatively different productions. In conditions in which all movements are of the same length, the rhythm of the sequences produced should remain isochronous, while conditions in which one movement is larger than others should move away from isochronous rhythms (**hypothesis 1**). In conditions in which all movements are of the same length, the amplitude of the movement would predict how long the IOIs are: shorter in the condition which includes only small movements than in the condition which includes only large movements (**hypothesis 2**). In conditions in which one movement is larger than the other ones included in the sequence, we can predict which IOI will be longer: it should be the first one in the sequence whenever the larger movement occurs first, and the third in the sequence whenever the larger movement occurs third (**hypothesis 3**). Together these three hypotheses predict, with precision, which rhythmical sequence participants at the end of the chains will produce based on the physical constraints they encounter in each condition.

Finally, we expect to have (1) different (between conditions) and (2) stable rhythms by the end of our transmission chains. The overall predicted pattern is one of divergence, i.e., the chains from different conditions will be less and less alike over time (i.e., experimental generations -**hypothesis 4**). The predicted emergence of stability means that we expect the amount of change –i.e., copying errors – to decrease through the chain (which is exactly the same as learnability or copying accuracy increasing) (**hypothesis 5**). This has been observed in previous experiments as well, e.g. by Ravignani et al (2016), and is usually interpreted as an increasing match between what participants have to reproduce and their own biases. We expect this type of gradual change to occur during our experiment as well, in all our conditions.

4. Results

Details on data analyses

Whenever a sequence was missing or had technical problems, one other trial among the 23 available ones from the same participant was randomly selected and passed on. This ensured that all participants went through the same number of trials, and no participants who had such doubles as input sequences noticed that they were actually identical. The total number of trials with problems (not recorded / recording not viable due to software issues) was 22 (over a total of 480 trials, i.e., amounting to 4.58% of the total number of trials, and distributed over all four conditions) – see Figure 3.

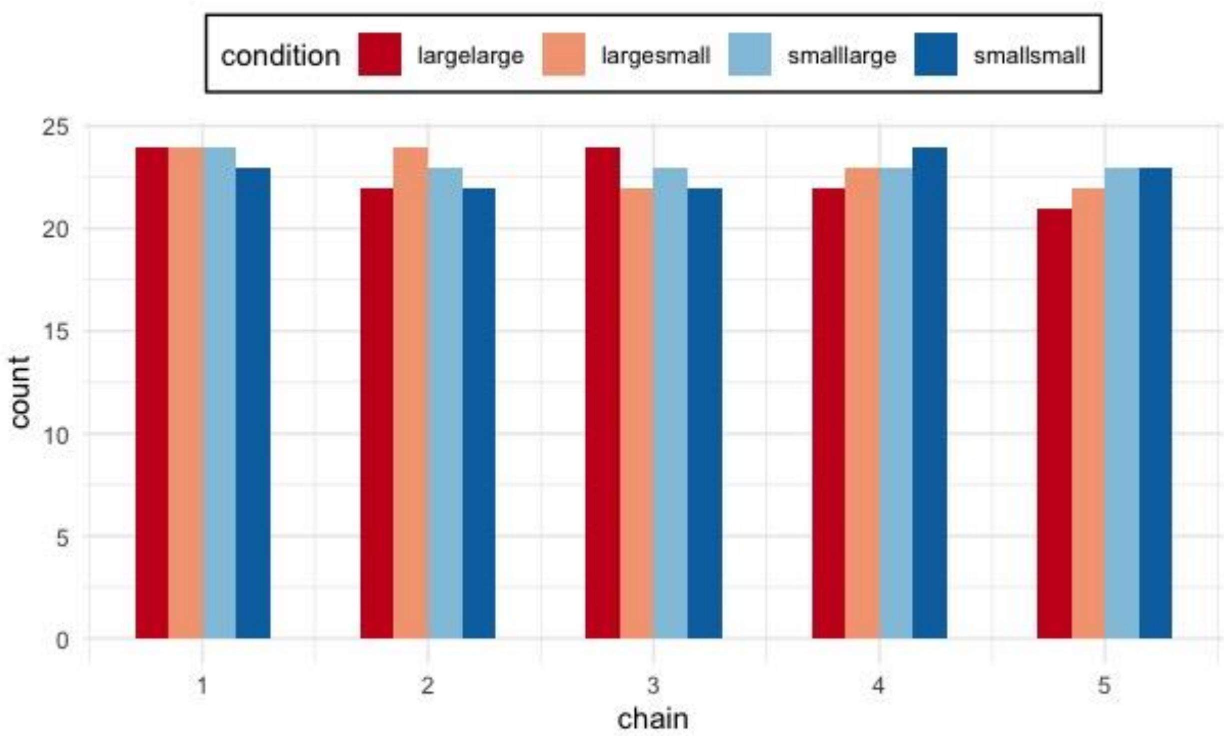


Figure 3. Number of complete chains of trials (there were 24 trials within each chain of each condition) included by condition and chain. No chain of participants was missing more than 3 trials.

Although participants had to reproduce a sequence of 13 taps, our predictions bear mainly on the central part of the sequence (apart from the first and last kinematic round), i.e., from the 4^{th} to the 10^{th} taps, which are not influenced by either the first or the last tap.

In order to illustrate our results, we created audio files that reflect the mean sequence produced by participants of each condition, at the first and the last generation, which can be listened to at https://osf.io/8p4mz/?view_only=ec41d8ed252d4cbf963aac4abc172a7b. These audio files were produced by averaging each IOI (out of the 12 included in each sequence of 13 taps) from all the sequences produced by all participants from all chains of the same condition at the same generation. An audio file is also available for the sequence used as the seed.

Hypothesis 1: All movements equal vs. not all movements equal

We predicted that both conditions including all equal movements (LARGE LARGE and SMALL SMALL) would show an increase in complexity under the form of a bimodal distribution of IOIs (i.e., a non-isochronous rhythmical sequence), whereas both conditions with not all equal movements (LARGE SMALL and SMALL LARGE conditions) wouldn't (i.e., they would produce non isochronous rhythmical sequences).

Visual inspection confirms that distributions of IOI in the conditions with two types of movements tended to become bimodal, whereas it wasn't the case for conditions in which all movements were of the same type (Figure 4).

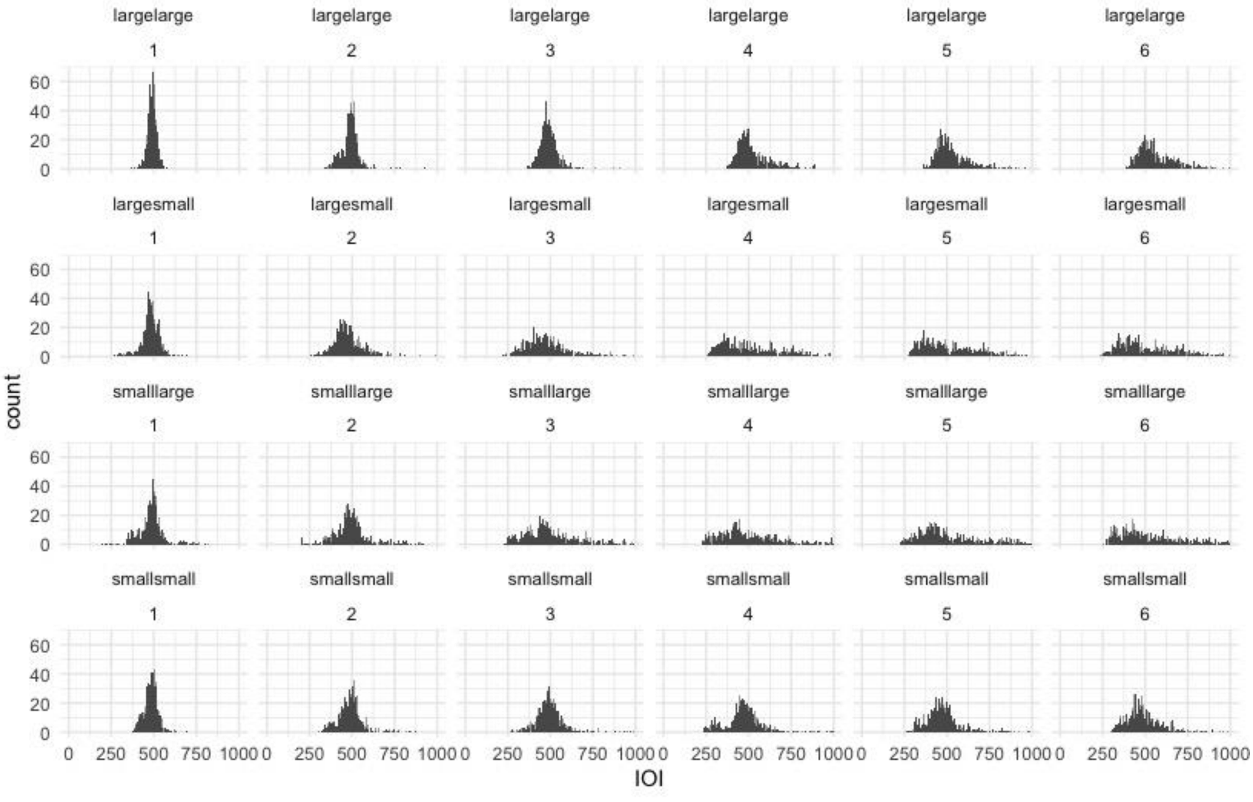


Figure 4. Histogram of IOIs per condition and generation.

Rhythmical structure was assessed using normalized pairwise calculations (nPC)(Condit-Schultz, 2019; Toussaint, 2012).

$$nPC = 200 * \left| \frac{antecedent\ IOI - consequent\ IOI}{antecedent\ IOI + consequent\ IOI} \right|$$

Visual inspection suggests that the distribution of nPC became bimodal for both conditions that mixed movements of both amplitude (i.e., LARGE SMALL and SMALL LARGE), but that this was not the case for conditions that included only movements of the same amplitude (LARGE LARGE and SMALL SMALL) – see Figure 5.

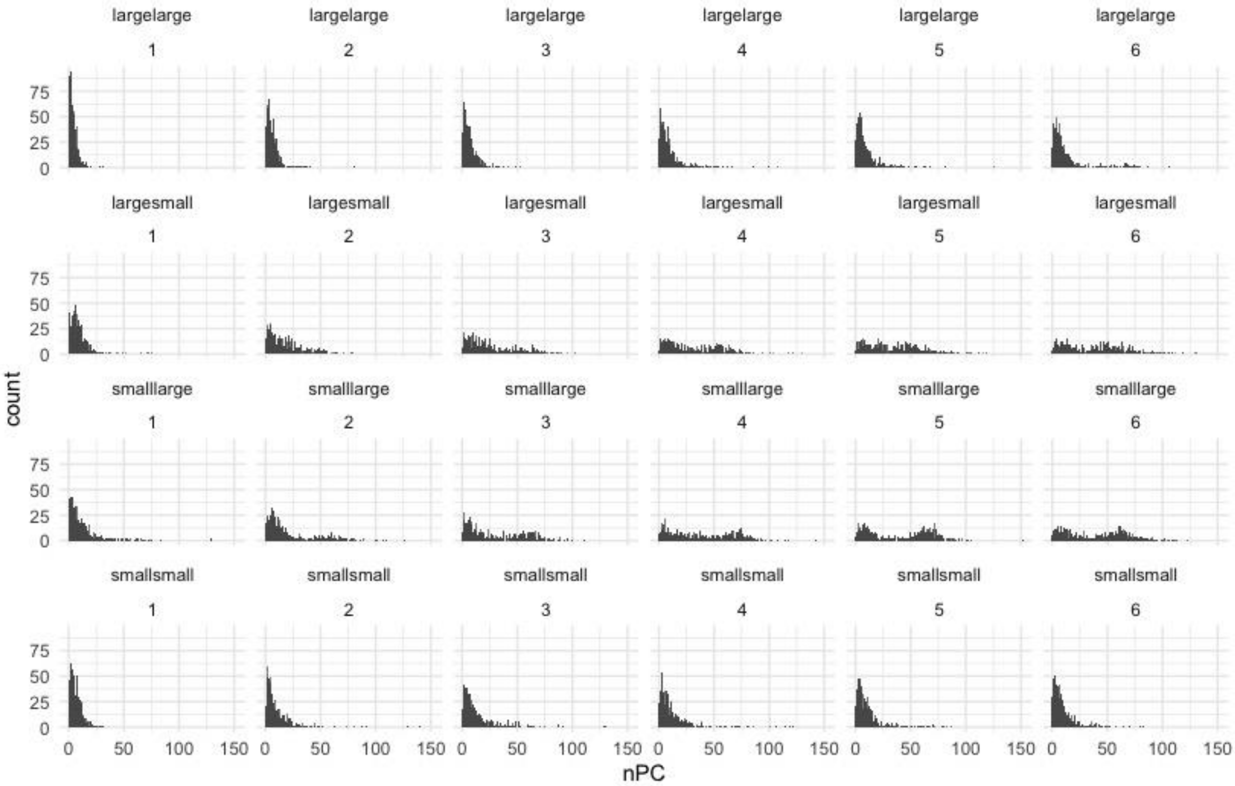


Figure 5. Histogram of nPCs per condition and generation.

In order to test for a difference in the types of rhythm, we computed the normalized pairwise variability index (nPVI, see below) for each sequence produced by participants.

$$nPVI = \frac{100}{m-1} * \sum_{k=1}^{m-1} \left| \frac{IOI_k - IOI_{k+1}}{\left(\frac{IOI_k + IOI_{k+1}}{2}\right)} \right|$$

The normalized pairwise variability index is a measure that allows for a minimal value of 0 when all IOIs are equal, and increases as a sequence gets more unequal IOIs. This distribution of nPVIs is used to test whether there is a change from the seed (i.e., metronome sequence): any divergence from this rhythm translates to an increase of the nPVIs. Overall, nPVI increased for both conditions with only one type of movement amplitude ($L = 829$, $k = 6$, $N = 10$, $p <.001$ including the first generation, $L = 478$, $k = 5$, $N = 10$, $p = 0.041$ excluding the first generation) and unequal conditions ($L = 870$, $k = 6$, $N = 10$, $p = 0.041$ including the first generation, $L = 511$, $k = 5$, $N = 10$, $p <.001$excluding the first generation).

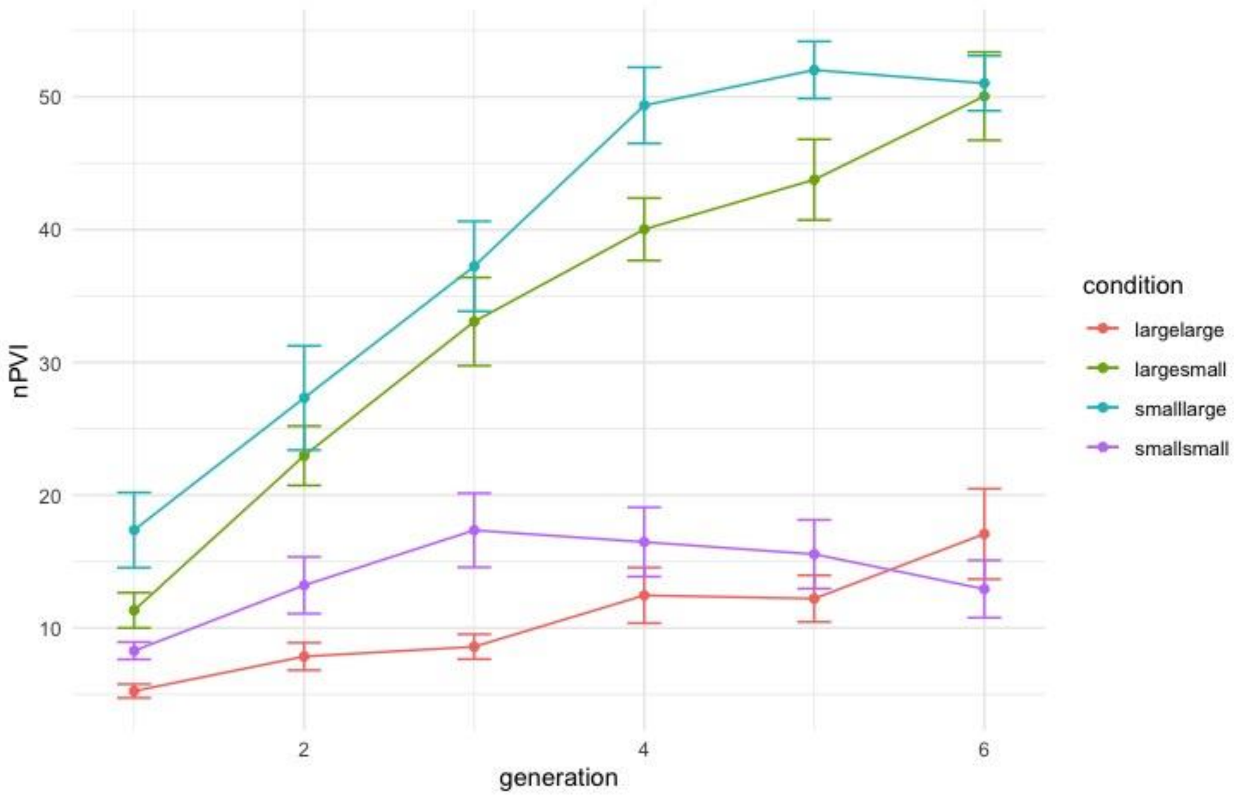


Figure 6. Normalized pairwise variability index (nPVI) by generation, colour represents the different conditions. Error bars represent standard 95% confidence intervals.

A Kolmogorov-Smirnov distance on distribution of nPVI at the last generation confirmed that the equal movement conditions (LARGE LARGE and SMALL SMALL) had a different nPVI from the unequal movement conditions (LARGE SMALL and SMALL LARGE), D = 0.84, p < .001. A t-test at the final generation suggested that unequal movement conditions (M = 50.54, SD = 14.96) had higher nPVIs than equal movement conditions (M = 15, SD = 15.41), t(455.00) = 25.03, *p* < .001, *d* = 2.34.

Hypothesis 2: Small versus large movements

We predicted that both conditions with all movements equal (LARGE LARGE and SMALL SMALL) would show isochronous rhythms, but with different IOIs (SMALL SMALL should have shorter IOIs than LARGE LARGE). A t-test at the final generation indicated that the SMALL SMALL condition (M = 494 ms, SD = 150 ms) had shorter ITIs than the LARGE LARGE condition (M = 582 ms, SD = 129 ms; t(1327.67) = 11.66, p <.001, d = 0.63), see Figure 4. IOIs were not normally distributed (Shapiro Wilk: W = 0.817, p < .001), but the difference between the IOI produced in both conditions were also confirmed by a Mann Whitney U test (U = 332850, p < .001): IOI produced in the LARGE LARGE condition (Med = 539 ms) were larger than the ones produced in the SMALL SMALL condition (Med = 472.5 ms).

A mixed-effects model [9], including condition and generation as main effects, and participants nested by chain as a random effect, showed that this pattern emerged over time There was a significant interaction effect between condition and generation (β = -18.991, SE = 9.387, t(55.971) = -2.023, p = .0479), indicating that as generations passed, the difference in IOI between the LARGE LARGE and the SMALL SMALL conditions increased. There was also a significant effect of generation (β = 19.791, SE = 6.638, t(55.984) = 2.982, p = .004), but not of condition (p = .28) - see Figure 7.

[9] This mixed effects model, as all others from this chapter use the package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015).

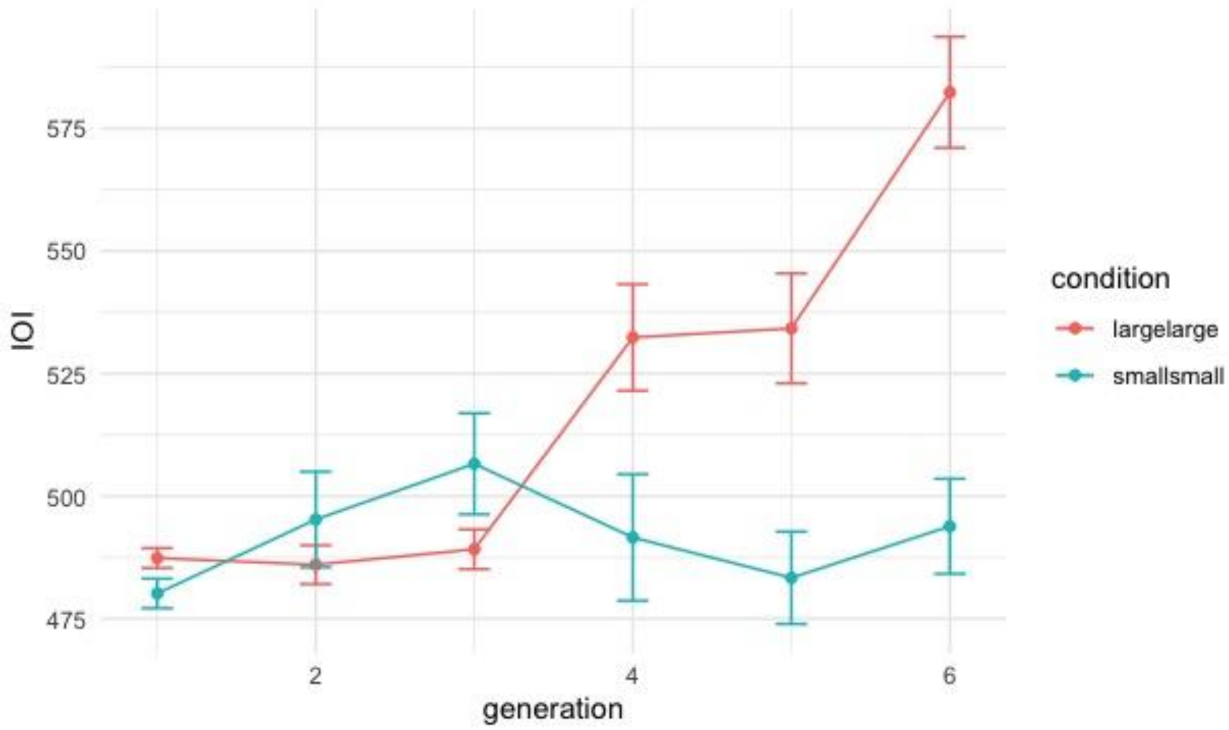


Figure 7. Mean InterOnset Intervals (IOIs) by condition (LARGE LARGE or SMALL SMALL) and generation (first to sixth). Error bars represent 95% confidence intervals.

Hypothesis 3: Large movement as the first or the third of the sequence

We predicted that while conditions that included both types of movements would show non-isochronous rhythms, they would have longer IOIs in different places. This longer IOI (out of three) should occur first when the large movement occurs first in the sequence (LARGE SMALL condition), and third when the large movement occurs last in the sequence (SMALL LARGE consequence).

We predicted that the condition (SMALL LARGE or LARGE SMALL) to impact which position in the sequence (i.e., MapIOI_Index) is associated with longer IOIs. We should observe an interaction effect between condition and MapIOI_Index, which reflects the order in a sequence. A sequence was understood as three consequent IOIs, to reflect the cycle of movements. MapIOI_Index could take the values 1, 2 or 3. Because we analysed sequences of two such cycles (taps 4 to 10, i.e., 6 IOIs), there were two IOIs per position in the cycle of movements per trial.

Condition and Order in the sequence (MapIOI_Index) were used as fixed effects, and participant nested by chain were used as random effects. On the last generation, the mixed effects model revealed significant effects of both order in sequence (β = -84.50, SE = 7.705, t(1374) = -10.967, p < .001), and condition (i.e., SMALL LARGE, compared to LARGE SMALL – β = -521.824, SE = 57.073, t(10.952) = -9.143, p < .001). An interaction effect between condition and position of the Tap confirmed our prediction (β = 279.06, SE = 10.873, t(1374) = 25.666, p < .001), see Figure 8.

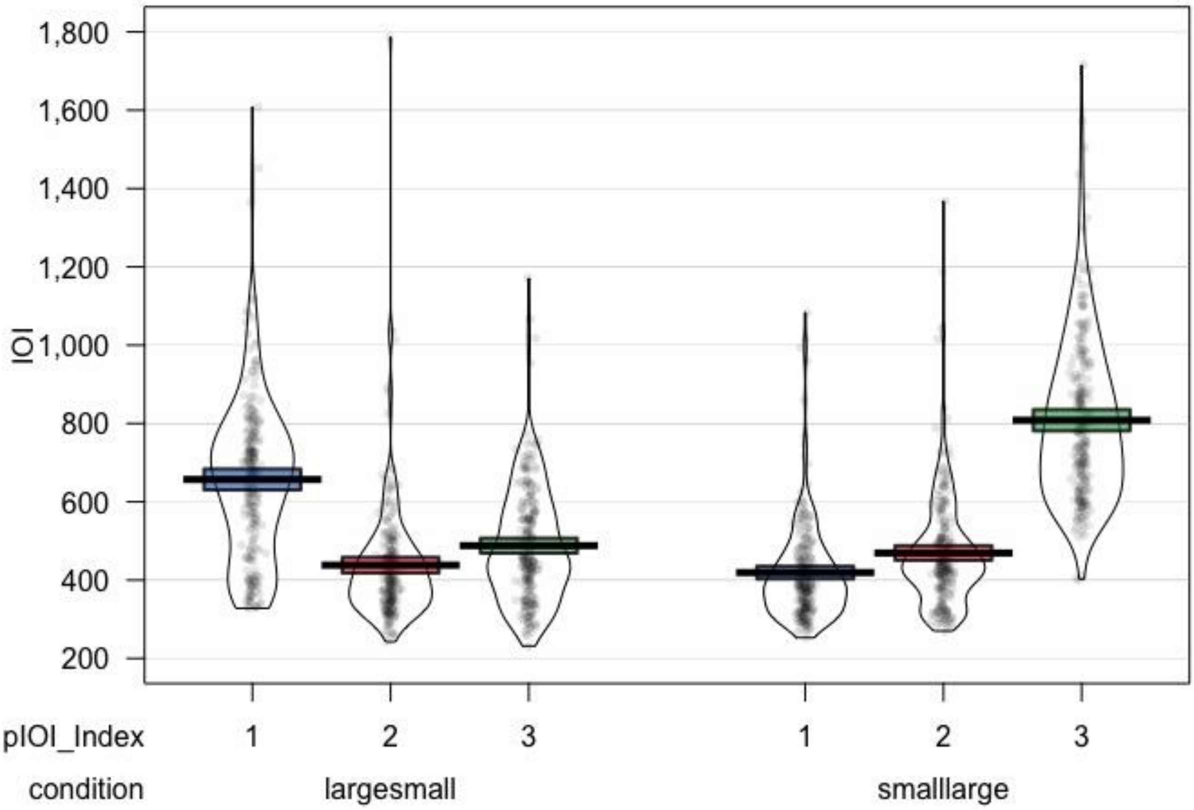


Figure 8. IOI by their position in the sequence, by condition, at the last generation. The coloured bands represent the 95% confidence intervals, points the raw data, and the external curve is the distribution's density.

A similar mixed effects model was run on data including all six generations, with generation as a main effect. Results of this model suggested that the difference emerged over the course of the experiment, as we observed a three-way interaction effect between condition, generation and order in the sequence (β = 41.734, SE = 2.173, t(8252) = 19.208, p < .001). This means that the effects of condition and order in the sequence became stronger as the generations passed. The mixed effects model also included significant effects of condition ($\beta = -94.90$, $SE = 44.23$, $t(76.884) = -2.146$, $p = 0.035$), generation ($\beta = 32.352$, $SE = 8.034$, $t(76.999) = 4.027$, $p < .001$), and order in sequence ($\beta = -28.933$, $SE = 5.996$, $t(8252) = -4.825$, $p < .001$), as well as interaction effects between condition and generation ($\beta = -80.605$, $SE = 11.357$, $t(76.884) = -7.097$, $p < .001$), between order in sequence and condition ($\beta = 53.978$, $SE = 8.462$, $t(8252) = 6.379$, $p < .001$), and finally order in sequence and generation ($\beta = -10.658$, $SE = 1.540$, $t(8252) = -6.922$, $p < .001$).

Hypothesis 4: Divergence

We predicted that chains become increasingly different between conditions. We tested this prediction by calculating the Jensen-Shannon divergence (JSD) on the list of IOIs. Here, we prefer the JSD as a measure of distance to the Kolmogorov-Smirnov equivalent because of its sensitivity to the order of the IOIs. The JSD was calculated between each trial to each trial from other conditions, at each generation. The average distance of a chain to other chains that aren't from the same condition (the divergence between conditions) increased over time. A Page trend test confirmed that the JSD between conditions increased over generations, whether we included the first generation ($L = 1778$, $k = 6$, $N = 20$, $p < .001$) or not ($L = 1059$, $k = 5$, $N = 20$, $p <.001$), see Figure 9.

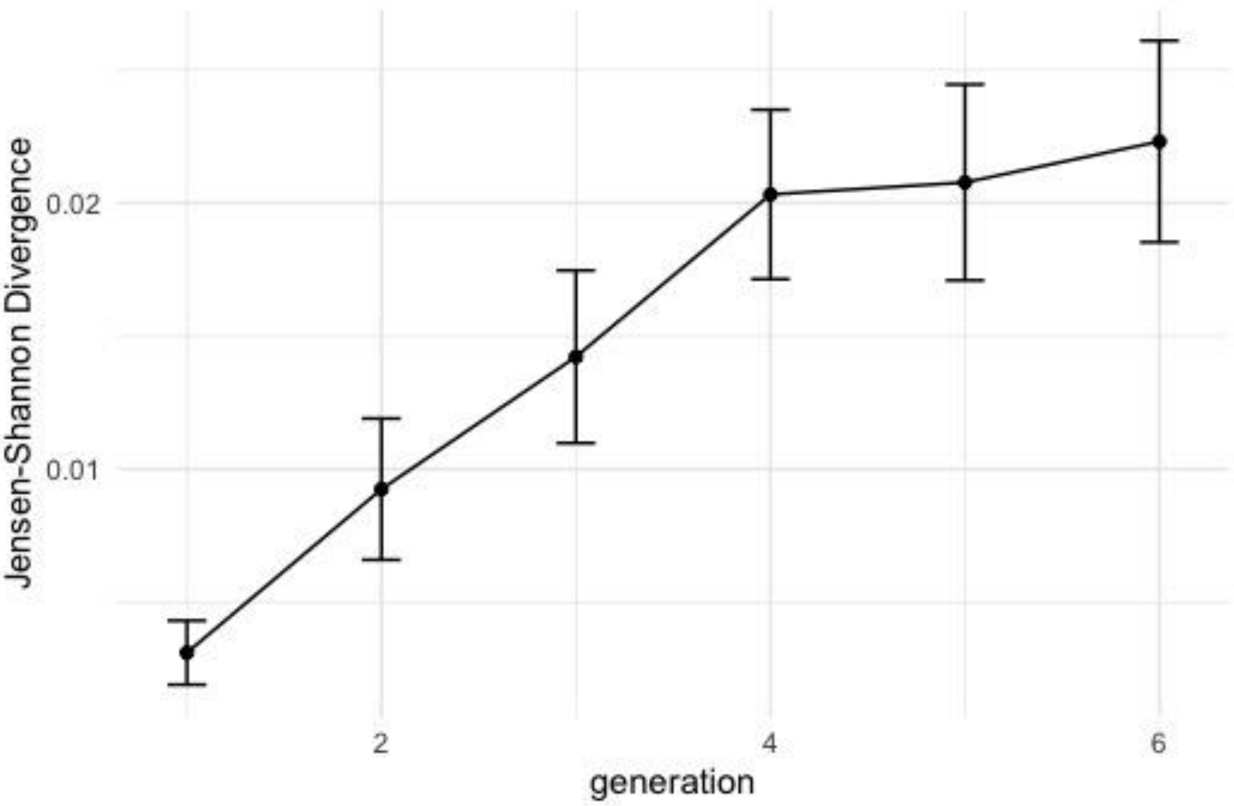


Figure 9. JSD calculated between each trial and all trials from different conditions at each generation, by generation. Error bars represent the 95% confidence intervals.

Another way to visualize such differences is to use ternary plots. We plotted the IOIs on a triangular simplex, such that each side of the simplex represents either the first IOI of the sequence, the second one, or the third one. As our design includes a cycle of three movements (3 IOIs, produced from 4 taps), this is particularly fitting and allows us to have a quick, visualization-based idea of how the rhythms evolved in the different conditions (Figure 10).

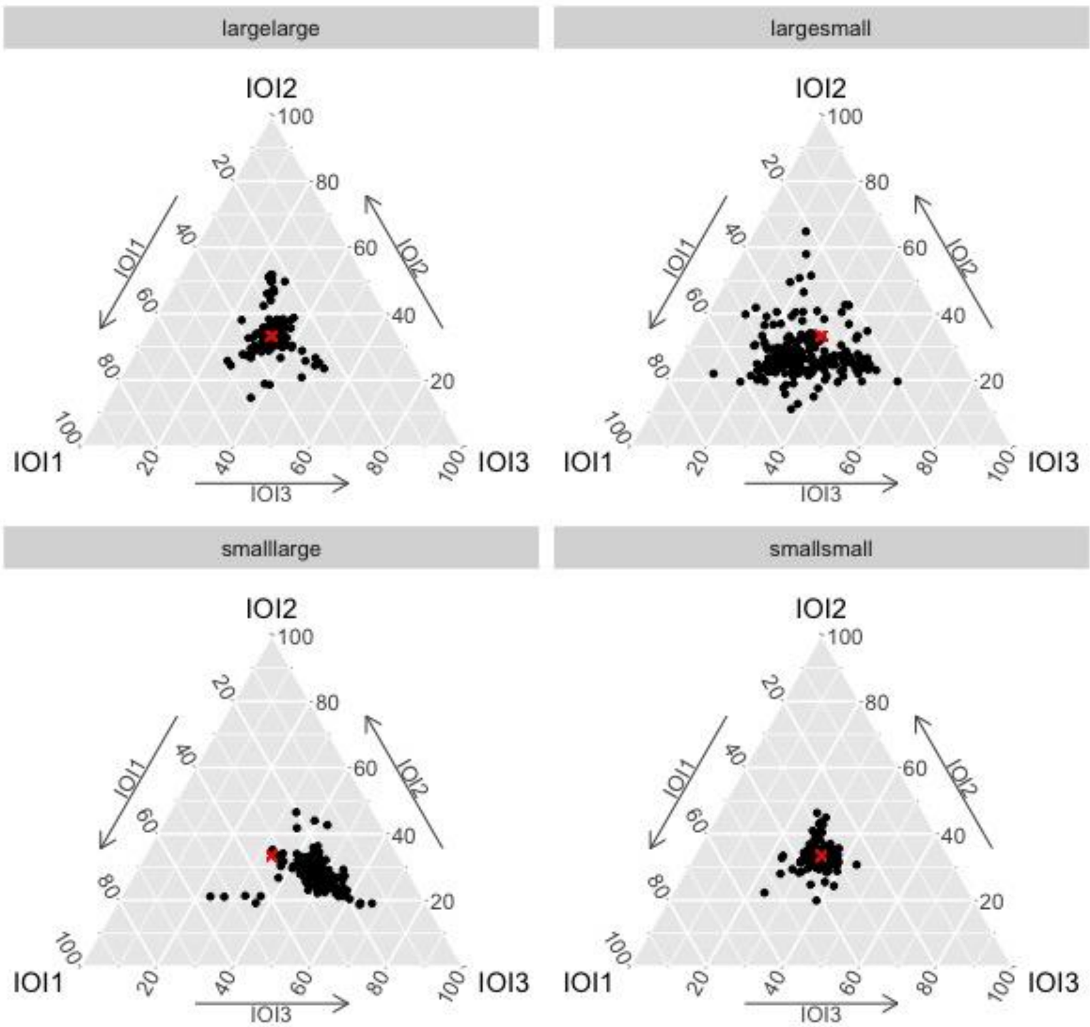


Figure 10. Ternary plots of the distribution of IOIs at the last generation, for each condition. The red cross indicates where is the 1:1:1 integer ratio (i.e., the metronome-like sequence with which the chains were seeded).

Hypothesis 5: Stability

Edit Time distance

We predicted that our experiment would produce *stable* rhythms, and that the rhythms produced by participants would become easier to reproduce. This is usually operationalized by using the (edit) time distance (Ravignani et al., 2016). The (edit) time distance from a participant to the next for increasing learning is defined by Ravignani et al. (2016)as the "total cost of the minimal cost set of substitutions, insertions or deletions among IOIs necessary to transform the pattern of durations a participant has heard into the pattern they have reproduced, where the edit costs are taken to be the absolute difference in time between duration". It is the same as copying accuracy in Jacoby & McDermott, 2017, i.e., the "distance between stimulus and reproduction".

This time distance increased over time, as confirmed by a Page Trend test, both when including the first generation (L = 1676, k = 6, N = 20, p <.001), or excluding the first generation (L = 985, k = 5, N = 20, p < .001), see Figure 11.

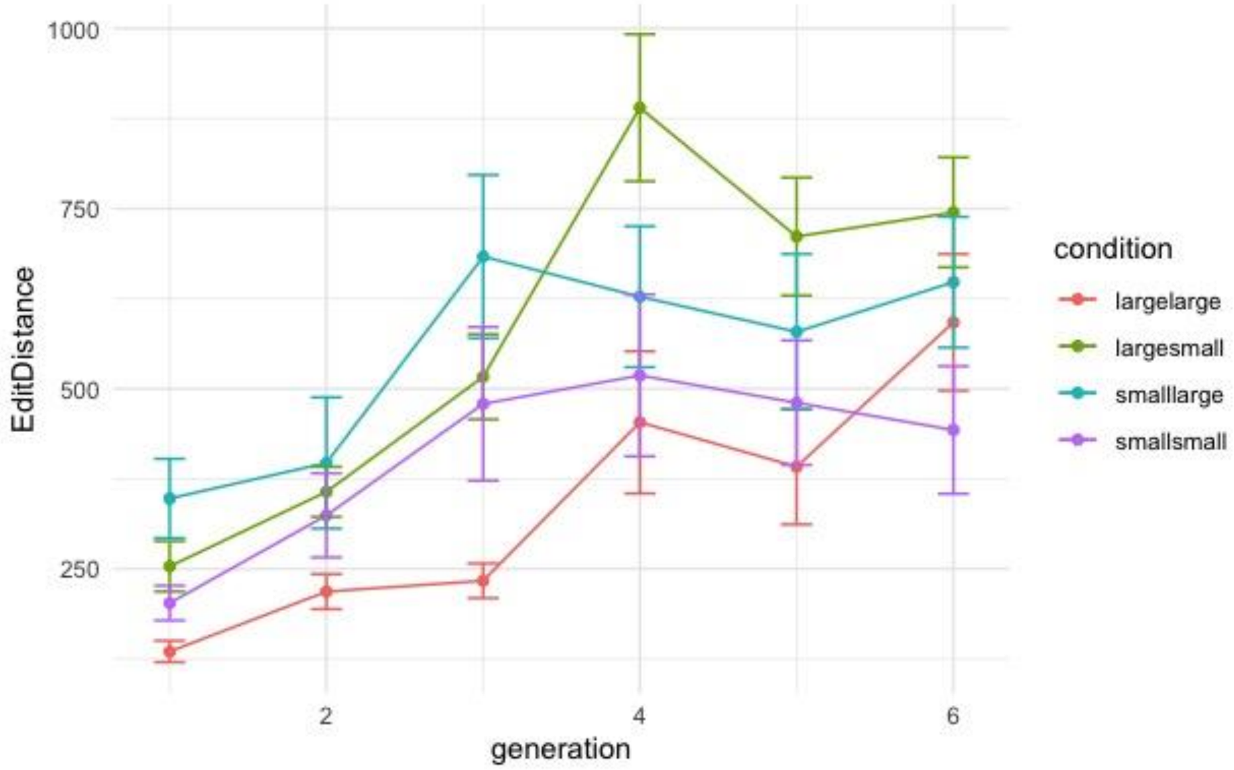


Figure 11. Edit time distance, by generation and condition. Error bars represent 95% confidence intervals.

Stability of the rhythmical structure: Edit nPC distance

Although the edit time distance is a straightforward and commonly used measure in the literature (Ravignani et al., 2016), it reflects similarity in terms of timing, but not necessarily in terms of rhythm. Edit time distance could show relatively high amounts of differences between two sequences that still have the exact same rhythmical structure. For instance, let's compare two sequences to a sequence of 4 IOI of 500 ms in a row (500 500 500 500). Both the sequence 520 580 500 600 and the sequence 550 550 550 550 would have the same time edit distance to it– yet, the latter's rhythm is much closer (i.e., equal IOIs) than the former (unequal IOIs) from the original sequence. Our prediction bears on the rhythmical sequence becoming stable – i.e., we predict a decrease in the quantity of change to the relation between IOI, not necessarily in the IOI themselves. We thus decided to create an nPC edit distance, which can be defined as the total cost of the minimal cost set of substitutions, insertions or deletions among nPCs necessary to transform the pattern of nPCs a participant has heard into the pattern they have reproduced.

Here too, the distance increased over time, as confirmed by a Page Trend test, both when including the first generation ($L = 1668$, $k = 6$, $N = 20$, $p < .001$) or excluding it ($L = 978$, $k = 5$, $N = 20$, $p < .001$), see Figure 12.

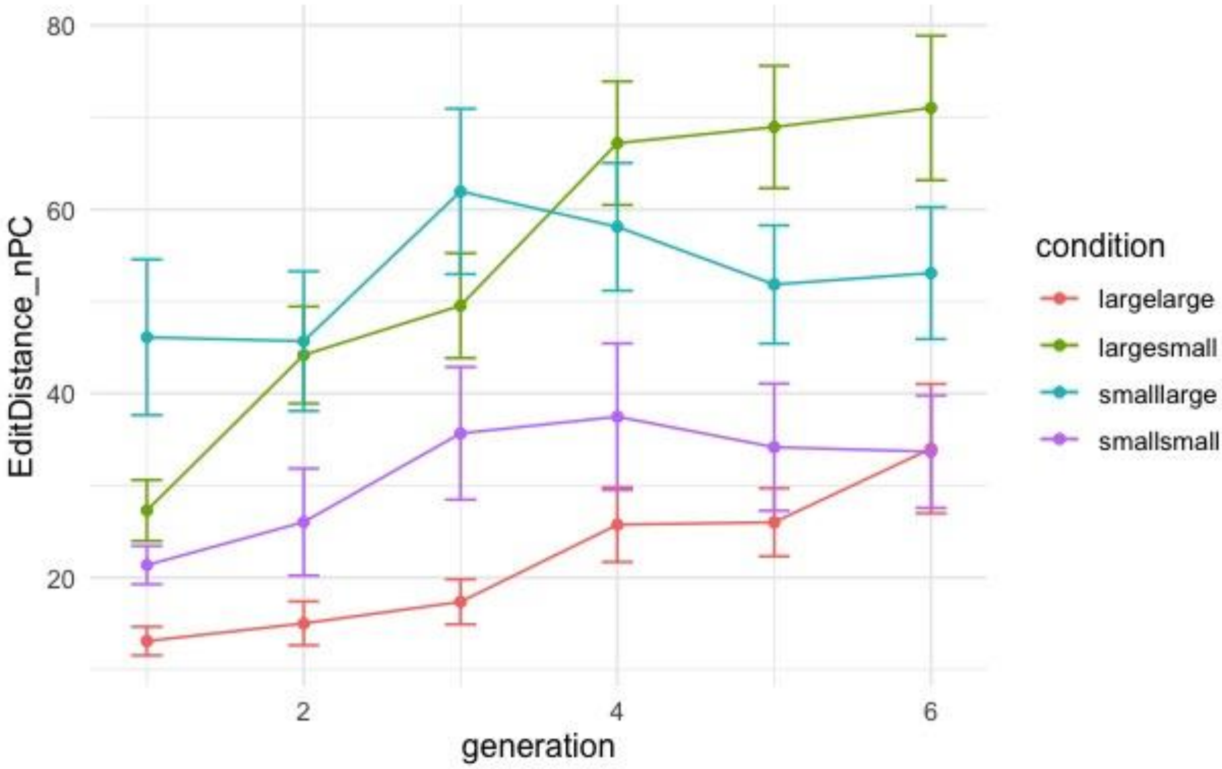


Figure 12. Edit distance based on nPC, by generation and condition. Error bars represent 95% confidence interval.

Relation between conditions and stability

As this pattern was rather unexpected – we predicted an increase in learnability, not an increase in the amount of change - we explored whether this effect was also driven by our conditions.

We ran a mixed effects model with edit distances as the dependent variable, condition and generation as fixed effects (independent variables), and participant and chain as nested random effects.

On the edit time distance, this mixed effects model confirmed that the edit time distance increased with generation ($\beta = 88.71$, $SE = 25.17$, $t(112.29) = 3.524$, $p < .001$). The only condition to significantly depart from LARGE LARGE (our baseline) was SMALL LARGE ($\beta = 316.95$, $SE = 138.52$, $t(111.93) = 138.52$, $p = 0.024$). No other main or interaction effect was significant (all *p*s > .141).

On the edit nPC distance, this mixed effects model revealed a significant effect of generation ($\beta = 4.269$, $SE = 1.736$, $t(112.787) = 2.459$, $p = 0.0155$), and of the SMALL LARGE ($\beta = 40.689$, $SE = 9.552$, $t(112.313) = 4.260$, $p <.001$), but not of LARGE SMALL ($\beta = 16.578$, $SE = 9.556$, $t(112.472) = 1.735$, $p = 0.0855$) conditions, but not of SMALL SMALL condition (p >. 1086). These results suggest that conditions including both amplitudes of movements led to higher edit nPC distances, i.e., the difference between what participants heard and produced was higher in those conditions than in conditions including only small or only large movements. There was a trend for an additional interaction effect between generation and the LARGE SMALL condition ($\beta = 4.621$, $SE = 2.454$, $t(112.472) = 1.883$, $p = 0.0623$), but no effect was significant (all *p*s >. 1086). Edit distances depended, to some extent, on both generation and condition (more so when analyses included all 13 taps – see Appendix E).

Relation to difficulty ratings

Overall, participants judged the experiment to be roughly mid-way between very easy (0) and very difficult (7), with a mean of 4.19 ($SD = 2.21$). Difficulty ratings increased over generations: the first generation rated the task as relatively easy ($M = 1.5$, $SD = 1.70$) while it was considered more difficult by participants in the last generation ($M = 4.4$, $SD = 1.82$; $t(37.83) = -5.21$, $p < .001$, $d = -1.65$).

5. Discussion

This study had four different conditions, which differ in the exact sequence of movements required to reproduce a rhythmical sequence. All conditions started with the same seed: an isochronous sequence of 13 taps equally spaced by 500 ms. We predicted that differences in physical affordances and movements produced by participants (an ecological factor of attraction) would determine which rhythmical sequences were produced in the different conditions (different attractors). This study follows the logic of creating an out-of-equilibrium system: the different conditions could either fit or not fit with the rhythm of the seed. Conditions that included a mix of movements didn't fit well with the initial sequence, unlike conditions with only one type of movement. Physical constraints and their fit with different rhythmical sequences influenced both how much transformation took place and which productions were stable. Physical constraints had an effect on three features: (1) whether the rhythm produced would be isochronous or not; (2) the length of the IOIs; and (3) the locations in a musical sequence at which longer or shorter IOIs occur. The differences in motor affordances created an overall pattern of divergence: the distance between rhythmical sequences produced by participants across different conditions increased over experimental generations. This study provides a proof of concept of the influence of a type of causal factor – an 'ecological' factor of attraction - instantiated by physical affordances in this particular case.

Surprisingly, we did not find an increase in learnability. Both measures of change between what was heard by participants and what they produced (either on the ITIs directly or a measure of rhythmical structure – nPCs) did not show a decrease over generations. The physical constraints impacted not only which rhythms were produced in our experimental populations, but also how stable these rhythms were.

Cognitive factors might have also played a role in at least some of our results. Most of the IOIs produced by the participants remained around 500ms / 120bpm (cf. Figure 3). For the two conditions including movements of different lengths, the rhythmical sequences produced by participants moved closer to the next possible integer ratio (1:1:2 or 2:1:1, equivalent to an nPVI of 66.67). This could also correspond to an attractor, since such small integer ratios are known as a universal feature of human songs (Savage et al., 2015). Participants tend to converge towards these ratios after a few iterations of reproducing an initially random rhythmical sequence (Jacoby & McDermott, 2017).

A great deal of work has focused on cognitive aspects of music-making (e.g., Jacoby & McDermott, 2017). The present study provides a proof of concept of the role of physical constraints in musical production. If such constraints have played a role in the cultural evolution of musical practices, this should be observable through patterns of co-evolution of music and instruments. Thus our results may have implications for the study of the evolution of instruments.

Evolutionary changes should flow in both directions: from the instruments to the music produced, and from the music to be produced to the instruments. The evolution of violins might provide a relevant example : When female voices became more popular in Baroque music, violins were made to sound closer to such voices (Tai, Shen, Lin, & Chung, 2018). In our experiment, all three drum pads produced the same sound. In actual instruments, on the other hand, we should expect to see the physical characteristics of instruments to be manipulated in relation to sounds' properties. Instruments themselves might reflect organisations making them easier to use, depending on what – the way some string instruments, like guitars, can be tuned so that some melodies become easier to play.

Finally, there are a number of features of the present study that contrast with music-making in the wild. This study recruited novice participants, who are less able to adapt to different physical constraints than experts. The question of the extent to which expert performance is influenced by physical constraints remains. Even if expert music production might be less influenced by such constraints, physical affordances might still play a role in the acquisition of expertise. Music is also, more often than not, produced collectively, by two or more individuals involved in joint action (D'Ausilio, Novembre, Fadiga, & Keller, 2015). Joint action, and coordination with one or several partners is itself a complex, at least partly ecological factor of attraction (Keller et al., 2007). It can be expected to have influenced the evolution of several aspects of musical productions, including, for instance, the tempo (Wolf, Vesper, Sebanz, Keller, & Knoblich, 2019), in non-trivial ways.

Conclusion

As announced in the introduction, each chapter aimed both to address specific questions and to be understood in relation to the framework presented in the introduction. Here, after briefly summing up their respective results and contributions, I highlight how they enrich and illustrate the framework sketched in the introduction.

Chapter 1: Cumulative culture in the laboratory

This first chapter reviewed the use of cultural transmission experiments (transmission chains, replacement, closed groups and seeded groups) in studying cumulative cultural evolution. Cumulative cultural evolution is usually defined as the process by which traditions are gradually modified (and improved, in the case of technological traditions). This chapter identified several mismatches between theoretical definitions of cumulative culture and their implementation in cultural transmission experiments:

- Observed performance improvement can result from participants learning faster in a group context rather than being evidence of a genuine cumulative effect;
- Participants are asked to complete quite simple tasks, which can undermine the evidential value of the diagnostic criterion traditionally used for cumulative culture (i.e. that cumulative culture is a process that produces solutions that no single individual could have invented on their own).
- The use of unidimensional metrics of cumulativeness drastically curtails the variation that may be observed, which raises specific issues in the interpretation of the experimental evidence.

This chapter also suggested possible solutions to reduce these mismatches (adapt controls to have comparable learning times and increase task complexity and ecological validity) and to develop the use of design spaces in experimental studies of cumulative culture. While illustrating the utility of cultural transmission experiments and proposing ways to enhance it, this chapter should also be read as a plea for more diversified methodologies.

Chapter 2: When iconicity stands in the way of abbreviation

Zipf's law of abbreviation, relating more frequent signals to shorter signal lengths, has been shown to apply to sounds in a variety of communication systems, both human and non-human. This study documented an exception to this law of abbreviation. Observing European heraldic motifs, whose frequency of use was documented over the whole continent in two large corpus (total N = 25115), one medieval, one early modern, we found that they do not obey a robust law of abbreviation. In our early modern corpus, motif complexity and motif frequency were positively, not negatively, correlated, a result driven by iconic motifs. In both our corpus, iconic motifs tended to be frequent and complex. They grew in popularity after the invention of printing. Our results suggested that lacking –or at least losing- iconicity may be a precondition for Zipf's Law of Abbreviation to obtain in a graphic code.

Since this chapter was written, another exception to Zipf law of Abbreviation has now surfaced, in chimpanzee gestural communication (Heesen, Hobaiter, Ferrer-i-Cancho, & Semple, 2019).

Chapter 3: Complex Writing

What determines the visual complexity of written characters? How does inventory size (the number of characters in a script), type (what kind of linguistic units are represented by the script's characters), and phylogenetic influence (which script each script descended from, and where it was localised)? Do characters become simpler when they belong to scripts that have been exposed to evolutionary pressures for longer amounts of time? This chapter tested these hypotheses using a standardized collection of 47,880 pictures from 133 writing systems, and two measures of visual complexity (algorithmic and perimetric).

Our results support the conclusions that (1) the size of a script's inventory influences character complexity, (2) one of the main determinant of character complexity is the script's type (e.g., alphabetic, syllabic), and (3) there is a surprising lack of evolutionary change in character complexity.

Chapter 4: Forward bias in human portraits

Chapter 4 reviews the evidence for existing biases in spatial composition that may be at play in human profile portraits. Such biases would predict that two main types of composition are attractive: sitters centered in their frame, or ex-centered with more space in front of the sitter than behind her (a forward bias). This chapter evidenced the existence of a forward bias in human profile-oriented portraits: there is a widespread tendency (total N = 1833, from 582 unique painters) to represent sitters with more space in front of them than behind them. It also suggested that this bias became more frequently and more strongly expressed over time.

Chapter 5: The role of motor /physical constraints in the production of rhythms

Finally, chapter 5 aimed at testing whether and how different physical constraints influence the rhythms naïve participants produce in a transmission chain experiment. Participants were asked to reproduce rhythmical sequences: metronome sequences for the first generation, and sequences produced by previous participants in the chains, for subsequent generations. The amplitude of movements (small or large) influenced the time between two taps. Whether a condition included one or two types of movement determined whether the rhythms produced remained isochronous or became non-isochronous over time (i.e., experimental generations). Finally, different physical constraints led to different levels of stability: participants in conditions that included both small and large movements produced rhythms less similar to the ones they heard than participants in conditions including only one type of movement (small or large).

Contrasts between studies: robustness across cultural domains, methods, and diversity in factors of attraction

In addition to their contribution to their own research questions, chapters 2 to 5 also illustrate two aspects of the framework presented in the introduction: (1) the robustness of the approach for very different cultural domains, and (2) the diversity in the types of causal factors relevant to understanding the success of given cultural types.

The methodological framework presented in the introduction, and used throughout all case studies presented in this thesis, has shown robustness across domains, and across types of causal factors – see Table 1.

First, these case studies covered a variety of cultural domains. The first two studies (chapters 2 and 3) focused on graphic communication systems, although they differed on their specific uses. Heraldic motifs, the focus of chapter 2, were used to identify families (coats of arms). Writing, by contrast, supports the transcription of spoken communication and has allowed forms of written communication without oral counterpart. Chapter 4 dealt with biases in spatial composition and aesthetic sensibilities applied to human profile-oriented portraits. Chapter 5 tackles physical constraints impacting the production of rhythms with a musical apparatus. These case studies add up to previous works, which also appealed to the notion of cultural attraction to explain cultural phenomena (Morin, 2016) including gaze-orientation in portraits (Morin, 2013) and the use of cardinal and oblique strokes in writing systems (Morin, 2018). Medical beliefs can also be added to that list – both bloodletting and anti-vaccination beliefs can find reasonable explanations within a cultural attraction framework (Miton, Claidière, & Mercier, 2015; Miton & Mercier, 2015).

Chapters 2 to 5 also cover a range of different factors of attraction. Both chapter 2 on heraldic motifs and chapter 3 on written characters both test possible determinants of visual complexity, and build on known effects of visual complexity on the cognitive processing of such stimuli. In particular the frequencies of heraldic motifs were also influenced by another cognitive factor–the possibility to produce iconic motifs, which are also easy to process despite higher complexity—and production costs, which decreased with the invention of printing (a more ecological factor of attraction).

Chapter 4 on human profile-oriented portraits also focused on a cognitive factor of attraction: the disposition to pay more attention to what is in front of an agent and, in particular to what she may be seeing, than to what is behind her and out of her line of vision. Such a cognitive disposition is useful to infer the agent's perceptions and thoughts and to anticipate her actions. This cognitive disposition is a factor of attraction in the production and appreciation of human portraits. This, of course, was only one causal factor in the evolution of portrait painting, other factors including socially distributed and historically diverse norms of composition. Both the pressure from social norms and the availability of materials, including format of canvas, are more ecological than cognitive factors of attraction (with the qualification that once internalized by individual painters, norms weigh also as psychological factors).

Finally, chapter 5 focused on affordances and constraints from a physical apparatus, and thus on an ecological factor of attraction. At the same time, we noted that rhythmical patterns produced by participants were kept in the vicinity of IOIs (Inter Onset Intervals) known to be particularly cognitively easy and appreciated (500ms, Collyer, Broadbent, & Church, 1994; Moelants, 2002). The participants' productions we observe do not, however, drift to the point of reaching another integer ratio – which could have been the case if participants had optimised ease (and cognitive appeal, to some extent).

The case studies also illustrated that the distinction between ecological and cognitive factors of attraction is a blurry one, a matter of degree more than a categorical distinction.

	Cultural phenomenon	Factor(s) of attraction	Methods (type of data)	Domain
Chapter 2	Relation between frequency and complexity in heraldic motifs	Visual complexity Iconicity Production costs	Large-scale (total N = 25115) historical, one communication system	Graphic communication (coats of arms, family identifiers)
Chapter 3	Scripts with different degrees of complexity	Number of characters included in a script, linguistic unit represented by characters, ancestry	Large-scale (47 880 characters, 133 scripts), across different systems	Graphic communication, transcribing spoken language
Chapter 4	Human profile-oriented portraits	Cognitive disposition to attend to what is in front of agents Historical norms	Large-scale (1831 paintings, 582 painters), historical (1425 to 2018)	Visual art
Chapter 5	Rhythmical sequences	Physical constraints (ecological)	Experimental (transmission chains)	Music

Table 1. A summary of the four case studies included in this thesis (i.e., chapters 2 to 5), including which cultural phenomenon was studied, and the influence of which factors of attraction were tested, along with the methods used and cultural domain in which they took place.

By mixing different types of factors of attraction, these chapters echo another point made in the introduction: all those chapters studied attractors, i.e., fixation points determined by an array of causal factors. The distribution of occurrences (e.g., clusters) around attractors (fixation points) is dynamic – and factors of attraction here play a role both in ensuring stability and change: Whenever a factor of attraction changes in which content it favours (orientation) or how strongly (strength), this shifts the location of the attractor in the variation space. Within the chapter on heraldry, the cost of complex motifs limits the appeal of iconic complex motifs before the invention of printing. After the invention of printing, the cost of production for complex motifs decreased, and opened the possibility for iconic motifs (which are complex, yet easy to process in virtue of their iconicity) to gain in popularity. Production and processing costs here are a factor that explains both stability and change at a populational level. In chapter 4, the weakening of cultural norms for centering sitters play a role in explaining why sitters became more ex-centered (with more space in front of them) over time. Finally, the experiment described in chapter 5 had four different conditions, based on four different movement sequences. They all started with the same sequence to reproduce: 13 taps spaced by 500 ms. It followed the logic of creating an out-of-equilibrium system: the different conditions could either match or not the type of rhythm from the seed. Conditions that included a mix of movements were less well 'matched' to the sequence to reproduce than conditions with only one type of movement. Physical constraints and their fit with different rhythmical sequences influenced both how much change there was and which productions were stable.

5. Epilogue: the Gym card

Figure 1 is my university gym card. Can you tell what is wrong with it?

Figure 1. The author's university gym card.

My last name has been spelled wrong. There is no 'l' in Miton – it's a 5-letters long family name of French origin. The gym card was written with little time pressure, and directly copied from a correct occurrence (the student card issued by the university). It occurred right after a verbal interaction that would leave the copier quite certain that I am not a native speaker of English. She wasn't either.

This is also an addition: adding this letter slightly increased the time it took to write the whole name, and it cannot be explained as 'simply' forgetting, as an omission could. It makes the whole copying slightly longer.

This misspelling of my last name is actually pretty frequent and occurred repeatedly in contexts in which the right spelling was equally available and salient. It happened at conferences, when people would tweet about a talk I was giving – the correct spelling would still be displayed on my first slide and/or on the program, yet the tweet would show that mysterious L appearing between the I and the T. The two spellings would even sometimes co-exist in the very same text. It happened in emails, in my last name is spelled right in the address but not in the core of the email. It happened when some of my articles would be cited – and this, despite the fact that citations are largely automatized by reference managers, which are supposedly less prone to errors than humans.

Anecdote aside, what is there to take from that example for the study of cultural evolution?

First, (high-fidelity) copying does not necessarily occur, even when all conditions are present for it to happen - transformations are still persistent and frequent. Second, most transformations are not simply random mistakes: this misspelling is anything but random. On purely theoretical grounds, there are at least two possible types of transformations (a letter could have been omitted, rather than added), times 26 candidate letters times 4 possible places in the name (after the first, second, third or fourth letter). This would make a total of 208 possible transformations – and this is a low estimate, as we excluded any substitution, or the possibility of change at the first or last letters. Yet, *one* transformation is much more likely than any other.

Constraining possible transformations to phonologically acceptable combinations would of course limit this combinatorial explosion – but it is already part of the answer. It amounts to restricting the possible variation space to a *plausible* subset, and to start narrowing down on which causal factors are involved.

"Miton", just as any proper or common name, could in principle be misspelled in many ways. Only one of many possible misspelling, "Milton," actually occurs with a frequency that makes it noticeable. This misspelling is far from being a random mutation. The name of the author of *Paradise Lost* is a strong attractor in the vicinity of my name. Many quite diverse factors of attraction—different ones for different cases—may help explain the existence of clusters of such tokens in the vicinity of attractors.

As sketched in the introduction, it is because of the extreme variety of cultural phenomena to be explained and of the diversity of causal factors that may explain them that we should start off with a very flexible framework. In order to account for the variety of causal factors and phenomena to explain within the realm of culture, a variety of models will be required, some of which might not exemplify Darwinian principles – thus making them slightly 'outside' of the field of cultural evolution as standardly understood. I have presented an attempt at developing an alternative approach. I have described some building blocks of a minimalist (and non-exhaustive) ontology of culture: attractors – what needs to be explained, factors of attraction – how to explain them, and cultural causal chains – how to describe the trajectories of cultural contents. These concepts and the relations between them serve as a basis for an operational framework both robust and flexible, as I have tried to illustrate across cultural domains, types of causal factors, and methods.

References

Acerbi, A., Charbonneau, M., Miton, H., & Scott-Phillips, T. (n.d.). *Cultural stability without copying*.

Acerbi, A., Tennie, C., & Mesoudi, A. (2016). Social learning solves the problem of narrow-peaked search landscapes: Experimental evidence in humans. *Royal Society Open Science*, *3*, 160215.

Acerbi, A., Tennie, C., & Nunn, C. L. (2011). Modeling imitation and emulation in constrained search spaces. *Learning & Behavior*, *39*(2), 104–114.

Aleem, H., Correa-Herran, I., & Grzywacz, N. M. (2017). Inferring Master Painters' Esthetic Biases from the Statistics of Portraits. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00094

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Anyinam, C. (1995). Ecology and ethnomedicine: Exploring links between current environmental crisis and indigenous medical practices. *Social Science & Medicine*, *40*(3), 321–329. https://doi.org/10.1016/0277-9536(94)E0098-D

Apfelbach, R., & Wester, U. (1977). The quantitative effect of visual and tactile stimuli on the prey-catching behaviour of ferrets (Putorius furo L.). *Behavioural Processes*, *2*(2), 187–200. https://doi.org/10.1016/0376-6357(77)90020-1

Arend, U., Muthig, K.-P., & Wandmacher, J. (1987). Evidence for global feature superiority in menu selection by icons. *Behaviour & Information Technology*, *6*(4), 411–426. https://doi.org/10.1080/01449298708901853

Arnheim, R. (1965). Art and Visual Perception: A Psychology of the Creative Eye. University of California Press.

Arnheim, R. (1983). The Power of the Center: A Study of Composition in the Visual Arts. University of California Press.

Arnoult, M. D. (1960). Prediction of Perceptual Responses from Structural Characteristics of the Stimulus. *Perceptual and Motor Skills*, *11*(3), 261–268. https://doi.org/10.2466/pms.1960.11.3.261

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*(6), 452–471. https://doi.org/10.1037/h0044049

Balfour, H. (1893). The evolution of decorative art: An essay upon its origin and development as illustrated by the art of modern races of mankind. Rivington, Percival & Company.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 424–434. https://doi.org/10.3758/BF03195471

Bartlett, S. F. C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*(7), 1–23.

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

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, *16*(4), 391–399. https://doi.org/10.1016/j.conb.2006.07.001

Bennett, C. H., Gacs, P., Ming Li, Vitanyi, P. M. B., & Zurek, W. H. (1998). Information distance. *IEEE Transactions on Information Theory*, *44*(4), 1407–1423. https://doi.org/10.1109/18.681318

Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society of London B: Biological Sciences*, *271*(1547), 1443–1450. https://doi.org/10.1098/rspb.2004.2746

Bentz, C., & Ferrer-i-Cancho, R. (2016). Zipf's law of abbreviation as a language universal. Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics, Leiden, The Netherlands, 26–30 October 2015. Presented at the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics, Tübingen.

Beppu, A., & Griffiths, T. L. (2009). Iterated Learning and the Cultural Ratchet. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2089–2094). Amsterdam: Cognitive Science Society.

Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target–flanker similarity. *Journal of Vision*, *11*(8), 1–1. https://doi.org/10.1167/11.8.1

Bernstein, L. J., & Cooper, L. A. (1997). Direction of motion influences perceptual identification of ambiguous figures. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 721–737. https://doi.org/10.1037/0096-1523.23.3.721

Bertamini, M., Bennett, K. M., & Bode, C. (2011). The anterior bias in visual art: The case of images of animals. *Laterality: Asymmetries of Body, Brain and Cognition*, *16*(6), 673–689. https://doi.org/10.1080/1357650X.2010.508219

Bezerra, B. M., Souto, A. S., Radford, A. N., & Jones, G. (2011). Brevity is not always a virtue in primate communication. *Biology Letters*, *7*(1), 23–25. https://doi.org/10.1098/rsbl.2010.0455

Bode, C., Bertamini, M., & Helmy, M. S. (2016). Left–Right Position in Moving Images: An Analysis of Face Orientation, Face Position, and Movement Direction in Eight Action Films. *Art & Perception*, *4*(3), 241–263. https://doi.org/10.1163/22134913-00002053

Bond, Jr., Charles F., & Titus, L. J. (1983). Social Facilitation: A Meta-Analysis of 241 Studies. *Psychological Bulletin*, *94*(2), 265–292.

Boyd, R., & Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy*, *88*, 77–93.

Boyer, P. (2007). Religion explained: The evolutionary origins of religious thought. Basic books.

Brown, S. (2000). Evolutionary Models of Music: From Sexual Selection to Group Selection. In F. Tonneau & N. S. Thompson (Eds.), *Perspectives in Ethology: Evolution, Culture, and Behavior* (pp. 231–281). https://doi.org/10.1007/978-1-4615-1221-9_9

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. https://doi.org/10.1177/0305735611425896

Buskell, A. (2017). What are cultural attractors? *Biology & Philosophy*, *32*(3), 377–394. https://doi.org/10.1007/s10539-017-9570-6

Buskell, A. (2019). Looking for middle ground in cultural attraction theory. *Evolutionary Anthropology: Issues, News, and Reviews*, *28*(1), 14–17. https://doi.org/10.1002/evan.21762

Byrne, M. D. (1993). Using icons to find documents: Simplicity is critical. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*, 446–453. https://doi.org/10.1145/169059.169369

Caldwell, C. A., Atkinson, M., & Renner, E. (2016). Experimental Approaches to Studying Cumulative Cultural Evolution. *Current Directions in Psychological Science*, *25*(3), 191–195.

Caldwell, C. A., Cornish, H., & Kandler, A. (2016). Identifying innovation in laboratory studies of cultural evolution: Rates of retention and measures of adaptation. *Philosophical Transactions of the Royal Society B*, *371*(20150193).

Caldwell, C. A., & Eve, R. M. (2014). Persistence of Contrasting Traditions in Cultural Evolution: Unpredictable Payoffs Generate Slower Rates of Cultural Change. *PLoS ONE*, *9*(6), e99708.

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

Caldwell, C. A., & Millen, A. E. (2008b). Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society B*, *363*, 3529–3539.

Caldwell, C. A., & Millen, A. E. (2009). Social Learning Mechanisms and Cumulative Cultural Evolution: Is Imitation Necessary? *Psychological Science*, *20*(12), 1478–1483.

Caldwell, C. A., & Millen, A. E. (2010a). Conservatism in laboratory microsocieties: Unpredictable payoffs accentuate group-specific traditions. *Evolution and Human Behavior*, *31*, 123–130.

Caldwell, C. A., & Millen, A. E. (2010b). Human Cumulative Culture in the Laboratory: Effects of (Micro) Population Size. *Learning and Behavior*, *38*, 310–318.

Caldwell, C. A., & Smith, K. (2012). Cultural Evolution and Perpetuation of Arbitrary Communicative Conventions in Experimental Microsocieties. *PLoS ONE*, *7*(8), e43807. https://doi.org/10.1371/journal.pone.0043807

Catania, K. C. (2009). Tentacled snakes turn C-starts to their advantage and predict future prey behavior. *Proceedings of the National Academy of Sciences*, *106*(27), 11183–11187. https://doi.org/10.1073/pnas.0905183106

Chanceaux, M., Mathôt, S., & Grainger, J. (2014). Effects of number, complexity, and familiarity of flankers on crowded letter identification. *Journal of Vision*, *14*(6), 7–7. https://doi.org/10.1167/14.6.7

Chang, L.-Y., Chen, Y.-C., & Perfetti, C. A. (2018). GraphCom: A multidimensional measure of graphic complexity applied to 131 written languages. *Behavior Research Methods*, *50*(1), 427–449. https://doi.org/10.3758/s13428-017-0881-y

Chang, L.-Y., Plaut, D. C., & Perfetti, C. A. (2016). Visual complexity in orthographic learning: Modeling learning across writing system variations. *Scientific Studies of Reading*, *20*(1), 64–85. https://doi.org/10.1080/10888438.2015.1104688

Changizi, M., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1560), 267–275. https://doi.org/10.1098/rspb.2004.2942

Changizi, M., Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *The American Naturalist*, *167*(5), E117-139. https://doi.org/10.1086/502806

Charbonneau, M. (2015). All innovations are equal, but some more than others: (Re)integrating modification processes to the origins of cumulative culture. *Biological Theory*, *10*(4), 322–335.

Charbonneau, M. (2016). Modularity and Recombination in Technological Evolution. *Philosophy & Technology*, *29*(4), 373–392. https://doi.org/10.1007/s13347-016-0228-0

Charbonneau, M. (2018a). Technical Constraints on the Convergent Evolution of Technologies. In M. J. O'Brien, B. Buchanan, & M. I. Eren (Eds.), *Convergent Evolution in Stone-Tool Technology* (pp. 73–89). Cambridge, MA: MIT Press.

Charbonneau, M. (2018b). Understanding cultural fidelity [Published Article or Volume]. Retrieved June 16, 2019, from The British Journal for the Philosophy of Science website: http://philsci-archive.pitt.edu/14745/

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. https://doi.org/10.1016/S1364-6613(02)00005-0

Chrisomalis, S. (2004). A Cognitive Typology for Numerical Notation. *Cambridge Archaeological Journal*, *14*(1), 37–52. https://doi.org/10.1017/S0959774304000034

Claidière, N., Amedon, G. K., André, J.-B., Kirby, S., Smith, K., Sperber, D., & Fagot, J. (2018). Convergent transformation and selection in cultural evolution. *Evolution and Human Behavior*, *39*(2), 191–202. https://doi.org/10.1016/j.evolhumbehav.2017.12.007

Claidière, N., Kodjo-kuma Amedon, G., André, J.-B., Kirby, S., Smith, K., Sperber, D., & Fagot, J. (2018). Convergent transformation and selection in cultural evolution. *Evolution and Human Behavior*, *39*(2), 191–202.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution? *Phil. Trans. R. Soc. B*, *369*(1642), 20130368. https://doi.org/10.1098/rstb.2013.0368

Claidière, N., Smith, K., Kirby, S., & Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society B*, *281*(281), 1797.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1–2), 89–111. https://doi.org/10.1163/156853707X171829

Claidière Nicolas, & Sperber Dan. (2010). Imitation explains the propagation, not the stability of animal culture. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 651–659. https://doi.org/10.1098/rspb.2009.1615

Clark, G. (2016, June 16). Lincoln Green and Real Dutch Java Prints: Cloth Selvedges as Brands in International Trade,. https://doi.org/10.4324/9781315430898-14

Clemmensen, S. (2017). *Armorial.dk: A medieval armorials database.* Retrieved from http://www.armorial.dk/

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 6008–6010. https://doi.org/10.1073/pnas.0811884106

Cohen Priva, U. (2017). Informativity and the actuation of lenition. *Language*, *93*(3), 569–597. https://doi.org/10.1353/lan.2017.0037

Collyer, C. E., Broadbent, H. A., & Church, R. M. (1994). Preferred rates of repetitive tapping and categorical time production. *Perception & Psychophysics*, *55*(4), 443–453. https://doi.org/10.3758/BF03205301

Condit-Schultz, N. (2019). Deconstructing the nPVI: A Methodological Critique of the Normalized Pairwise Variability Index as Applied to Music. *Music Perception: An Interdisciplinary Journal*, *36*(3), 300–313. https://doi.org/10.1525/mp.2019.36.3.300

Conner, R. N. (1985). Vocalizations of Common Ravens in Virginia. *The Condor*, *87*(3), 379–388. https://doi.org/10.2307/1367219

Cooper, W. E. (1981). Visual guidance of predatory attack by a scincid lizard, Eumeces laticeps. *Animal Behaviour*, *29*(4), 1127–1136. https://doi.org/10.1016/S0003-3472(81)80065-6

Crema, E. R., Kandler, A., & Shennan, S. (2016). Revealing patterns of cultural transmission from frequency data: Equilibrium and non-equilibrium assumptions. *Scientific Reports*, *6*, 39122. https://doi.org/10.1038/srep39122

Cross, I., & Morley, I. (n.d.). The evolution of music: Theories, definitions and the nature of the evidence. 22.

Culbertson, J., & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01964

Cultural Evolution Society. (n.d.). Retrieved September 28, 2019, from culturalevolutionsociety.org/

Daniels, P. T., & Bright, W. (1996). *The World's Writing Systems*. Oxford University Press.

D'Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, *19*(3), 111–114. https://doi.org/10.1016/j.tics.2015.01.005

Davis, S. J., Vale, G. L., Schapiro, S. J., Lambeth, S. P., & Whiten, A. (2016). Foundations of cumulative culture in apes: Improved foraging efficiency through relinquishing and combining witnessed behaviours in chimpanzees (Pan troglodytes). *Scientific Reports*, *6*(35953).

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture. *Science*, *335*, 1114–1118.

Dean, Lewis G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews*, *89*, 284–301.

Dehaene, S. (2010). Reading in the brain: The new science of how we read. New York: Viking.

Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

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

Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature Communications*, *6*, 8398.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *PNAS*, *113*(11), 2982–2987.

Derex, M., Feron, R., Godelle, B., & Raymond, M. (2015). Social learning and the replication process: An experimental investigation. *Proceedings of the Royal Society B*, *282*, 20150719.

Derex, M., Godelle, B., & Raymond, M. (2013). Social learners require process information to outperform individual learners. *Evolution*, *67*(3), 688–697.

Derex, M., Godelle, B., & Raymond, M. (2014). How does competition affect the transmission of information? *Evolution and Human Behavior*, *35*, 89–95.

Donderi, D.C., & McFadden, S. (2005). Compressed file length predicts search time and errors on visual displays. *Displays*, *26*(2), 71–78. https://doi.org/10.1016/j.displa.2005.02.002

Donderi, Don C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73–97. https://doi.org/10.1037/0033-2909.132.1.73

Enfield, N. J. (2014). Natural Causes of Language: Frames, biases and cultural transmission.

Enquist, M., Strimling, P., Eriksson, K., Laland, K., & Sjostrand, J. (2010). One cultural parent makes no culture. *Animal Behaviour*, *79*(6), 1353–1362. https://doi.org/10.1016/j.anbehav.2010.03.009

Ericsson, K. A. (1996). The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games. New York: Psychology Press.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints. *Annual Review of Psychology*, *47*, 273–305.

Eriksson, K., & Coultas, J. (2012). The advantage of multiple cultural parents in the cultural transmission of stories. *Evolution and Human Behavior*, *33*, 251–259.

Eriksson, K., & Coultas, J. C. (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. https://doi.org/10.1163/15685373-12342107

Farroni, T., Johnson, M. H., Brockbank, M., & Simion, F. (2000). Infants' use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition*, *7*(6), 705–718. https://doi.org/10.1080/13506280050144399

Ferdinand, V. A. (2015). *Inductive evolution: Cognition, culture, and regularity in language*. Retrieved from https://www.era.lib.ed.ac.uk/handle/1842/11741

Ferrer-i-Cancho, R., & Hernández-Fernández, A. (2013). The failure of the law of brevity in two New World primates. Statistical caveats. *Glottotheory*, *4*(1), 44–55. https://doi.org/10.1524/glot.2013.0004

Ferrer-i-Cancho, R., & Lusseau, D. (2009). Efficient coding in dolphin surface behavioral patterns. *Complexity*, *14*(5), 23–25. https://doi.org/10.1002/cplx.20266

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/h0055392

Flynn, E. (2008). Investigating children as cultural magnets: Do young children transmit redundant information among diffusion chains? *Philosophical Transactions of the Royal Society B*, *363*, 3541–3551.

Fox-Davies, A. C. (1900). *Heraldry Explained*. London: T.C. & E.C. Jack.

Gandon, E., Roux, V., & Coyle, T. (2014). Copying errors of potters from three cultures: Predictable directions for a so-called random phenomenon. *Journal of Anthropological Archaeology*, *33*, 99–107.

García, M., Badre, A. N., & Stasko, J. T. (1994). Development and validation of icons varying in their abstractness. *Interacting with Computers*, *6*(2), 191–211. https://doi.org/10.1016/0953-5438(94)90024-8

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007a). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science*, *31*(6), 961–987. https://doi.org/10.1080/03640210701703659

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007b). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science*, *31*(6), 961–987. https://doi.org/10.1080/03640210701703659

Gelb, I. J. (1963). *A Study of Writing*. Chicago: Univ of Chicago Pr.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2160–2170. https://doi.org/10.1098/rstb.2012.0118

Goldin, C. (2006). The Quiet Revolution That Transformed Women's Employment, Education, and Family. 96(2), 21.

Goldin, C., & Katz, L. F. (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy*, *110*(4), 730–770. https://doi.org/10.1086/340778

Gould, S. J. (1979). Mickey mouse meets konrad lorenz. *Natural History*, *88*(5), 30–36.

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze Following Is Not Dependent on Ostensive Cues: A Critical Test of Natural Pedagogy. *Child Development*, *89*(6), 2091–2098. https://doi.org/10.1111/cdev.13026

Green, A., & Barnard, P. (1990). Graphical and iconic interfacing: The role of icon distinctiveness and fixed or variable screen locations. In D. Diaper, D. Gilmore, G. Cockton, & B. Shackel (Eds.), *INTERACT '90 Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction* (pp. 457–462). Retrieved from http://oro.open.ac.uk/26443/

Guidi, S., & Palmer, S. E. (2015). The Internal Structure of a Rectangular Frame: Position, Orientation, and Direction. *Art & Perception*, *3*(1), 19–52. https://doi.org/10.1163/22134913-00002013

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. H. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience*, *1*(3), 235–241. https://doi.org/10.1038/681

Hamilton, M. J., & Buchanan, B. (2009). The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: Quantifying Clovis evolutionary dynamics. *Journal of Anthropological Archaeology*, *28*, 55–69.

Harmon-Vukić, M., Upal, M. A., & Sheehan, K. J. (2012). Understanding the memory advantage of counterintuitive concepts. *Religion, Brain & Behavior*, *2*(2), 121–139. https://doi.org/10.1080/2153599X.2012.672816

Hedström, P., & Bearman, P. (2011). *The Oxford Handbook of Analytical Sociology*. OUP Oxford.

Heesen, R., Hobaiter, C., Ferrer-i-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896), 20182900. https://doi.org/10.1098/rspb.2018.2900

Heintz, C., Blancke, S., & Scott-Phillips, T. (2019). Methods for studying cultural attraction. *Evolutionary Anthropology: Issues, News, and Reviews*, *28*(1), 18–20. https://doi.org/10.1002/evan.21764

Henrich, J. (2004). Demography and Cultural Evolution: How Adaptative Cultural Processes can Produce Maladaptive Losses: The Tasmanian Case. *American Antiquity*, *69*(2), 197–214.

Henrich, J. (2010). The Evolution of Innovation-Enhancing Institutions. In M. J. O'Brien & S. Shennan (Eds.), *Innovation in Cultural Systems: Contributions from Evolutionary Anthropology* (pp. 99–120). Cambridge, MA: MIT Press.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *12*(3), 123–135. https://doi.org/10.1002/evan.10110

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, *22*(4), e12779. https://doi.org/10.1111/desc.12779

Hernik, M., Fearon, P., & Csibra, G. (2014). Action anticipation in human infants reveals assumptions about anteroposterior body-structure and action. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1781), 20133205. https://doi.org/10.1098/rspb.2013.3205

Herzog, H. A., Bentley, R. A., & Hahn, M. W. (2004). Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society B: Biological Sciences*, *271*(Suppl 5), S353–S356.

Heyes, C. (2018). Enquire within: Cultural evolution and cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170051. https://doi.org/10.1098/rstb.2017.0051

Hinde, R. A., & Barden, L. A. (1985). The evolution of the teddy bear. *Animal Behaviour*, *33*(4), 1371–1373. https://doi.org/10.1016/S0003-3472(85)80205-0

Hori, T. (2005). Cultural note on dreaming and dream study in the future: Release from nightmare and development of dream control technique. *Sleep and Biological Rhythms*, *3*(2), 49–55. https://doi.org/10.1111/j.1479-8425.2005.00163.x

Hruschka, D. J. (2010). Friendship: Development, Ecology, and Evolution of a Relationship. University of California Press.

Huffman, D. (1952). A Method for the Construction of Minimum-Redundancy Codes. *PROCEEDINGS OF THE I.R.E.*, *40*, 1098–1101.

Hurley, S. L., & Chater, N. (2005). Perspectives on Imitation: Imitation, human development, and culture. MIT Press.

ISO 15924 Registration Authority, Mountain View, CA. (2013). *ISO 15924*. Retrieved from http://www.unicode.org/iso15924/iso15924-codes.html

Itakura, S. (2004). Gaze-following and joint visual attention in nonhuman animals. *Japanese Psychological Research*, *46*(3), 216–226. https://doi.org/10.1111/j.1468-5584.2004.00253.x

Jacoby, N., & McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, *27*(3), 359–370. https://doi.org/10.1016/j.cub.2016.12.031

Jahanian, A., Vishwanathan, S. V. N., & Allebach, J. P. (2015). Learning visual balance from large-scale datasets of aesthetically highly rated images. *Human Vision and Electronic Imaging XX*, *9394*, 93940Y. https://doi.org/10.1117/12.2084548

Jappy, T. (2013). *Introduction to Peircean Visual Semiotics*. London: Bloomsbury Academic.

Johnson, C., Bishop, P., & Kelly, S. (2010). Measuring The Mnemonic Advantage of Counter-intuitive and Counter-schematic Concepts. *Journal of Cognition and Culture*, *10*(1–2), 109–121. https://doi.org/10.1163/156853710X497194

Jones, B. F. (2010). Age and Great Invention. *The Review of Economics and Statistics*, *92*(1), 1–14.

Jordan, P., & Shennan, S. (2003). Cultural transmission, language, and basketry traditions amongst the California Indians. *Journal of Anthropological Archaeology*, *22*(1), 42–74. https://doi.org/10.1016/S0278-4165(03)00004-7

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–294. https://doi.org/10.3758/BF03194066

Kandler, A., & Shennan, S. (2013). A non-equilibrium neutral model for analysing cultural change. *Journal of Theoretical Biology*, *330*, 18–25. https://doi.org/10.1016/j.jtbi.2013.03.006

Kandler, A., Wilder, B., & Fortunato, L. (2017). Inferring individual-level processes from population-level patterns in cultural evolution. *Royal Society Open Science*, *4*(170949).

Kano, F., Moore, R., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2018). Human ostensive signals do not enhance gaze following in chimpanzees, but do enhance object-oriented attention. *Animal Cognition*, *21*(5), 715–728. https://doi.org/10.1007/s10071-018-1205-z

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52. https://doi.org/10.1016/j.cognition.2017.05.001

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Conciousness and Cognition*, *16*(1), 102–111.

Kelly, P., Winters, J., Miton, H., & Morin, O. (submitted). The predictable evolution of letter shapes: An emergent script of West Africa recapitulates historical change in writing systems.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic Typology and Efficient Communication. *Annual Review of Linguistics*, *4*(1), 109–128. https://doi.org/10.1146/annurev-linguistics-011817-045406

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

Kempe, M., & Mesoudi, A. (2014b). Experimental and theoretical models of human cultural evolution. *Wires Cognitive Science*.

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247–254.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences*, *22*(7), 651–665. https://doi.org/10.1016/j.tics.2018.04.003

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., … Gavin, M. C. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLOS ONE*, *11*(7), e0158391. https://doi.org/10.1371/journal.pone.0158391

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. https://doi.org/10.1016/j.cognition.2015.03.016

Krakauer, D. C. (2011). The Star Gazer and the Flesh Eater: Elements of a Theory of Metahistory. *Cliodynamics*, *2*(1). https://doi.org/10.21237/C7clio21207

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest.' *R Package Version*, *2*(0).

Labar, M. (1973). Turning the Left Cheek Examined using Modern Photography. *Nature*, *245*(5424), 338–338. https://doi.org/10.1038/245338a0

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14. https://doi.org/10.3758/BF03196002

Laland, K. N. (2018). Darwin's Unfinished Symphony: How Culture Made the Human Mind. Princeton University Press.

Langlois, T. A. (2018). *Uncovering Human Visual Priors* (UC Berkeley). Retrieved from https://escholarship.org/uc/item/7843w9q3

Lehman, H. C. (1953). *Age and Achievement*. Princeton, NJ: Princeton University Press.

Lewens, T. (2015). Cultural Evolution: Conceptual Challenges. OUP Oxford.

Leyssen, M. H. R., Linsen, S., Sammartino, J., & Palmer, S. E. (2012). Aesthetic Preference for Spatial Composition in Multiobject Pictures. *I-Perception*, *3*(1), 25–49. https://doi.org/10.1068/i0458aap

Lindell, A. K. (2013). The silent social/emotional signals in left and right cheek poses: A literature review. *Laterality: Asymmetries of Body, Brain and Cognition*, *18*(5), 612–624. https://doi.org/10.1080/1357650X.2012.737330

Liu, T., Chen, W., Liu, C. H., & Fu, X. (2012). Benefits and costs of uniqueness in multiple object tracking: The role of object complexity. *Vision Research*, *66*, 31–38. https://doi.org/10.1016/j.visres.2012.06.009

Lloyd, S. (2001). Measures of complexity: A nonexhaustive list. *IEEE Control Systems Magazine*, *21*(4), 7–8. https://doi.org/10.1109/MCS.2001.939938

Locher, P., Cornelis, E., Wagemans, J., & Stappers, P. J. (2001). Artists' Use of Compositional Balance for Creating Visual Displays. *Empirical Studies of the Arts*, *19*(2), 213–227. https://doi.org/10.2190/EKMD-YMN5-NJUG-34BK

Locher, P. J., Jan Stappers, P., & Overbeeke, K. (1998). The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays. *Acta Psychologica*, *99*(2), 141–161. https://doi.org/10.1016/S0001-6918(98)00008-0

Lomax, A., & Berkowitz, N. (1972). The Evolutionary Taxonomy of Culture. *Science*, *177*(4045), 228–239. Retrieved from JSTOR.

Luo, B., Jiang, T., Liu, Y., Wang, J., Lin, A., Wei, X., & Feng, J. (2013). Brevity is prevalent in bat short-range communication. *Journal of Comparative Physiology A*, *199*(4), 325–333. https://doi.org/10.1007/s00359-013-0793-y

Mace, R., Pagel, M., Bowen, J. R., Otterbein, K. F., Ridley, M., Schweizer, T., & Voland, E. (1994). The Comparative Method in Anthropology [and Comments and Reply]. *Current Anthropology*, *35*(5), 549–564. https://doi.org/10.1086/204317

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, *97*(8), 4398–4403.

Martin, D., Hutchison, J., Slessor, G., Urquhart, J., Cunningham, S. J., & Smith, K. (2014). The Spontaneous Formation of Stereotypes via Cumulative Cultural Evolution. *Psychological Science*, *25*(9), 1777–1786.

Mavrides, C. M., & Brown, D. R. (1969). Discrimination and reproduction of patterns: Feature measures and constraint redundancy as predictors. *Perception & Psychophysics*, *6*(5), 276–280. https://doi.org/10.3758/BF03210098

McBeath, M. K., Morikawa, K., & Kaiser, M. K. (1992). Perceptual Bias for Forward-Facing Motion. *Psychological Science*, *3*(6), 362–367. https://doi.org/10.1111/j.1467-9280.1992.tb00048.x

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. https://doi.org/10.1016/j.conb.2008.09.005

Mcdougall, S. J. P., Curry, M. B., & de Bruijn, O. (1999). Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 487–519. https://doi.org/10.3758/BF03200730

McDougall, S. J. P., de Bruijn, O., & Curry, M. B. (2000). Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity, and distinctiveness. *Journal of Experimental Psychology: Applied*, *6*(4), 291–306. https://doi.org/10.1037/1076-898X.6.4.291

McGuigan, N., Burdett, E., Burgess, V., Dean, L. G., Lucas, A., Vale, G., & Whiten, A. (2017). Innovation and social transmission in experimental micro-societies: Exploring the scope of cumulative culture in young children. *Philosophical Transactions of the Royal Society B*, *372*(20160425).

McManus, I. C., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, *76*(3), 311–324. https://doi.org/10.1111/j.2044-8295.1985.tb01955.x

McManus, I C, Stöver, K., & Kim, D. (2011). Arnheim's Gestalt Theory of Visual Balance: Examining the Compositional Structure of Art Photographs and Abstract Images. *I-Perception*, *2*(6), 615–647. https://doi.org/10.1068/i0445aap

McManus, I Christopher, & Humphrey, N. K. (1973). Turning the left cheek. *Nature*, *243*(5405), 271.

McManus, I Christopher, & Thomas, P. (2007). Eye Centring in Portraits: A Theoretical and Empirical Evaluation. *Perception*, *36*(2), 167–182. https://doi.org/10.1068/p5578

Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and Function in Human Song. *Current Biology*, *28*(3), 356-368.e5. https://doi.org/10.1016/j.cub.2017.12.042

Merker, B., Morley, I., & Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140095. https://doi.org/10.1098/rstb.2014.0095

Mesoudi, A. (2008). An experimental simulation of the "copy-successful-individuals" cultural learning strategy: Adaptive landscapes, producer-scrounger dynamics, and informational access costs. *Evolution and Human Behavior*, *29*, 350–363.

Mesoudi, A. (2011a). An experimental comparison of human social learning strategies: Payoff-biased social learning is adaptive but underused. *Evolution and Human Behavior*, *32*, 334–342.

Mesoudi, A. (2011b). Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences. Chicago ; London: University Of Chicago Press.

Mesoudi, A. (2011c). Variable Cultural Acquisition Costs Constrain Cumulative Cultural Evolution. *PLoS ONE*, *6*(3), e18239.

Mesoudi, A. (2016a). Cultural Evolution: A Review of Theory, Findings and Controversies. *Evolutionary Biology*, *43*(4), 481–497. https://doi.org/10.1007/s11692-015-9320-0

Mesoudi, A. (2016b). Cultural evolution: Integrating psychology, evolution and culture. *Current Opinion in Psychology*, *7*, 17–22. https://doi.org/10.1016/j.copsyc.2015.07.001

Mesoudi, A. (2016c). Cultural evolution: Integrating psychology, evolution and culture. *Current Opinion in Biotechnology*, *7*, 17–22.

Mesoudi, A. (2017). Pursuing Darwin's curious parallel: Prospects for a science of cultural evolution. *Proceedings of the National Academy of Sciences*, *114*(30), 7853–7860. https://doi.org/10.1073/pnas.1620741114

Mesoudi, A., & O'Brien, M. J. (2008). The Learning and Transmission of Hierarchical Cultural Recipes. *Biological Theory*, *3*, 63–72.

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1880), 20180712. https://doi.org/10.1098/rspb.2018.0712

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

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, *97*(3), 405–423. https://doi.org/10.1348/000712605X85871

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, *29*(4), 329–347. https://doi.org/10.1017/S0140525X06009083

Metcalf, A. (2010). *OK: The Improbable Story of America's Greatest Word*. New York, NY: Oxford University Press.

Miton, H., & Charbonneau, M. (2018). Cumulative culture in the laboratory: Methodological and theoretical challenges. *Proc. R. Soc. B*, *285*(1879), 20180677. https://doi.org/10.1098/rspb.2018.0677

Miton, H., Claidière, N., & Mercier, H. (2015). Universal cognitive mechanisms explain the cultural success of bloodletting. *Evolution and Human Behavior*, *36*(4), 303–312. https://doi.org/10.1016/j.evolhumbehav.2015.01.003

Miton, H., & Mercier, H. (2015). Cognitive Obstacles to Pro-Vaccination Beliefs. *Trends in Cognitive Sciences*, *19*(11), 633–636. https://doi.org/10.1016/j.tics.2015.08.007

Miton, H., & Morin, O. (2019). When iconicity stands in the way of abbreviation: No Zipfian effect for figurative signals. *PLOS ONE*, *14*(8), e0220793. https://doi.org/10.1371/journal.pone.0220793

Mitsui, K., & Noguchi, K. (2002). "Searching for gestalt rules on aesthetic arrangement"Proceedings of the 6th Asian Design Conference.

Moelants, D. (2002). Preferred tempo reconsidered. Proceedings of the 7th International Conference on Music Perception and Cognition / C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.). - Sydney, Adelaide, Causal Productions, 2002, 580–583.

Moore, B. C. J., & Sęk, A. (2009). Sensitivity of the human auditory system to temporal fine structure at high frequencies. *The Journal of the Acoustical Society of America*, *125*(5), 3186–3193. https://doi.org/10.1121/1.3106525

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., … Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, *6*, 6029.

Morin, O. (2013). How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution. *Evolution and Human Behavior*, *34*(3), 222–229. https://doi.org/10.1016/j.evolhumbehav.2013.01.004

Morin, O. (2016). *How Traditions Live and Die*. Oxford University Press.

Morin, O. (2018). Spontaneous Emergence of Legibility in Writing Systems: The Case of Orientation Anisotropy. *Cognitive Science*, *42*(2), 664–677. https://doi.org/10.1111/cogs.12550

Morin, O., Kelly, P., & Winters, J. (2019). Writing, graphic codes, and asynchronous communication. *Topics in Cognitive Science*.

Morin, O., & Miton, H. (2018). Detecting wholesale copying in cultural evolution. *Evolution and Human Behavior*, *39*(4), 392–401. https://doi.org/10.1016/j.evolhumbehav.2018.03.004

Mueller, S. T., & Weidemann, C. T. (2012). Alphabetic letter identification: Effects of perceivability, similarity, and bias. *Acta Psychologica*, *139*(1), 19–37. https://doi.org/10.1016/j.actpsy.2011.09.014

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Murray, D. R., Fessler, D. M. T., Kerry, N., White, C., & Marin, M. (2017). The kiss of death: Three tests of the relationship between disease threat and ritualized physical contact within traditional cultures. *Evolution and Human Behavior*, *38*(1), 63–70. https://doi.org/10.1016/j.evolhumbehav.2016.06.008

Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society B*, *371*(20150192).

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2013). Sociality influences cultural complexity. *Proceedings of the Royal Society B*, *281*(20132511).

Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual Complexity and Beauty Appreciation: Explaining the Divergence of Results. *Empirical Studies of the Arts*, *28*(2), 173–191. https://doi.org/10.2190/EM.28.2.d

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52. https://doi.org/10.3758/BF03193811

Nicholls M. E. R., Clode D., Wood S. J., & Wood A. G. (1999). Laterality of expression in portraiture: Putting your best cheek forward. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1428), 1517–1522. https://doi.org/10.1098/rspb.1999.0809

Norenzayan, A., Atran, S., Faulkner, J., & Schaller, M. (2006). Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives. *Cognitive Science*, *30*(3), 531–553. https://doi.org/10.1207/s15516709cog0000_68

Noveck, I. A., & Sperber, D. (2004). *Experimental pragmatics*. Springer.

O'Brien, M. J., Boulanger, M. T., Buchanan, B., Bentley, R. A., Lyman, R. L., Lipo, C. P., … Eren, M. I. (2016). Design Space and Cultural Transmission: Case Studies from Paleoindian Eastern North America. *Journal of Archaeological Method and Theory*, *23*, 692–740.

O'Brien, M. J., Buchanan, B., & Eren, M. I. (2018). *Convergent Evolution in Stone-Tool Technology*. Cambridge, MA: MIT Press.

Palmer, S. E., & Guidi, S. (2011). Mapping the Perceptual Structure of Rectangles through Goodness-of-Fit Ratings. *Perception*, *40*(12), 1428–1446. https://doi.org/10.1068/p7021

Palmer, S. E., & Langlois, T. A. (2017). Effects of Implied Motion and Facing Direction on Positional Preferences in Single-Object Pictures. *Perception*, *46*(7), 815–829. https://doi.org/10.1177/0301006617694189

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual Aesthetics and Human Preference. *Annual Review of Psychology*, *64*(1), 77–107. https://doi.org/10.1146/annurev-psych-120710-100504

Palmer, S., Gardner, J., & Wickens, T. (2008). Aesthetic issues in spatial composition: Effects of position and direction on framing single objects. *Spatial Vision*, *21*(3–5), 421–449.

Parkes, M. B. (2008). Their Hands Before Our Eyes: A Closer Look at Scribes : the Lyell Lectures Delivered in the University of Oxford, 1999. Ashgate Publishing, Ltd.

Pastoureau, M. (2007). *Traité d'héraldique* (5e édition). Paris: Editions A & J Picard.

Pavlova, M., Krägeloh-Mann, I., Birbaumer, N., & Sokolov, A. (2002). Biological Motion Shown Backwards: The Apparent-Facing Effect. *Perception*, *31*(4), 435–443. https://doi.org/10.1068/p3262

Peirce, C. S. (1974). *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*(28), 4646–4674. https://doi.org/10.1016/j.visres.2006.04.023

Perfetti, C. A., & Harris, L. N. (2013). Universal Reading Processes Are Modulated by Language and Writing System. *Language Learning and Development*, *9*(4), 296–316. https://doi.org/10.1080/15475441.2013.813828

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Prentiss, A. M., Skelton, R. R., Eldredge, N., & Quinn, C. (2011). Get Rad! The Evolution of the Skateboard Deck. *Evolution: Education and Outreach*, *4*(3), 379–389. https://doi.org/10.1007/s12052-011-0347-0

Puffer, E. D. (1903). Studies in symmetry. *The Psychological Review: Monograph Supplements*, *4*(1), 467–539.

Puttfarken, T. (2000). The discovery of pictorial composition: Theories of visual order in painting, 1400-1800.

R Core Team. (2018). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Austria, 2015. ISBN 3-900051-07-0: URL http://www. R-project. org.

Ravignani, A., Bowling, D. L., & Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01118

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1), 1–7. https://doi.org/10.1038/s41562-016-0007

Ravignani, A., Thompson, B., Grossi, T., Delgado, T., & Kirby, S. (2017). Evolving building blocks of rhythm: How human cognition creates music via cultural transmission. *BioRxiv*, *198390*.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48. Retrieved from JSTOR.

Reindl, E., Apperly, I. A., Beck, S. R., & Tennie, C. (2017). Young children copy cumulative technological design in the absence of action information. *Scientific Reports*, *7*(1788).

Reindl, E., & Tennie, C. (2018). Young children fail to generate an additive ratchet effect in an open-ended construction task. *PLOS ONE*, *13*(6), e0197828. https://doi.org/10.1371/journal.pone.0197828

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., … Laland, K. N. (2010). Why Copy Others? Insights from the Social Learning Strategies Tournament. *Science*, *328*(5975), 208–213. https://doi.org/10.1126/science.1184719

Renesse, T. (1894). *Dictionnaire des figures héraldiques* (Vols. 1–7). Bruxelles: Oscar Schepens.

Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Rietstap, J.-B. (1884). *Armorial Général, précédé d'un dictionnaire des termes du blason* (2d ed., Vols. 1–2). Gouda: G. B. van Goor Zonen.

Rogers, E. M. (2003). *Diffusion of Innovations* (Fifth). New York: Free Press.

Rogers, Y. (1989). Icon Design for the User Interface. *International Reviews of Ergonomics: 3*, 129–155.

Rogers, Yvonne. (1986). Evaluating the meaningfulness of icon sets to represent command operations. Proceedings of the Second Conference of the British Computer Society, Human Computer Interaction Specialist Group on People and Computers: Designing for Usability, 586–603. Cambridge University Press.

Rogers, Yvonne, & Oborne, D. J. (1987). Pictorial communication of abstract verbs in relation to human-computer interaction. *British Journal of Psychology*, *78*(1), 99–112. https://doi.org/10.1111/j.2044-8295.1987.tb02229.x

Rohr, G., & Keppel, E. (1984). Iconic interfaces: Where to use and how to construct. In *Human factors in organizational design and management* (pp. 269–275). The Netherlands.

Rolland, V. (1909). *Planches de l'Armorial Général de J.-B. Rietstap* (Vols. 1–6). Paris: Institut Héraldique.

Rolland, V., & Rolland, H. (1969). *Supplement to the Armorial Général by J.-B. Rietstap* (2d ed., Vols. 1–8). London: Heraldry Today.

Rovenchak, A. (2011). Phoneme distribution, syllabic structure, and tonal patterns in Nko texts. *Mandenkan: Bulletin Semestriel d'études Linguistiques Mandé*, *47*, 77–96.

Rovenchak, A., Mačutek, J., & Riley, C. (2009). Distribution of complexities in the Vai script. *Glottometrics*, *18*, 1–12.

Rubin, D. C. (1995). Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes. Oxford University Press on Demand.

Sadakata, M., Desain, P., & Honing, H. (2006). The Bayesian Way to Relate Rhythm Perception and Production. *Music Perception: An Interdisciplinary Journal*, *23*(3), 269–288. https://doi.org/10.1525/mp.2006.23.3.269

Sadakata, M., Ohgushi, K., & Desain, P. (2004). A cross-cultural comparison study of the production of simple rhythmic patterns. *Psychology of Music*, *32*(4), 389–403. https://doi.org/10.1177/0305735604046097

Sammartino, J., & Palmer, S. E. (2012). Aesthetic issues in spatial composition: Effects of vertical position and perspective on framing single objects. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 865–879. https://doi.org/10.1037/a0027736

Sammler, D., Novembre, G., Koelsch, S., & Keller, P. E. (2013). Syntax in a pianist's hand: ERP signatures of "embodied" syntax processing in music. *Cortex*, *49*(5), 1325–1339.

Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature Communications*, *8*(15049).

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Savage, P. E., Merritt, E., Rzeszutek, T., & Brown, S. (2012). *CantoCore: A New Cross-Cultural Song Classification Scheme*. 25.

Savage, P. E., Tierney, A. T., & Patel, A. D. (2017). Global Music Recordings Support the Motor Constraint Hypothesis for Human and Avian Song Contour. *Music Perception: An Interdisciplinary Journal*, *34*(3), 327–334. https://doi.org/10.1525/mp.2017.34.3.327

Schmandt-Besserat, D. (2010). *How Writing Came About*. Austin, TX: University of Texas Press.

Scott-Phillips, T., Blancke, S., & Heintz, C. (2018). Four misunderstandings about cultural attraction. *Evolutionary Anthropology: Issues, News, and Reviews*, *27*(4), 162–173. https://doi.org/10.1002/evan.21716

Scott-Phillips, T. C. (2017). A (Simple) Experimental Demonstration that Cultural Evolution is not Replicative, but Reconstructive—And an Explanation of Why this Difference Matters. *Journal of Cognition and Culture*, *17*(1–2), 1–11. https://doi.org/10.1163/15685373-12342188

Selinger, P. (2003). Potrace: A polygon-based tracing algorithm.

Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, *6*(4), 469–471. https://doi.org/10.1098/rsbl.2009.1062

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Shennan, S. J., & Wilkinson, J. R. (2001). Ceramic Style Change and Neutral Evolution: A Case Study from Neolithic Europe. *American Antiquity*, *66*(4), 577–593. https://doi.org/10.2307/2694174

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of School Chinese: Implications for Learning to Read. *Child Development*, *74*(1), 27–47. https://doi.org/10.1111/1467-8624.00519

Sigurd, B., Eeg-Olofsson, M., & Weijer, J. V. (2004). Word length, sentence length and frequency – Zipf revisited. *Studia Linguistica*, *58*(1), 37–52. https://doi.org/10.1111/j.0039-3193.2004.00109.x

Simonton, D. K. (1996). Creative Expertise. In K. A. Ericsson (Ed.), *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (pp. 227–254). New York: Psychology Press.

Smith, R. D. (2012). Distinct word length frequencies: Distributions and symbol entropies. *ArXiv:1207.2334 [Physics]*. Retrieved from http://arxiv.org/abs/1207.2334

Smith, S. M. (1973). A Study of Prey-Attack Behaviour in Young Loggerhead Shrikes, Lanius ludovicianus L. *Behaviour*, *44*(1/2), 113–141.

Sperber, D. (1996a). *Explaining culture: A naturalistic approach*. Cambridge, MA: Cambridge.

Sperber, D. (1996b). Why are perfect animals, hybrids, and monsters food for symbolic thought? *Method & Theory in the Study of Religion*, *8*(2), 143–169. https://doi.org/10.1163/157006896X00170

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

Stammers, R., & Hoffman, J. (1991). Transfer between Icon Sets and Ratings of Icon Concreteness and Appropriateness. *Proceedings of the Human Factors Society Annual Meeting*, *35*(5), 354–358. https://doi.org/10.1177/154193129103500525

Sterelny, K. (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, 42–50. https://doi.org/10.1016/j.shpsc.2016.12.005

Stout, D. (2002). Skill and Cognition in Stone Tool Production: An Ethnographic Case Study from Irian Jaya. *Current Anthropology*, *43*(5), 693–722.

Stout, D. (2005). The Social and Cultural Context of Stone-knapping Skill Acquisition. In V. Roux & B. Bril (Eds.), *Stone knapping: The necessary conditions for a uniquely hominin behaviour* (pp. 331–340). Cambridge: McDonald Institute for Archaeological Research.

Strauss, U., Grzybek, P., & Altmann, G. (2006). Word Length and Word Frequency. In P. Grzybek (Ed.), *Contributions to the Science of Text and Language: Word Length Studies and Related Issues* (pp. 277–294). https://doi.org/10.1007/978-1-4020-4068-9_13

Stubbersfield, J. M., Flynn, E. G., & Tehrani, J. J. (2017). Cognitive Evolution and the Transmission of Popular Narratives: A Literature Review and Application to Urban Legends. *Evolutionary Studies in Imaginative Culture*, *1*(1), 121–136. https://doi.org/10.26613/esic.1.1.20

Stubbersfield, J. M., Tehrani, J. J., & Flynn, E. G. (2015). Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends. *British Journal of Psychology*, *106*(2), 288–307. https://doi.org/10.1111/bjop.12073

Tai, H.-C., Shen, Y.-P., Lin, J.-H., & Chung, D.-T. (2018). Acoustic evolution of old Italian violins from Amati to Stradivari. *Proceedings of the National Academy of Sciences*, *115*(23), 5926–5931. https://doi.org/10.1073/pnas.1800666115

Tamaoka, K., & Kiyama, S. (2013). The Effects of Visual Complexity for Japanese Kanji Processing with High and Low Frequencies. *Reading and Writing: An Interdisciplinary Journal*, *26*(2), 205–223. https://doi.org/10.1007/s11145-012-9363-x

Tamariz, M., & Kirby, S. (2015). Culture: Copying, Compression, and Conventionality. *Cognitive Science*, *39*(1), 171–183. https://doi.org/10.1111/cogs.12144

Tan, R., & Fay, N. (2011). Cultural transmission in the laboratory: Agent interaction improves the intergenerational transfer of information. *Evolution and Human Behavior*, *32*, 399–406.

Tehrani, J. J. (2013). The Phylogeny of Little Red Riding Hood. *PLOS ONE*, *8*(11), e78871. https://doi.org/10.1371/journal.pone.0078871

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Tennie, C., Walter, V., Gampe, A., Carpenter, M., & Tomasello, M. (2014). Limitations to the cultural ratchet effect in young children. *Journal of Experimental Child Psychology*, *126*, 152–160.

Testolin, A., Stoianov, I., & Zorzi, M. (2017). Letter perception emerges from unsupervised deep learning and recycling of natural image features. *Nature Human Behaviour*, *1*(9), 657–664. https://doi.org/10.1038/s41562-017-0186-2

The Unicode Standard, Version 11.0, (Mountain View, CA: The Unicode Consortium, 2018. ISBN 978-1-936213-19-1). (n.d.).

Thomas, E., Vandebroek, I., Van Damme, P., Goetghebeur, P., Douterlungne, D., Sanca, S., & Arrazola, S. (2009). The relation between accessibility, diversity and indigenous valuation of vegetation in the Bolivian Andes. *Journal of Arid Environments*, *73*(9), 854–861. https://doi.org/10.1016/j.jaridenv.2009.03.010

Thomas, H. (1968). Spatial Models and Multidimensional Scaling of Random Shapes. *The American Journal of Psychology*, *81*(4), 551–558. https://doi.org/10.2307/1421059

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535. https://doi.org/10.1073/pnas.1523631113

Thornton, T. P. (1996). *Handwriting in America: A cultural history*. Yale University Press.

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, *108*(37), 15510–15515. https://doi.org/10.1073/pnas.1103882108

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552.

Toussaint, G. T. (2012). The Pairwise Variability Index as a Tool in Musical Rhythm Analysis.

Treiman, R., & Kessler, B. (2011). Similarities Among the Shapes of Writing and Their Effects on Learning. *Written Language and Literacy*, *14*(1), 39–57. https://doi.org/10.1075/wll.14.1.03tre

Trudgill, P. (2011). *Sociolinguistic Typology*. Oxford: Oxford University Press.

Tyler, C. (2007). Some principles of spatial organization in art. *Spatial Vision*, *20*(6), 509–530.

Tyler, C. W. (1998). Painters centre one eye in portraits. *Nature*, *392*(6679), 877–878. https://doi.org/10.1038/31833

Tyler, C. W. (2007). Eye-Centering in Portraits: Reply to McManus and Thomas. *Perception*, *36*(2), 183–188. https://doi.org/10.1068/p5578b

Vale, G. L., Davis, S. J., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2017). Acquisition of a socially learned tool use sequence in chimpanzees: Implications for cumulative culture. *Evolution and Human Behavior*, *38*, 635–644.

van der Helm, P. A. (2014). Simplicity in Vision: A Multidisciplinary Account of Perceptual Organization. Cambridge University Press.

Vitányi, P. M. B., Balbach, F. J., Cilibrasi, R. L., & Li, M. (2009). Normalized Information Distance. In F. Emmert-Streib & M. Dehmer (Eds.), *Information Theory and Statistical Learning* (pp. 45–82). https://doi.org/10.1007/978-0-387-84816-7_3

von Hippel, E. (1988). *Sources of Innovation*. Oxford: Oxford University Press.

Wasielewski, H. (2014). Imitation Is Necessary for Cumulative Cultural Evolution in an Unfamiliar, Opaque Task. *Human Nature*, *25*, 161–179.

Watson, A. (2012). Perimetric Complexity of Binary Digital Images. *The Mathematica Journal*, *14*, 1–40. https://doi.org/10.3888/tmj.14-5

Wengrow, D. (2008). Prehistories of Commodity Branding. *Current Anthropology*, *49*(1), 7–34. https://doi.org/10.1086/523676

White, C., Marin, M., & Fessler, D. M. T. (2017). Not Just Dead Meat: An Evolutionary Account of Corpse Treatment in Mortuary Rituals. *Journal of Cognition and Culture*, *17*(1–2), 146–168. https://doi.org/10.1163/15685373-12342196

Whiten, A., Caldwell, C. A., & Mesoudi, A. (2016). Cultural diffusion in humans and other animals. *Current Opinion in Psychology*, *8*, 15–21.

Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(8), 1186–1203. https://doi.org/10.1037/xhp0000213

Wilson, B. (2012). Consider the fork: A history of how we cook and eat. Basic books.

Wimmer, G., & Altmann, G. (1996). The theory of word length: Some results and generalizations. *Glottometrika*, *15*, 112–133.

Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, *176*, 15–30. https://doi.org/10.1016/j.cognition.2018.03.002

Wolf, T., Vesper, C., Sebanz, N., Keller, P. E., & Knoblich, G. (2019). Combining Phase Advancement and Period Correction Explains Rushing during Joint Rhythmic Activities. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-45601-5

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*, 543–578.

Xiao, W., & Treiman, R. (2012). Iconicity of simple Chinese characters. *Behavior Research Methods*, *44*(4), 954–960. https://doi.org/10.3758/s13428-012-0191-3

Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1758), 20123073. https://doi.org/10.1098/rspb.2012.3073

Yue, G. (2015). The China Girl on the Margins of Film. *October*, *153*, 96–116. https://doi.org/10.1162/OCTO_a_00228

Zhang, J.-Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2007). Legibility Variations of Chinese Characters and Implications for Visual Acuity Measurement in Chinese Reading Population. *Investigative Ophthalmology & Visual Science*, *48*(5), 2383–2390. https://doi.org/10.1167/iovs.06-1195

Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human ecology. New York: Addison-Wesley.

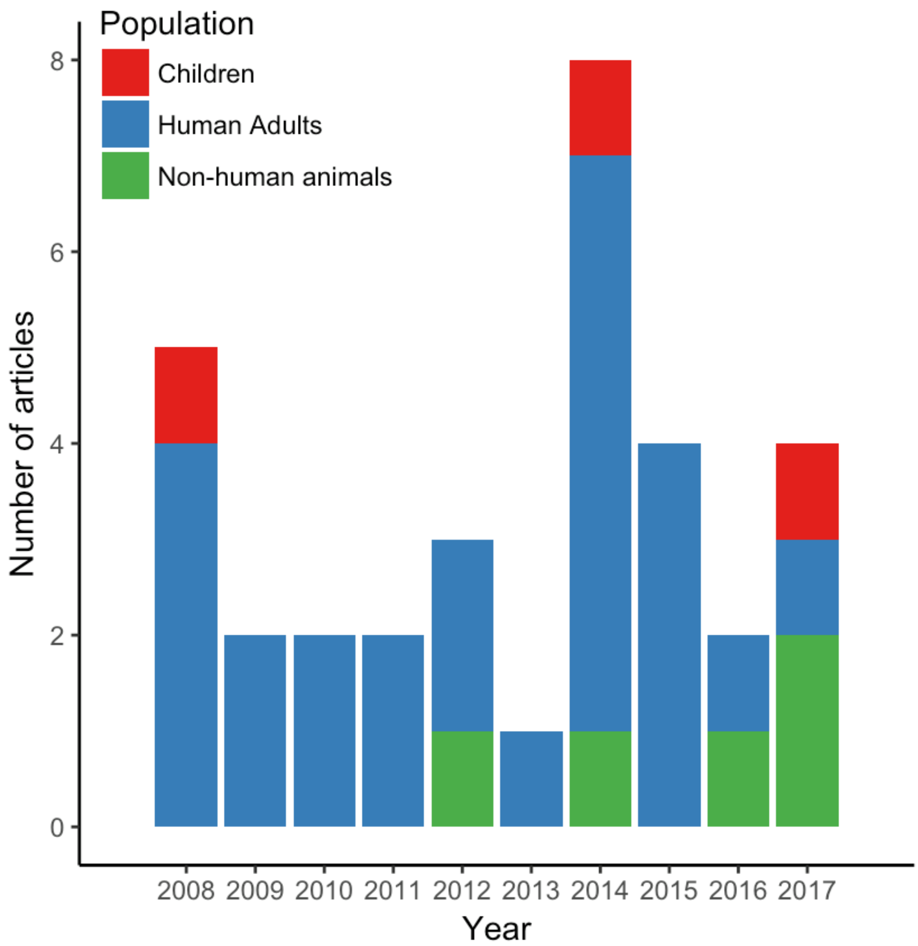
Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: Teaching is useful but not essential. *Scientific Reports*, *5*(16781).

Appendices

Appendix A – Supplementary materials for Cumulative culture in the laboratory

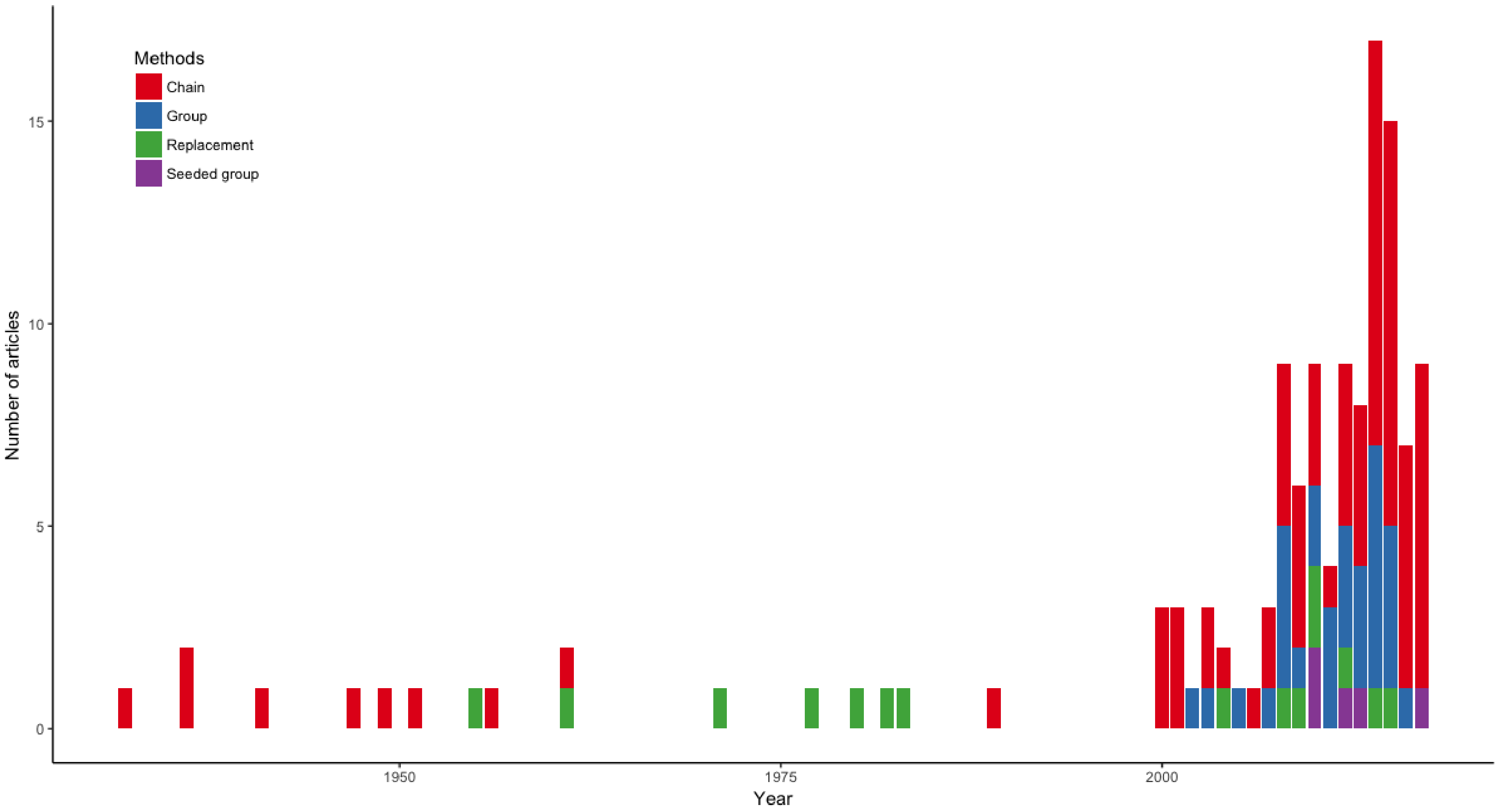
ESM 1 – Cultural transmission experiments incidence and additional bibliography

1. Incidence of cultural transmission experiments testing for cumulative culture by population



Number of cultural transmission experiments testing for cumulative culture by population from 2008 to 2017. N= 33, including 25 experiments with human adults, 3 experiments with children, and 5 experiments with non-humans (references available in ESM-2).

2. Incidence of cultural transmission experiments



Number of cultural transmission experiments ran on adult human participants by type (i.e. transmission chain, closed group, or replacement methods), from 1932 to 2017. N= 127, including 76 transmission chains experiments, 31 closed groups, 15 replacement methods and 5 seeded groups. This figure was compiled from the table of references below.

3. Cultural transmission experiments with humans (supplementary bibliography)

This list includes cultural transmission experiments (i.e., transmission chains, replacement, closed group, seeded group) with human participants, whether they test for cumulative culture or not. Experiments with children are marked by an asterisk. The list is not meant to be exhaustive.

Author	Year	Method	Full reference
Allport & Postman	1947	Chain	Allport GW, Postman L. The psychology of rumor. Oxford: Henry Bolt. 1947.
Atkinson et al.	2012	Closed Group	Atkisson C, O'Brien MJ, Mesoudi A. Adult learners in a novel environment use prestige-biased social learning. Evolutionary psychology. 2012 Jul 1;10(3):147470491201000309.
Bangerter	2000	Chain	Bangerter A. Transformation between scientific and social representations of conception: The method of serial reproduction. British Journal of Social Psychology. 2000 Dec 1;39(4):521-35.
Barrett & Nyhof	2001	Chain	Barrett JL, Nyhof MA. Spreading non-natural concepts: The role of intuitive conceptual structures in memory and transmission of cultural materials. Journal of cognition and culture. 2001 Feb 1;1(1):69-100.
Bartlett	1932	Chain	Bartlett FC. Remembering: An experimental and social study. Cambridge: Cambridge University. 1932.
Baum et al.	2004	Replacement	Baum WM, Richerson PJ, Efferson CM, Paciotti BM. Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. Evolution and Human Behavior. 2004 Sep 30;25(5):305-26.
Bebbington et al	2017	Chain	Bebbington K, MacLeod C, Ellison TM, Fay N. The sky is falling: evidence of a negativity bias in the social transmission of information. Evolution and Human Behavior. 2017 Jan 31;38(1):92-101.
Beppu & Griffiths	2009	Chain	Beppu, A., & Griffiths, T. L. (2009). Iterated Learning and the Cultural Ratchet. Proceedings of the 31st annual conference of the cognitive science society, 2089-2094.
Brissey	1961	Chain	Brissey FL. The factor of relevance in the serial reproduction of information. Journal of communication. 1961 Dec 1;11(4):211-9.
Caldwell & Eve	2014	Chain	Caldwell CA, Eve RM. Persistence of contrasting traditions in cultural evolution: Unpredictable payoffs generate slower rates of cultural change. PloS one. 2014 Jun 18;9(6):e99708.
Caldwell & Millen	2008	Replacement	Caldwell CA, Millen AE. Experimental models for testing hypotheses about cumulative cultural evolution. Evolution and Human Behavior. 2008 May 31;29(3):165-71.
Caldwell & Millen	2009	Replacement	Caldwell CA, Millen AE. Social learning mechanisms and cumulative cultural evolution: is imitation necessary?. Psychological Science. 2009 Dec;20(12):1478-83.
Caldwell & Millen	2010	Replacement	Caldwell, C. A., & Millen, A. E. (2010). Conservatism in laboratory microsocieties: unpredictable payoffs

			accentuate group-specific traditions. Evolution and human behavior, 31(2), 123-130.
Caldwell & Millen	2010	Replacement	Caldwell CA, Millen AE. Human cumulative culture in the laboratory: effects of (micro) population size. Learning & Behavior. 2010 Aug 1;38(3):310-8.
Caldwell & Smith	2012	Replacement	Caldwell CA, Smith K. Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. PloS one. 2012 Aug 23;7(8):e43807.
Canini et al.	2014	Chain	Canini KR, Griffiths TL, Vanpaemel W, Kalish ML. Revealing human inductive biases for category learning by simulating cultural transmission. Psychonomic bulletin & review. 2014 Jun 1;21(3):785-93.
Carr et al	2017	Chain	Carr JW, Smith K, Cornish H, Kirby S. The cultural evolution of structured languages in an open-ended, continuous world. Cognitive science. 2017 May 1;41(4):892-923.
Clark & Kashima	2007	Chain	Clark AE, Kashima Y. Stereotypes help people connect with others in the community: a situated functional analysis of the stereotype consistency bias in communication. Journal of personality and social psychology. 2007 Dec;93(6):1028.
Connor et al.	2016	Chain	Connor P, Harris E, Guy S, Fernando J, Shank DB, Kurz T, Bain PG, Kashima Y. Interpersonal communication about climate change: how messages change when communicated through simulated online social networks. Climatic change. 2016 Jun 1;136(3-4):463-76.
Cook et al.	2014	Closed Group	Cook JL, den Ouden HE, Heyes CM, Cools R. The social dominance paradox. Current Biology. 2014 Dec 1;24(23):2812-6.
Derex & Boyd	2015	Closed Group	Derex M, Boyd R. The foundations of the human cultural niche. Nature communications. 2015 Sep 24;6.
Derex & Boyd	2016	Closed Group	Derex M, Boyd R. Partial connectivity increases cultural accumulation within groups. Proceedings of the National Academy of Sciences. 2016 Mar 15;113(11):2982-7.
Derex et al.	2012	Closed Group	Derex M, Godelle B, Raymond M. Social learners require process information to outperform individual learners. Evolution. 2013 Mar 1;67(3):688-97.
Derex et al.	2013	Closed Group	Derex M, Beugin MP, Godelle B, Raymond M. Experimental evidence for the influence of group size on cultural complexity. Nature. 2013 Nov 21;503(7476):389.
Derex et al.	2014	Closed Group	Derex M, Godelle B, Raymond M. How does competition affect the transmission of information?. Evolution and Human Behavior. 2014 Mar 31;35(2):89-95.
Derex et al.	2015	Closed Group	Derex M, Feron R, Godelle B, Raymond M. Social learning and the replication process: an experimental investigation. InProc. R. Soc. B 2015 Jun 7 (Vol. 282, No. 1808, p. 20150719). The Royal Society.
DiFonzo et al.	2013	Closed Group	DiFonzo N, Bourgeois MJ, Suls J, Homan C, Stupak N, Brooks BP, Ross DS, Bordia P. Rumor clustering, consensus, and polarization: Dynamic social impact and

			self-organization of hearsay. Journal of Experimental Social Psychology. 2013 May 31;49(3):378-99.
DiFonzo et al.	2014	Closed Group	DiFonzo N, Suls J, Beckstead JW, Bourgeois MJ, Homan CM, Brougher S, Younge AJ, Terpstra-Schwab N. Network structure moderates intergroup differentiation of stereotyped rumors. Social Cognition. 2014 Oct;32(5):409-48.
Efferson et al.	2007	Closed Group	Efferson C, Richerson PJ, McElreath R, Lubell M, Edsten E, Waring TM, Paciotti B, Baum W. Learning, productivity, and noise: an experimental study of cultural transmission on the Bolivian Altiplano. Evolution and Human Behavior. 2007 Jan 31;28(1):11-7.
Efferson et al.	2008	Closed Group	Efferson C, Lalive R, Richerson PJ, McElreath R, Lubell M. Conformists and mavericks: the empirics of frequency-dependent cultural transmission. Evolution and Human Behavior. 2008 Jan 31;29(1):56-64.
Eriksson & Coultas	2012	Chain	Eriksson K, Coultas JC. The advantage of multiple cultural parents in the cultural transmission of stories. Evolution and human behavior. 2012 Jul 31;33(4):251-9.
Eriksson & Coultas	2014	Chain	Eriksson K, Coultas JC. Corpses, maggots, poodles and rats: emotional selection operating in three phases of cultural transmission of urban legends. Journal of Cognition and Culture. 2014 Jan 30;14(1-2):1-26.
Fay et al.	2010	Closed Group	Fay N, Garrod S, Roberts L, Swoboda N. The interactive evolution of human communication systems. Cognitive science. 2010 Apr 1;34(3):351-86.
* Flynn	2008	Chain	Flynn E. Investigating children as cultural magnets: do young children transmit redundant information along diffusion chains?. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2008 Nov 12;363(1509):3541-51.
* Flynn & Whiten	2008	Chain	Flynn E, Whiten A. Cultural transmission of tool use in young children: A diffusion chain study. Social Development. 2008 Aug 1;17(3):699-718.
*Flynn & Whiten	2010	Seeded group	Flynn E, Whiten A. Studying children's social learning experimentally "in the wild". Learning & Behavior. 2010 Aug 1;38(3):284-96.
*Flynn & Whiten	2012	Seeded group	Flynn E, Whiten A. Experimental "microcultures" in young children: Identifying biographic, cognitive, and social predictors of information transmission. Child development. 2012 May 1;83(3):911-25.
Griffiths, Christian & Kalish	2008	Chain	Griffiths TL, Christian BR, Kalish ML. Using category structures to test iterated learning as a method for identifying inductive biases. Cognitive Science. 2008 Jan 2;32(1):68-107.
Griffiths, Lewandowsky & Kalish	2013	Chain	Griffiths TL, Lewandowsky S, Kalish ML. The effects of cultural transmission are modulated by the amount of information transmitted. Cognitive science. 2013 Jul 1;37(5):953-67.
Hall	1951	Chain	Hall KR. The effect of names and titles upon the serial reproduction of pictorial and verbal material. British Journal of Psychology. 1950 Dec 1;41(3-4):109-21.

Heath et al	2001	Chain	Heath C, Bell C, Sternberg E. Emotional selection in memes: the case of urban legends. Journal of personality and social psychology. 2001 Dec;81(6):1028.
Hopper et al.	2010	Chain	Hopper LM, Flynn EG, Wood LA, Whiten A. Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. Journal of experimental child psychology. 2010 May 31;106(1):82-97.
Hunzaker	2014	Chain	Hunzaker MF. Making sense of misfortune: Cultural schemas, victim redefinition, and the perpetuation of stereotypes. Social Psychology Quarterly. 2014 Jun;77(2):166-84.
Hunzaker	2016	Chain	Hunzaker MF. Cultural Sentiments and Schema-Consistency Bias in Information Transmission. American Sociological Review. 2016 Dec;81(6):1223-50.
Hutchison et al	2017	Chain	Hutchison J, Cunningham SJ, Slessor G, Urquhart J, Smith K, Martin D. Context and Perceptual Salience Influence the Formation of Novel Stereotypes via Cumulative Cultural Evolution. Cognitive science. 2017 Nov 2.
Imada & Yussen	2012	Chain	Imada T, Yussen SR. Reproduction of cultural values: A cross-cultural examination of stories people create and transmit. Personality and Social Psychology Bulletin. 2012 Jan;38(1):114-28.
Insko et al.	1980	Replacement	Insko CA, Thibaut JW, Moehle D, Wilson M, Diamond WD, Gilmore R, Solomon MR, Lipsitz A. Social evolution and the emergence of leadership. Journal of Personality and Social Psychology. 1980 Sep;39(3):431.
Insko et al.	1982	Replacement	Insko CA, Gilmore R, Moehle D, Lipsitz A, Drenan S, Thibaut JW. Seniority in the generational transition of laboratory groups: The effects of social familiarity and task experience. Journal of Experimental Social Psychology. 1982 Nov 1;18(6):557-80.
Insko et al.	1983	Replacement	Insko CA, Gilmore R, Drenan S, Lipsitz A, Moehle D, Thibaut J. Trade versus expropriation in open groups: A comparison of two types of social power. Journal of Personality and Social Psychology. 1983 May;44(5):977.
Jacobs & Campbell	1961	Replacement	Jacobs RC, Campbell DT. The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. The Journal of Abnormal and Social Psychology. 1961 May;62(3):649.
Jacoby & McDermott	2017	Chain	Jacoby N, McDermott JH. Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. Current Biology. 2017 Feb 6;27(3):359-70.
Kalish, Griffiths & Lewandowsky	2007	Chain	Kalish ML, Griffiths TL, Lewandowsky S. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. Psychonomic Bulletin & Review. 2007 Apr 1;14(2):288-94.
Kameda & Nakanishi	2002	Closed Group	Kameda T, Nakanishi D. Cost–benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. Evolution and Human Behavior. 2002 Sep 30;23(5):373-93.

Kameda & Nakanishi	2003	Closed Group	Kameda T, Nakanishi D. Does social/cultural learning increase human adaptability?: Rogers's question revisited. Evolution and Human Behavior. 2003 Jul 31;24(4):242-60.
Kashima	2000	Chain	Kashima Y. Maintaining cultural stereotypes in the serial reproduction of narratives. Personality and Social Psychology Bulletin. 2000 May;26(5):594-604.
Kashima et al.	2013	Chain	Kashima Y, Lyons A, Clark A. The maintenance of cultural stereotypes in the conversational retelling of narratives. Asian Journal of Social Psychology. 2013 Mar 1;16(1):60-70.
Kempe & Mesoudi	2014	Chain	Kempe M, Mesoudi A. An experimental demonstration of the effect of group size on cultural accumulation. Evolution and Human Behavior. 2014 Jul 31;35(4):285-90.
* Kempe et al.	2015	Chain	Kempe V, Gauvrit N, Forsyth D. Structure emerges faster during cultural transmission in children than in adults. Cognition. 2015 Mar 31;136:247-54.
Kempe et al.	2012	Chain	Kempe M, Lycett S, Mesoudi A. An experimental test of the accumulated copying error model of cultural mutation for Acheulean handaxe size. PLoS One. 2012 Nov 8;7(11):e48333.
Kirby et al.	2008	Chain	Kirby, S., Cornish, H., & Smith, K. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. Proceedings of the National Academy of Sciences 2008 105, 10681-10686.
Kirby et al.	2015	Chain	Kirby S, Tamariz M, Cornish H, Smith K. Compression and communication in the cultural evolution of linguistic structure. Cognition. 2015 Aug 31;141:87-102.
Kurke, Weick & Ravlin	1989	Chain	Kurke LB, Weick KE, Ravlin EC. Can Information Loss Be Reversed: Evidence for Serial Reconstruction. Communication Research. 1989 Feb;16(1):3-24.
Lewandowsky, Griffiths, & Kalish	2009	Chain	Lewandowsky S, Griffiths TL, Kalish ML. The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. Cognitive science. 2009 Aug 1;33(6):969-98.
Lyons & Kashima	2003	Chain	Lyons A, Kashima Y. How are stereotypes maintained through communication? The influence of stereotype sharedness. Journal of personality and social psychology. 2003 Dec;85(6):989.
Lyons & Kashima	2001	Chain	Lyons A, Kashima Y. The reproduction of culture: Communication processes tend to maintain cultural stereotypes. Social cognition. 2001 Jul 1;19(3: Special issue):372-94.
Martin et al.	2014	Chain	Martin D, Hutchison J, Slessor G, Urquhart J, Cunningham SJ, Smith K. The spontaneous formation of stereotypes via cumulative cultural evolution. Psychological Science. 2014 Sep;25(9):1777-86.
Matthews et al.	2012	Chain	Matthews C, Roberts G, Caldwell CA. Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. Evolution and Human Behavior. 2012 Nov 30;33(6):759-70.
Maxwell	1936	Chain	Maxwell RS. Remembering in different social groups. British Journal of Psychology. 1936 Jul 1;27(1):30-40.

McElreath et al.	2005	Closed Group	McElreath R, Lubell M, Richerson PJ, Waring TM, Baum W, Edsten E, Efferson C, Paciotti B. Applying evolutionary models to the laboratory study of social learning. Evolution and Human Behavior. 2005 Nov 30;26(6):483-508.
McElreath et al.	2008	Closed Group	McElreath R, Bell AV, Efferson C, Lubell M, Richerson PJ, Waring T. Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2008 Nov 12;363(1509):3515-28.
* McGuigan & Graham	2009	Chain	McGuigan N, Graham M. Cultural transmission of irrelevant tool actions in diffusion chains of 3-and 5-year-old children. European Journal of Developmental Psychology. 2010 Sep 1;7(5):561-77.
* McGuigan & Cubillo	2013	Seeded group	McGuigan N, Cubillo M. Information transmission in young children: When social information is more important than nonsocial information. The Journal of genetic psychology. 2013 Nov 1;174(6):605-19.
* McGuigan et al	2017	Seeded group	McGuigan N, Burdett E, Burgess V, Dean L, Lucas A, Vale G, Whiten A. Innovation and social transmission in experimental micro-societies: exploring the scope of cumulative culture in young children. Phil. Trans. R. Soc. B. 2017 Dec 5;372(1735):20160425.
Mesoudi	2008	Closed Group	Mesoudi A. An experimental simulation of the "copy-successful-individuals" cultural learning strategy: adaptive landscapes, producer–scrounger dynamics, and informational access costs. Evolution and Human Behavior. 2008 Sep 30;29(5):350-63.
Mesoudi	2011	Closed Group	Mesoudi A. An experimental comparison of human social learning strategies: payoff-biased social learning is adaptive but underused. Evolution and Human Behavior. 2011 Sep 30;32(5):334-42.
Mesoudi & O'Brien	2008	Closed Group	Mesoudi A, O'Brien MJ. The cultural transmission of Great Basin projectile-point technology I: an experimental simulation. American Antiquity. 2008 Jan;73(1):3-28.
Mesoudi & Whiten	2004	Chain	Mesoudi A, Whiten A. The hierarchical transformation of event knowledge in human cultural transmission. Journal of cognition and culture. 2004 Mar 1;4(1):1-24.
Mesoudi et al.	2015	Closed Group	Mesoudi A, Chang L, Murray K, Lu HJ. Higher frequency of social learning in China than in the West shows cultural variation in the dynamics of cultural evolution. Proceedings of the Royal Society of London B: Biological Sciences. 2015 Jan 7;282(1798):20142209.
Mesoudi, Whiten & Dunbar	2006	Chain	Mesoudi A, Whiten A, Dunbar R. A bias for social information in human cultural transmission. British Journal of Psychology. 2006 Aug 1;97(3):405-23.
Miton et al.	2015	Chain	Miton H, Claidière N, Mercier H. Universal cognitive mechanisms explain the cultural success of bloodletting. Evolution and Human Behavior. 2015 Jul 31;36(4):303-12.
Molleman et al.	2014	Closed Group	Molleman L, Van den Berg P, Weissing FJ. Consistent individual differences in human social learning strategies. Nature Communications. 2014 Apr 1;5:3570.

Morgan et al.	2015	Chain	Morgan TJ, Uomini NT, Rendell LE, Chouinard-Thuly L, Street SE, Lewis HM, Cross CP, Evans C, Kearney R, De La Torre I, Whiten A. Experimental evidence for the co-evolution of hominin tool-making teaching and language. Nature communications. 2015 Jan 13;6.
Morgan et al.	2012	Closed Group	Morgan TJ, Rendell LE, Ehn M, Hoppitt W, Laland KN. The evolutionary basis of human social learning. Proceedings of the Royal Society of London B: Biological Sciences. 2012 Feb 22;279(1729):653-62.
Moussaid et al.	2015	Chain	Moussaïd M, Brighton H, Gaissmaier W. The amplification of risk in experimental diffusion chains. Proceedings of the National Academy of Sciences. 2015 May 5;112(18):5631-6.
Moussaid et al.	2017	Chain	Moussaïd M, Herzog SM, Kämmer JE, Hertwig R. Reach and speed of judgment propagation in the laboratory. Proceedings of the National Academy of Sciences. 2017 Apr 3:201611998.
Muthukrishna et al.	2014	Chain	Muthukrishna M, Shulman BW, Vasilescu V, Henrich J. Sociality influences cultural complexity. Proceedings of the Royal Society of London B: Biological Sciences. 2014 Jan 7;281(1774):20132511.
Northway	1936	Chain	Northway ML. The influence of age and social group on children's remembering. British Journal of Psychology. 1936 Jul 1;27(1):11-29.
Oishi et al.	2014	Chain	Oishi S, Kesebir S, Eggleston C, Miao FF. A hedonic story has a transmission advantage over a eudaimonic story. Journal of Experimental Psychology: General. 2014 Dec;143(6):2153.
Rafferty et al.	2013	Chain	Rafferty AN, Griffiths TL, Ettlinger M. Greater learnability is not sufficient to produce cultural universals. Cognition. 2013 Oct 31;129(1):70-87.
Ravignani et al	2016	Chain	Ravignani A, Delgado T, Kirby S. Musical evolution in the lab exhibits rhythmic universals. Nature Human Behaviour. 2016 Dec 19;1:0007.
Reali & Griffiths	2009	Chain	Reali F, Griffiths TL. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. Cognition. 2009 Jun 30;111(3):317-28.
Rose & Felton	1955	Replacement	Rose E, Felton W. Experimental histories of culture. American Sociological Review. 1955 Aug 1;20(4):383-92.
Schillinger et al.	2016	Chain	Schillinger K, Mesoudi A, Lycett SJ. Copying error, evolution, and phylogenetic signal in artifactual traditions: An experimental approach using “model artifacts”. Journal of Archaeological Science. 2016 Jun 30;70:23-34.
Schotter & Sopher	2003	Chain	Schotter A, Sopher B. Social learning and coordination conventions in intergenerational games: An experimental study. Journal of Political Economy. 2003 Jun;111(3):498-529.
Scott-Phillips	2017	Chain	Scott-Phillips TC. A (simple) experimental demonstration that cultural evolution is not replicative, but reconstructive—and an explanation of why this difference matters. Journal of Cognition and Culture. 2017 Feb 8;17(1-2):1-1.

Silvey et al	2015	Chain	Silvey C, Kirby S, Smith K. Word meanings evolve to selectively preserve distinctions on salient dimensions. Cognitive science. 2015 Jan 1;39(1):212-26.
Smith & Wonnacott	2010	Chain	Smith K, Wonnacott E. Eliminating unpredictable variation through iterated learning. Cognition. 2010 Sep 30;116(3):444-9.
Smith et al	2017	Chain	Smith K, Perfors A, Fehér O, Samara A, Swoboda K, Wonnacott E. Language learning, language use and the evolution of linguistic variation. Phil. Trans. R. Soc. B. 2017 Jan 5;372(1711):20160051.
Stubbersfield et al.	2015	Chain	Stubbersfield JM, Tehrani JJ, Flynn EG. Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends. British Journal of Psychology. 2015 May 1;106(2):288-307.
Stubbersfield et al.	2017	Chain	Stubbersfield JM, Tehrani JJ, Flynn EG. Chicken Tumours and a Fishy Revenge: Evidence for Emotional Content Bias in the Cumulative Recall of Urban Legends. Journal of Cognition and Culture. 2017 Feb 8;17(1-2):12-26.
Talland	1956	Chain	Talland GA. Cultural differences in serial reproduction. The Journal of Social Psychology. 1956 Feb 1;43(1):75-81.
Tamariz & Kirby	2015	Chain	Tamariz M, Kirby S. Culture: copying, compression, and conventionality. Cognitive science. 2015 Jan 1;39(1):171-83.
Tamariz et al.	2014	Closed Group	Tamariz M, Ellison TM, Barr DJ, Fay N. Cultural selection drives the evolution of human communication systems. Proceedings of the Royal Society of London B: Biological Sciences. 2014 Aug 7;281(1788):20140488.
Tamariz et al	2016	Chain	Tamariz M, Kirby S, Carr JW. Cultural evolution across domains: language, technology and art. InProceedings of the 38th Annual Conference of the Cognitive Science Society 2016. Austin, TX: Cogn. Sci. Soc. In press.
Tamariz et al	2017	Chain	Tamariz M, Roberts SG, Martínez JI, Santiago J. The Interactive Origin of Iconicity. Cognitive Science. 2017 May 15.
Tan & Fay	2011	Chain	Tan R, Fay N. Cultural transmission in the laboratory: agent interaction improves the intergenerational transfer of information. Evolution and Human Behavior. 2011 Nov 30;32(6):399-406.
Tennie et al.	2014	Chain	Tennie C, Walter V, Gampe A, Carpenter M, Tomasello M. Limitations to the cultural ratchet effect in young children. Journal of experimental child psychology. 2014 Oct 31;126:152-60.
Thompson, Judd & Park	2000	Chain	Thompson MS, Judd CM, Park B. The consequences of communicating social stereotypes. Journal of Experimental Social Psychology. 2000 Nov 30;36(6):567-99.
Toelch et al.	2009	Closed Group	Toelch U, van Delft MJ, Bruce MJ, Donders R, Meeus MT, Reader SM. Decreased environmental variability induces a bias for social information use in humans. Evolution and Human Behavior. 2009 Jan 31;30(1):32-40.
Toelch et al.	2010	Closed Group	Toelch U, Bruce MJ, Meeus MT, Reader SM. Humans copy rapidly increasing choices in a multiarmed bandit

			problem. Evolution and human behavior. 2010 Sep 30;31(5):326-33.
Toelch et al.	2011	Closed Group	Toelch U, Bruce MJ, Meeus MT, Reader SM. Social performance cues induce behavioral flexibility in humans. Frontiers in psychology. 2011;2.
Toelch et al.	2014	Closed Group	Toelch U, Bruce MJ, Newson L, Richerson PJ, Reader SM. Individual consistency and flexibility in human social information use. Proceedings of the Royal Society of London B: Biological Sciences. 2014 Feb 7;281(1776):20132864.
Tresselt & Spragg	1941	Chain	Tresselt ME, Spragg SD. Changes occurring in the serial reproduction of verbally perceived materials. The Pedagogical Seminary and Journal of Genetic Psychology. 1941 Jun 1;58(2):255-64.
van den Berg et al.	2015	Closed Group	van den Berg P, Molleman L, Weissing FJ. Focus on the success of others leads to selfish behavior. Proceedings of the National Academy of Sciences. 2015 Mar 3;112(9):2912-7.
Verhoef et al.	2014	Chain	Verhoef T, Kirby S, de Boer B. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. Journal of Phonetics. 2014 Mar 31;43:57-68.
Ward	1949	Chain	Ward TH. An experiment on serial reproduction with special reference to the changes in the design of early coin types. British Journal of Psychology. 1949 Mar 1;39(3):142-7.
Wasielewski	2014	Replacement	Wasielewski H. Imitation is necessary for cumulative cultural evolution in an unfamiliar, opaque task. Human Nature. 2014 Mar 1;25(1):161-79.
Weick & Gilfillan	1971	Replacement	Weick KE, Gilfillan DP. Fate of arbitrary traditions in a laboratory microculture. Journal of personality and Social Psychology. 1971 Feb;17(2):179.
*Whiten & Flynn	2010	Seeded group	Whiten A, Flynn E. The transmission and evolution of experimental microcultures in groups of young children. Developmental psychology. 2010 Nov;46(6):1694.
Winters et al	2015	Chain	Winters J, Kirby S, Smith K. Languages adapt to their contextual niche. Language and Cognition. 2015 Sep;7(3):415-49.
Wisdom et al.	2013	Closed Group	Wisdom TN, Song X, Goldstone RL. Social learning strategies in networked groups. Cognitive Science. 2013 Nov 1;37(8):1383-425.
Wisdom & Goldstone	2011	Closed Group	Wisdom TN, Goldstone RL. Innovation, Imitation, and Problem Solving in a Networked Group. Nonlinear Dynamics-Psychology and Life Sciences. 2011 Apr 1;15(2):229.
Xu & Griffths	2010	Chain	Xu J, Griffiths TL. A rational analysis of the effects of memory biases on serial reproduction. Cognitive psychology. 2010 Mar 31;60(2):107-26.
Xu et al.	2013	Chain	Xu J, Dowman M, Griffiths TL. Cultural transmission results in convergence towards colour term universals. InProc. R. Soc. B 2013 May 7 (Vol. 280, No. 1758, p. 20123073). The Royal Society.
Yeung & Griffiths	2015	Chain	Yeung S, Griffiths TL. Identifying expectations about the strength of causal relationships. Cognitive psychology. 2015 Feb 28;76:1-29.

Zucker	1977	Replacement	Zucker LG. The role of institutionalization in cultural persistence. American sociological review. 1977 Oct 1:726-43.
Zwirner & Thornton	2015	Chains	Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: teaching is useful but not essential. Scientific Reports, 5(16781).

ESM 2 - Learning times in cultural transmission experiments testing for cumulative culture

1. How to calculate learning times

We refer to the sum of all the learning and trial time invested by an individual participant in a non-social condition as the *total learning time of an individual*. The total learning time of an individual is obtained by summing up the time of all trials an individual goes through during the full length of the experiment.

We refer to the sum of the time spent learning and performing a task by all the participants of a chain or group during one run of an experiment as the *total additive learning time of a tradition*. The total additive learning time of a tradition is obtained by adding the learning time of each individual taking part in the tradition.

For transmission chains and closed group experiments, all participants in a chain or in a group all have equal individual learning times. Consequently, the additive learning time of a tradition equals the individual learning time multiplied by the number of participants in the chain or in the group.

In the case of replacement designs, not all participants have an equal individual learning times. The initial and last participants often have less time than the participants in between. The total additive learning time of a tradition is calculated by adding the individual learning times of each individual.

For seeded group designs, the time spent by an individual observing and manipulating the apparatus is generally not reported. We thus calculated learning times by summing up the time available for participants to solve the task. Because the time spent training the seed participant is often not reported, we have decided not to include it in our calculations.

The learning times computed here are based on the time reported in the description of the methods for each paper. They are to be considered as a theoretical upper limit, however, since the effective time spent by each participant learning the task is not reported (see section 3 below). We included in the learning times both the time spent observing an individual doing a task and the time spent by the participant doing the task. Relevant comparison can then be made between the total learning time of an individual in non-social conditions (indicated in bold) and total additive learning times of traditions in social conditions.

2. Comparative table of learning times

We included all papers from ESM-1 (i.e., using one of the four CTE designs discussed in the main text) that were explicitly presented by their author(s) as testing for cumulative culture.

Article, *Method*	Individual condition?	Conditions	Total learning time of an individual	Total additive learning time of a tradition
Beppu & Griffiths (2009,) *Chain*	No	Two Examples	2 examples x 5 trials = 10 trials	30 generations x 10 trials = 300 trials total
		Four Examples	4 examples x 5 trials = 20 trials	15 generations x 20 trials = 300 trials total
		Six Examples	6 examples x 5 trials = 30 trials	10 generations x 30 trials = 300 trials total
Caldwell & Eve (2014), *Chains*	No	Predictable (Cubic / Tripod)	3 minutes observation time + 7 minutes trial time = 10 minutes	10 minutes x 5 generations = 50 minutes total
		Unpredictable (Cubic / Tripod)		
Caldwell & Millen (2008), *Replacement*	No	N/A	5 minutes observation, 5 minutes production = 10 minutes	5 minutes observation time, 5 minutes building time; except participant one with no learning time; except participant 2 with only 2.5 minutes of learning time;

			(except for the two first participants, whose observation time was less than that, respectively none and 2 minutes 30 seconds)	= 42.5 minutes of learning time total and 50 minutes of trial time (total time = 92.5 minutes)
Caldwell & Millen (2009), *Replacement*	No	Full-information (Actions Results Teaching)	5 minutes observation, 5 minutes building time = 10 Minutes, except participant 1, no observation time, and participant 2, only 2.5 minutes observation	10 generations x 10 minutes, except participant 1, no observation time, and participant 2, only 2.5 minutes observation = 92.5 minutes total (42.5 minutes observation, 50 minutes building)
		A condition (actions only)		
		AR condition (actions and results)		
		AT condition (actions and teaching)		
		R condition (results only)	5 minutes building time (total)	10 generations x 5 minutes = 50 minutes
		RT condition (results and teaching)		
		T condition (teaching only)		
Caldwell & Millen (2010conservatism), *Replacement*	No	Immediate measurement	5 minutes observation time, 5 minutes building time = 10 minutes (except participant one with no observation time = 5 minutes; except participant 2 with only 2.5 minutes of observation time = 7.5 minutes)	10 participants x 10 minutes except participant one with no observation time; except participant 2 with only 2.5 minutes of observation time = 42.5 minutes of observation time total and 50 minutes of trial time (total time = 92.5 minutes)
		Delayed measurement		

Caldwell & Millen (2010pop-size), *Replacement*	No	One-model condition	5 minutes observation + 5 minutes trial per participant = 10 minutes; except participant 1 that had no observation time in all conditions	10 participants x 10 minutes except participant one with no observation time = 45 minutes of observation time total and 50 minutes of trial time (total time = 95 minutes)
		Two-model condition	Same, except participant 2 (only 2.5 minutes observation time)	10 participants x 10 minutes except participant one with no observation time, and participant 2 with only 2.5 minutes = 42.5 minutes of observation time total and 50 minutes of trial time (total time = 92.5 minutes)
		Three-model condition	Same, except participants 2 and 3 (respectively 1.66 and 3.33 minutes)	10 participants x 10 minutes except participant one with no observation time, participant 2 with only 1.66 minutes and participant 3 with only 3.33 minutes observation in three-model condition = 40 minutes of observation time total and 50 minutes of trial time (total time = 90 minutes)
Caldwell & Smith 2012, *Replacement*	N/A	N/A	Groups of 4, drawer/matchers roles: from 1 round (drawing - participant 1, or guessing - participant 10) to 4 rounds (3 guessing, 1 drawing, participants 4 to 7)	7 Rounds drawing total + 3x7rounds guessing total = 28 rounds

Claidière et al (2014), *Chains*	N/A	N/A	25 900 random trials on average + 50 transmission trials / chain the participant takes part to = 25950 trials	25950 trials x 12 generations = 311 400 trials
Davis et al. 2016, *Seeded group*	No	Social information group	Training phase (3 demonstrations of the inefficient method + practice until they had successfully retrieved the token a minimum of 20 times over no fewer than 2 training sessions) + Demonstration phase, at least 10 demonstrations per participant on at least two sessions + Testing phase, Apparatus presented over ten hours	(Training phase + Testing phase) x group of 4 to 5
		Non-seeded group	Training phase (3 demonstrations of the inefficient method + practice until they had successfully retrieved the token a minimum of 20 times over no fewer than 2 training sessions) + Testing phase, Apparatus presented over ten hours	(Training phase + Testing phase) x group of 6
		Naïve group	Apparatus presented for the test phase during over ten hours	10 hours x groups of 5 = 50 hours

Dean et al. 2012, *Seeded group*	No	Chimpanzees, Open	30 hours of exposure	30 hours x groups of 8 & 10 = 240 & 300 hours
		Chimpanzees, Scaffolded	30 hours of exposure	30 hours of Groups of 8 & 7 = 240 & 210 hours
		Capuchins, Scaffolded	53 trials of 1 hour = 53 hours	53 hours x cohort of 22 (in2007) or 17 (in 2008) = 1166 or 901 hours
		Children (Open/Scaffolded)	5 trials x 30 minutes = 2 hours 30	2.5 hours x groups of 4 or 5 = 10 or 12.5 hours
		Chimpanzees, Seeded (Expt 2)	8 trials x 3 hours = 24 hours	24 hours x Groups of 8 to 13 = 192 to 312 hours
Derex et al. (2012) – process information, *Closed-group*	Yes	Individual learning	180s (construction period) + 30s (information period) X 15 trials = 52.5 minutes total	N/A
		Product-copying (Social learning)	180s (construction period) + 90s (information period) = 270 seconds X 15 trials = 67.5 minutes total	67.5 minutes x 5 participants per group = 337.5 minutes
		Process-copying (Social learning)	180s (construction period) + 90s (information period) = 270 seconds X 15 trials = 67.5 minutes total	67.5 minutes x 5 participants per group = 337.5 minutes

Derex et al. (2013) – group size complexity, *Closed Group*	No	Group size = 2	15 trials, each trial including an Information period of 70s + a Construction period of 90s, i.e., each trial = 160 seconds, 15 x 160 seconds = 40 minutes	40 minutes x 2 participants = 80 minutes
		Group size = 4		40 minutes x 4 participants = 160 minutes
		Group size = 8		40 minutes x 8 participants = 320 minutes
		Group size = 16		40 minutes x 16 participants = 640 minutes
Derex et al. (2014), *Closed Group*	No	Within-group competition	180s (construction period) + 90s (transaction period) X 15 trials = 67.5 minutes per participant	within-group competition treatment: 5 players/game [5 X 67.5 minutes] = 337.5 minutes total
		Between-group competition		between-group competition treatment: 5 players/game X 2 groups [5 X 2 X 67.5 minutes] = 675 minutes total
Derex & Boyd (2015), *Closed Group*	Yes	Individual learning	45 minutes	N/A
		Social learning, Group size = 3	45 minutes	3 participants x 45 minutes = 135 minutes
		Social learning, Group size = 6, Full connectivity	45 minutes	6 participants x 45 minutes = 270 minutes
		Social learning, Group size = 6, Partial connectivity	45 minutes	3 subgroups of 2 participants x 45 minutes = 270 minutes
Derex & Boyd (2016), *Closed group*	No	Full connectivity	30 seconds (25 construction, 5 information) per trial x 72 trials = 36 minutes	Groups of 6 participants x 36 minutes = 216 minutes
		Partial connectivity		3 subgroups of 2 participants x 36 minutes = 216 minutes

Derex et al. (2015), *Closed group*	Yes	Individual learning (Non-social)	180s (construction period) + 30s (information period) = 210 seconds X 15 trials = 52.5 minutes total	N/A
		Social learning (other group members' arrowheads performances were visible and players could choose to see the shape of one of them)	180s (construction period) + 30s (information period) = 210 seconds X 15 trials = 52.5 minutes total	Groups of 4 participants x 52.5 minutes = 210 minutes
		Performance cue (only other group members' arrowheads performances were visible).	180s (construction period) + 30s (information period) = 210 seconds X 15 trials = 52.5 minutes total	Groups of 4 participants x 52.5 minutes = 210 minutes
Flynn 2008, *Chains*	Yes	No-model control condition	Either one success, refusal to continue or 4 minutes	
		Diffusion chain	2 demonstrations + 2 attempts = 4 'trials'	4 trials x groups of 6 = 24 trials
Hutchison et al 2017, *Chains*	No	Experiments 1, 2, 3	Training phase + test phase, unreported length	7 generations
Kempe & Mesoudi (2014), *Chains*	No	1 model	12 minutes	4 generations x 12 minutes 12 minutes per generation, 4 generations
				= 48 minutes
		3 models	12 minutes	4 generations x 3 participants x 12 minutes
				= 144 minutes
Kirby et al. (2008), *Chains*	No	Experiment 1	Training of 3 rounds x 2x14 items per round = 84 training trials	10 generations x 84 trials = 840 trials
		Experiment 2		

Kirby et al. (2015), *Chains + Closed group*	No	Chain	Training phase of 6 blocks x 12 shape-meaning associations = 72 trials	2 participants (pair) x 6 generations x 72 trials = 864 trials
		Closed group		
Martin et al. (2014), *Chains*	No	N/A	Training phase of 3 blocks x 13 aliens = 39 trials, (Excluded Test phase = 27 trials)	7 generations x 39 trials = 273 trials total
McGuigan et al (2017), *Seeded group*	Yes	Group (test) condition	1 hour x 4 days, 1 session per day = 4 hours	4 hours x 8 to 25 children in group = 32 to 100
		Asocial (Individual) controls	15 min on the first day, if successful, 2x 15 minutes more = 45 minutes max, total	N/A
		Level-4 controls	4 x 1 hour sessions = 4 hours	4 hours x 15 children in group = 60 hours
Mesoudi (2008), *Closed group*	Yes	Individual controls	3 seasons of 30 hunts = 90 hunts	N/A
		"Cultural learners" (social condition)	3 seasons of 30 hunts = 90 hunts	Groups of 5 to 6 participants x 90 hunts = 450 to 540 hunts (but participants can access social information only on some specific trials, so this is an over-estimation)
Mesoudi (2011), *Closed group*	Yes	Individual learners ("demonstrators")	3 seasons of 30 hunts = 90 hunts	N/A
		"Cultural learners"	3 seasons of 30 hunts = 90 hunts	Groups of 6 (1 learner + 5 demonstrators) participants x 90 hunts = 540 hunts
Morgan et al. 2015, *Chains*	No	Reverse Engineering	up to 20 minutes	5 generations x 20 minutes = 100 minutes
		Imitation/Emulation	5 minutes + up to 20 minutes = 25 minutes	5 generations x 25 minutes = 125 minutes
		Basic Teaching		

		Gestural Teaching		
		Verbal Teaching		

Muthukrishna et al. (2014), *Chains*	No	1 model	Experiment 1: 25 minutes (recreating the target image	25 minutes x 1 model x 10 generations = 250 minutes
		5 models		25 minutes x 5 models x 10 generations = 1250 minutes
		1 model	Experiment 2: 50 minutes to recreate the knot	50 minutes x 1 model x 10 generations = 500 minutes
		5 models		50 minutes x 5 models x 10 generations = 2500 minutes
Sasaki & Biro (2017), *Replacement*	Yes	Solo (Individual control)	60 trials	N/A
		Experimental / Test	24 trials per pigeon (12 trials per generation, and replacement meaning that each pigeon 'lasts' for 2 generations, expect first and last pigeon, 12 trials each instead)	108 trials experimental with 12 trials per generation/5 generations x 2 individuals, except the first generation (individual on its own)
		Pair	60 trials	60 trials x 2 individuals = 120 trials
Tan & Fay 2011, *Chains*	No	No interaction	5 minutes to read the text (1st participant) or listening once to a recording	5 minutes reading (from the first participant) + 1 listening to recording x 3 later participants in the chain
		Interactive	5 minutes to read the text (1st participant) or listening to the previous participant retell the story	5 minutes to read the text (1st participant) + listening to the previous participant retell the story x 3 participants in the chain
Tennie et al. (2014), *Chains*	No	No demonstration	60 seconds trial	5 minutes (no demonstration to G1; 5 60s trials serving as demonstrations to next generations)

		Demonstration		6 minutes (60s demonstration to G1 + 5 60s trials serving as demonstrations to next generations)
Vale et al. 2017, *Seeded Group*	Yes	Seeded-All phases	Phase 1: 5x 2hours sessions + Phase 2: 10hours of video demonstrations + Phase 3: 10 x 1hour sessions (or depletion) = maximum of 30 hours	maximum of 30 hours x number of chimpanzees per group (not directly reported)
		Non-seeded	Phase 1: 5x 2hours sessions (phase 2: 10hours of video of a conspecific close to the juice container, but not interacting with it) + Phase 3: 10 x 1hour sessions (or depletion) = maximum of 30 hours	
		Phase-3 only Controls	30 hours of exposure (collectively)	30 hours of exposure (collectively) x number of chimpanzees in the group (unreported)
		Asocial Phases 1 & 3	Separate exposure (i.e., not in group) for 2x30 minutes to phase 1+ 2x 30 minutes to phase 3 = 2 hours	N/A
Wasielewski (2014), *Replacement*	Yes	Asocial condition	15 minutes building, what varies between condition is what they can observe while building	N/A
		End-product condition		10 generations x 15 minutes = 150 minutes
		Action condition		
		End-product and action condition		
Zwirner & Thornton (2015), *Chains*	Yes	Asocial (Individual control)	6 trials, 5 minutes each = 30 minutes	N/A

		Emulation	5 minutes trial (time during which production was on display or they were teaching isn't included)	6 participants x 5 minutes = 30 minutes
		Teaching		
		Imitation	5 minutes observing + 5 minutes trial = 10 minutes, except participant 1 (5 minutes total, no observation)	6 participants, 5 with 10 minutes, 1 with 5 minutes only = 55 minutes

3. Effective and opportunity learning times

As mentioned above, the learning times computed here were calculated on the basis of the description of the methods for each paper and they represent the upper limit of the time spent by participants in learning the task. They serve as a *theoretical* upper limit because they do not measure the actual time the participants in fact invested in learning the task. For instance, in seeded groups, the participants are not always actively trying to solve the task. Moreover, the group typically shares the task such that not all individuals can try to solve it at the same time.

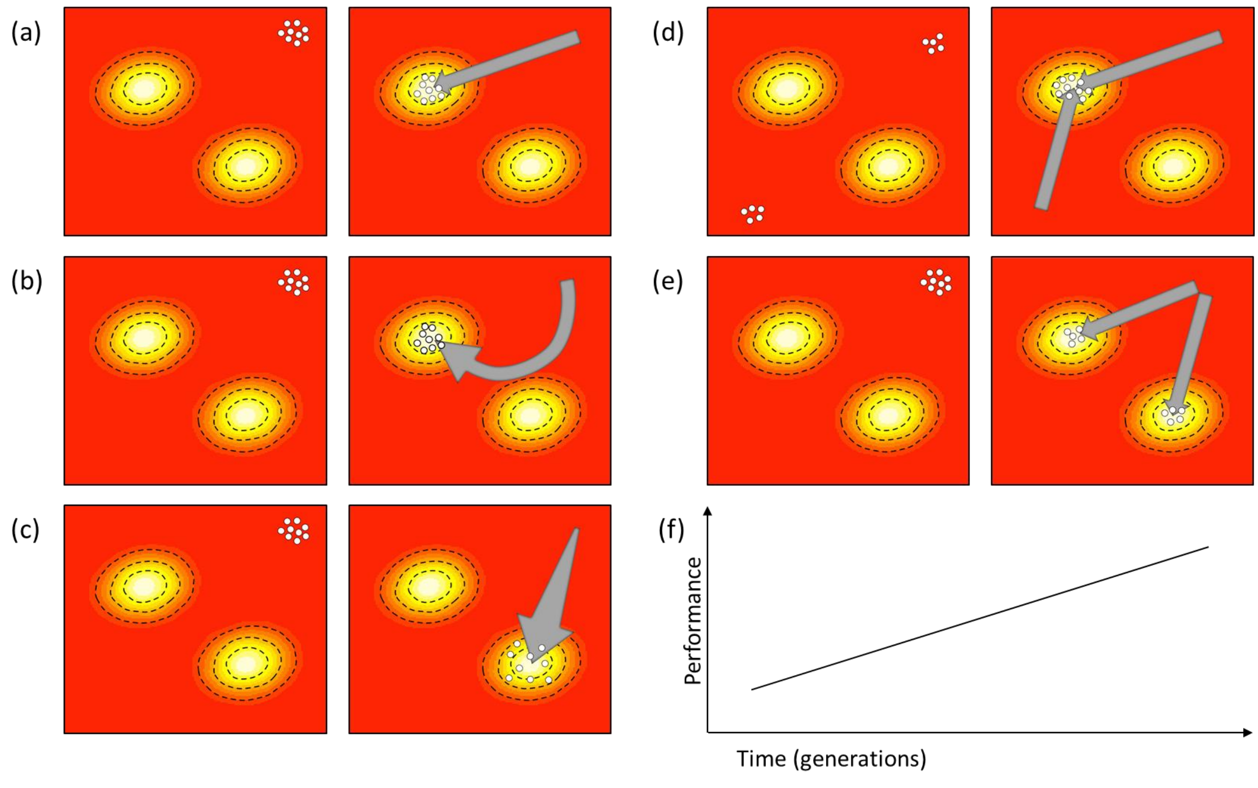
We can thus distinguish between *opportunity time* and *effective time*. Opportunity time is the total amount of time offered to each individual participant for learning and solving some task. The time computed for each experiment in the table above is thus opportunity time. In non-social conditions, opportunity time is equal to the total learning time of the individual. In social conditions, opportunity time refers to how much time each participant in the tradition has to solve the task, i.e., it is the total additive learning time of the tradition divided by the number of participants. In contrast, effective time refers to the total amount of the opportunity time that a participant *effectively* invests into the learning and solving of the task. There are several options available to measure the effective time invested by a participant into a task. For instance, effective time can be measured by the total time the participant is actively handling some object related to the task, such as the time using a computer mouse or the time spent manipulating the experimental apparatus (Bonawitz et al. 2011). The ratio between effective and opportunity time of a given participant is indicative of her motivation: a highly motivated participant will use most of her opportunity time (high effective/opportunity time ratio); a less motivated participant will make a less effective use of her opportunity time (lower effective/opportunity time ratio). Similar measures of motivation may be used for children and non-humans, even if we should expect cross-species and developmental differences in motivation (Dean et al. 2014).

Differences in motivation may offer an alternative explanation of why participants in non-social conditions fail to reach the same degree of performance improvement as those in the social condition. Indeed, if we follow the recommendations made in the previous section, the participants in the social condition will be individually required to invest less time in the experimental task than those in the non-social condition, and they will generally only have to solve the task once, whereas the participant in the non-social condition will have to repeatedly do so. Therefore, participants in the non-social condition may simply fail to reach the same degree of performance than the traditions not because they are unable to invent the solutions on their own, but rather because they get demotivated by the duration or repetitiveness of the experimental task, and therefore either slowing down their rate of improvement or prematurely ceasing to improve upon their solutions, or both. Thus differences in participants' motivation may in turn explain the difference in performances between the social and non-social conditions.

Bonawitz E., Shafto P., Gweon H., Goodman N.D., Spelke E., Schulz L. 2011 The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition* 120, 322-330

Dean L.G., Vale G.L., Laland K.N., Flynn E., Kendal R.L. 2014 Human cumulative culture: a comparative perspective. *Biological Reviews* 89, 284-301.

ESM 3 - Design spaces and evolutionary trajectories



An illustration of 5 different evolutionary trajectories of experimental traditions set in a visual depiction of design spaces with dimensions representing some arbitrary varying properties of solutions to a task. For each trajectory (a-e), the hypothetical populations are represented by the white dots, with the left box representing the distribution of the population at the beginning of the experiment, and the right box representing the distribution of the population at the end of the experiment, together with arrows representing the trajectory of the traditions during the experiments. The colour distribution represents the degree of performance of the different solutions in the design space, increasing from red to yellow, with two sets of solutions (yellow peaks) of equal, high, performance. (a) An example of a trajectory of cumulative culture. (b) Same end-result as in the previous case, but with a different evolutionary trajectory. (c) Straightforward trajectory, but with an increase in the dispersion (variance/spread) of the population through time. (d) Convergence of two populations to a same solution. (e) Divergence of a population on two solutions of equal performance. (f) All these scenarios (a-e) would qualify as exhibiting cumulative effects. Although they exhibit the exact same mean linear increase of performance (or complexity) through time, they follow very different evolutionary trajectories.

Appendix B – Supplementary materials for When Iconicity Stands in the way of Abbreviation

Contents

1. Guide to the OSF materials

The data files (dataClem.csv, and dataRenesse.csv), metadata (metadata_Zipfpaper.rtf**)** and R script (ZipfPaperScript.R) necessary to replicate all the findings presented in the manuscript are available here: https://osf.io/ykp37/

We also pre-registered all the work that went into this project on the Open Science Framework's website, where a complete, public record of the research process can be accessed. We release all the registrations and reports relevant to this study**.** The successive reports allow a complete picture of the research process involved in producing the study and results reported in the manuscript. All documents can be freely consulted on the Open Science Framework at the given URLs. They include:

- **zipf1-2016-08-03**
(https://mfr.osf.io/render?url=https://osf.io/2fbns/?action=download%26mode=render**)**: First pre-registration of this study[10]. It presents the rationale of the study and the study plan, including possible follow-ups.

- **zipf2-2016-10-06**
(https://mfr.osf.io/render?url=https://osf.io/tq97x/?action=download%26mode=render): This document explains how we built our two corpus and measured the motifs' visual complexity, with a description of our exclusion criteria (both pre- and post-measurements) and additional methodological details on the measurement of complexity, including the editing of motif pictures.[11]

- **zipf3-2017-10-23**
(https://mfr.osf.io/render?url=https://osf.io/42ser/?action=download%26mode=render**):** This file reports a few departures from the initial registered study plan (due to characteristics of the materials used), additional exclusions, a short erratum correcting a paragraph from a previous registration, and the correlation between the two visual complexity measures that were used. It also registers the next version of the study[12], which is the one we report in this manuscript.

- **zipf4-2018-04**
(https://mfr.osf.io/render?url=https://osf.io/jms8z/?action=download%26mode=render**)**: This report includes results for the following: (1) whether frequency distributions (in both the Renesse and the Clemmensen datasets) followed scale-free distributions, (2), whether the motifs' frequency was negatively correlated to visual complexity (i.e., what would have been predicted by a law of abbreviation), and (3) whether this correlation applied equally to both iconic and non-iconic motifs. It also states which files those results are based on. Finally, this report also stands as the registration for additional exploratory analyses that the results called for (mainly pertaining

10 Timestamped version can be found here : https://osf.io/ct28b
11 Timestamped registration can be found here: https://osf.io/jwrsv
12 Timestamped registration can be found here: https://osf.io/b3pdt

to the differences between the Renesse and Clemmensen datasets). Be aware that some of the results reported there were actually incorrect due to a small mistake in the R code (use of the wrong 'Frequency' variable). The correct results are the ones included in the manuscript.

- zipf5-2018-05-22(https://mfr.osf.io/render?url=https://osf.io/672qz/?action=download%26mode=render**):** Our latest report includes results for the following additional exploratory analyses: (1) whether there is a significant change in the proportion of iconic to non-iconic motifs between the two datasets, (2) whether complex motifs became more frequent and simpler motifs less frequent, independently of their iconicity, and finally (3) whether differences between Renesse & Clemmensen depend on the new motifs appearing in Renesse but not in Clemmensen.

2. Basic heraldry vocabulary

Term	**Definition**
Arms	Equivalent of coat of arms or shield.
Armorial	A collection of branches (part of families) names, with descriptions of their arms ("blasons").
Motif	Images that can appear on arms.
Charge	Iconic motif, can be placed anywhere on the shield, roughly corresponds to "meuble" in heraldic French.
Ordinary	Abstract motif. Includes "pièces", whose placement is constrained by rules, "partitions", which are divisions of the shield, and "rebattements", geometric patterns covering all or part of the arms.
Tincture	Ways in which any part of the arms can be colored, divided between colors, metals and furs. Furs (Hermine, Vair) are patterns, whereas colors and metals are plain colors.

3. Additional methodological details

We here present additional methodological details about data collection. They list all methodological decisions that were made related to processing our original (historical) materials.

3.1. Inventory constitution

A first list of motifs was constituted from the general index of Renesse's seven-volume compilation. The index was text-captured and rearranged to yield a list of motifs. From this list of motifs, we made a series of deletions, for the following reasons:

- Removing duplicates (i.e., motifs that Renesse seems to have mentioned twice by mistake);
- All "extraordinaires" and "divers/diverses" items were removed, as they represent miscellaneous, quaint, or heterogeneous motifs not clearly standing for a well-formed category;
- The tinctures (red, blue, yellow, etc. — corresponding to gules, azur, or, etc. in heraldic parlance) were removed;
- Categories which were referred to interchangeably were merged into one;
- Chapter headings, which referred to the general theme of the chapter's motifs, rather than to motifs themselves, were removed;
- Motifs that were not present alone on at least one otherwise empty shield (i.e., never present "in isolation"), were removed;
- Motifs that occurred in isolation, but only in groups of two or more repetitions, were removed;
- Geometric motifs occurring in groups of varying size were removed;
- Motifs that only occurred accompanied by other, different motifs, were removed.

Additionally, a few items were present inside Renesse's volumes but not in the index were added to the list of motifs. Particular additions or deletions were reported in the Open Science Framework reports at each stage.

A second step consisted in cross-checking the list obtained with Renesse's sections entitled "armes complètes". For each motif listed in his index, Renesse, when possible, devotes a section to an exhaustive list of all the arms bearing the motif in question and nothing else. At this stage, items were classified into two groups: the charges ("meubles", referred to as "iconic" motifs in our paper) on one side, the "pièces, partitions, rebattements, émaux" on the other (referred to as "non-iconic" motifs in our paper).

A third and last step required us to discard all the motifs that were represented by two arms or less, as the complexity measures were taken as the average of three shields. This happened either when the motif was featured on less than thee shields in total, or when the motif didn't occur on its own on three shields (i.e., occurred only in combinations).

3.2. The iconic / non-iconic distinction in heraldry and in our data

Renesse's inventory, following a long-established taxonomy, makes a sharp distinction between certain categories: "charges", which are any image that can be placed anywhere on the arms, on one hand, and "pièces", whose placement is constrained by rules, or "partitions", which are divisions of the arms. The English term "ordinaries" covers both "pièces" and "partitions" (Fox-Davies 1900). Ordinaries are abstract, geometric shapes that do not represent a natural object in any detail: saltires, bends, lozenges, etc. The subset of motifs they represent is referred to as non-iconic. By contrast, charges are essentially figurative motifs, representing mainly animals, plants and various artefacts and the subset they represent are referred to as iconic.

3.3. Shields selection

Three shields were collected among the Rolland's compendium of illustrations, for each motif. They were selected in the following way: (1) the shield bore the motif of interest and nothing else, and (2) shields where the motif is tinctured with white were given priority, unless the motif was tinctured with black 90% of the time or more. When there weren't enough white motifs, the next preferred tincture was yellow. This constraint made it easy for us to process the pictures: motifs were edited to remove the tincture markings added by the illustrators to their designs.

These markings were added by the Rollands to stand for tinctures, i.e., colours. White was coded with no marking, while yellow had the simplest of markings (dots). Finally, due to the structure of the source, choosing based the alphabetical order of the owners' name would have introduced a confound. Hence, the shields, as long as they satisfied the two previous criteria, were picked randomly.

3.4. Pictures' preparation

All pictures underwent a very light resizing to match a fixed 309 x 400 pixels template, which was able to accommodate all the shields that we selected. This resizing also made the various shields more easily comparable and compensated for irregularities in the original material. All shields also got their border erased, so that complexity measures could bear on the motifs themselves, and would not bear any additional noise due to irregularities in the borders' printing. After editing, pictures were saved as .pnm files, and then had the Potrace algorithm applied to them (Selinger, 2003).

3.5. Frequency measures in the Clemmensen corpus

Our measures of frequency were based on Steen Clemmensen's extensive database "armorial.dk" (version 12). We counted only the arms where the focal motif occurs in isolation: in one exemplar, and not accompanied by anything else.

Within this database, our measures were based on the table of branches, i.e., versions of a coat of arms possessed by a family, and presents a list of arms characterized by their design and the family or sub-lineage that carried them. This table includes 31 691 branches over 20 606 distinct families. Branches differ from families because a given family often came to possess various arms in the course of its history, either because the family branched out into distinct lineages, or because it decided to change the design of its arms.

Starting from the 31 691 branches listed by Clemmensen, were removed (1) all entries referring to mythical or heroic characters (corresponding to "MarchedArmes = _HERO"), and (2) all entries with no available arms (30 928 entries left).

Manual searches were then carried on both the English and the French versions of the list of arms, following criteria mimicking Renesse's classification as closely as possible. To ensure that our counts included only arms featuring the motif in isolation, the following were discarded:
- Arms mentioning two motifs (or more), or repetitions of one motif;
- Arms mentioning any kind of accessory, prop or support not part of the original motif. This includes (1) a mount, a hill or a terrace when a motif is poised on it, (2) any non-standard decoration accompanying a charge if that decoration isn't part of the charge by definition (e.g., a flowery bend), (3) any uncommon or unexpected object held or contained by an entity (e.g., a basket holding stars or a beaver eating a duck), (4) any pattern or motif superimposed on another: e.g., a checky lion.
- Any exceptional variant requiring a special mention in the database.

On the other hand, arms were retained when:
- They specified the motif's position on the shield: "per bend", "per fess", "in chief", etc. (This did not apply to partitions or charges whose positions on the shield is fixed by definition).
- Their orientation was inverted.
- They specified an animal's posture: *rampant, courant, passant*, etc., even when that indication was not part of the motif's definition, as long as such modificatory were not issant or naissant (these terms indicate that only one half of an animal is shown). The adjectives *issant* and *naissant* were treated as introducing a different kind of motif, different from the whole animal, as Renesse usually treats such motifs.

When in doubt, we referred to Renesse's inventory of heraldic motifs, attempting to stay as close as possible to his classifications. All counts were made for each motif in English and in French to ensure robustness. In case of conflict, the French version was systematically preferred, since our reference classification (Renesse's) is in French.

3.6. Frequency measures in the Renesse corpus

In Renesse, the only arms to be counted exhaustively are the ones present in the "Armes complètes" section, meaning they are coats of arms bearing the relevant motif, and only this motif. Such arms were counted manually by a research assistant, systematically going through the volumes of Renesse's inventory. Minor variations, signaled between parentheses in Renesse were counted in the focus motif's occurrences, whereas more major variations, signaled by being mentioned in different paragraphs (or sub-paragraphs).

4. Results of analyses on the datasets without applying the exclusion criterion

Test	With exclusions (reported in the main text)	Without exclusions
Correlation between perimetric and descriptive complexity	r_τ = .69, p < .001 95%CI = [0.657, 0.725]	r_τ = .70, p < .001 95%CI = [0.665, 0.729]
Difference in perimetric complexity between iconic and non iconic motifs	U = 32160, p < .01	U = 34052, p < .01
Difference in descriptive complexity between iconic and non-iconic motifs	U = 36110, p < .01	U = 38110, p < .01
Overall correlation between perimetric complexity and frequency - Clemmensen corpus	r_τ= -.02, p = .633 95%CI = [-0.103, 0.065]	r_τ= -.03, p = .401 95%CI = [-0.116,0.049]
Overall correlation between descriptive complexity and frequency - Clemmensen corpus	r_τ= -.09, p = .018 95%CI = [-0.177, -0.011]	r_τ= -.11, p = .007 95%CI = [-0.189,-0.025]
Overall correlation between perimetric complexity and frequency - Renesse corpus	r_τ = .12, p < .001 95%CI = [0.058, 0.186]	r_τ= .12, < .001 95%CI = [0.062,0.186]
Overall correlation between descriptive complexity and frequency - Renesse corpus	r_τ = .08, p = .008 95%CI=[0.02, 0.149]	r_τ= .09, p = .004 95%CI = [0.026,0.153]
Correlation between perimetric complexity and frequency for non-iconic motifs only - Clemmensen corpus	r_τ= -.08, p = .233 95%CI = [-0.211, 0.058]	r_τ= -.07, p = .273 95%CI = [-0.204,0.063]
Correlation between descriptive complexity and frequency for non-iconic motifs only - Clemmensen corpus	r_τ= -.14, p = .035 95%CI = [-0.27, -0.001]	r_τ= -.13, p = .045 95%CI = [-0.262,0.005]

Correlation between perimetric complexity and frequency for non-iconic motifs only - Renesse corpus	r_τ = -.03, p = .534, 95%CI = [-0.147, 0.078]	r_τ= -.04, p = .504 95%CI = [-0.148,0.075]
Correlation between descriptive complexity and frequency for non-iconic motifs only - Renesse corpus	r_τ = -.10, p = .07 95%CI = [-0.211, 0.011]	r_τ= -.10, p = .066 95%CI = [-0.212,0.009]
Correlation between perimetric complexity and frequency for iconic motifs only - Clemmensen corpus	r_τ= .12, p < .05 95%CI = [0.021, 0.226]	r_τ= .09, p = .067 95%CI = [-0.01, 0.194]
Correlation between descriptive complexity and frequency for iconic motifs only - Clemmensen corpus	(r_τ= .08, p = .14 95%CI = [-0.028, 0.18]	r_τ= .05, p = .359 95%CI = [-0.057,0.149]
Correlation between perimetric complexity and frequency for iconic motifs only - Renesse corpus	r_τ = .22, p < .001 95%CI = [0.143, 0.291]	r_τ= .21, p < .001 95%CI = [0.14,0.281]
Correlation between descriptive complexity and frequency for iconic motifs only - Renesse corpus	r_τ = .18, p < .001 95%CI = [0.108, 0.253]	r_τ= .18, p < .001 95%CI = [0.111,0.253]
Change in frequency for iconic versus non-iconic motifs	U = 16004, p < .001	U = 16962, p < .001
Correlation between change in frequency and perimetric complexity	r_τ = .16, p < .001 95%CI = [0.076, 0.242]	r_τ= .16, p < .001 95%CI = [0.08,0.242]
Correlation between change in frequency and descriptive complexity	r_τ = .22, p < .001 95%CI = [0.144, 0.297]	r_τ= .22, p < .001 95%CI = [0.149,0.298]
Correlation between change in frequency and perimetric complexity - Iconic motifs only	r_τ = .11, p = .034 95%CI = [0.005, 0.206]	r_τ= .11, p = .027 95%CI = [0.012,0.205]
Correlation between change in frequency and descriptive complexity - Iconic motifs only	r_τ = .08, p = .108 95%CI = [-0.011, 0.172]	r_τ= .09, p = .067 95%CI = [0,0.179]
Correlation between change in frequency and perimetric complexity - Non-iconic motifs	r_τ = .03, p = .590 95%CI = [-0.105, 0.173]	r_τ= .02, p = .710 95%CI = [-0.115,0.162]

Correlation between change in frequency and descriptive complexity - Non-iconic motifs	r_τ = .09, p = .161 95%CI = [-0.049, 0.227]	r_τ= .08, p = .226 95%CI = [-0.061,0.214]

Appendix C - For Chapter 3

Exclusions

Exclusions from criteria at the inventory level

Aran	Nastaliq (typographic) variant of Arab
Bagam	Not completely deciphered
Balti B	Insufficient documentation (based on only one manuscript)
Blis	Not a visual writing system
Bopo	Phonetic alphabet not used independently from other scripts
Brai	Not a visual writing system
Cirt	Fictional writing system.
Cpmn	Undeciphered, and not in Unicode (only a very preliminary proposal)
Dupl	Stenography
Egyd	Not in Unicode, and without a valid proposal
Egyh	Unified with Egyp
Hanb	Han with Bopomofo (alias for Han + Bopomofo) - Han covered with Hani, and Bopomofo excluded for other reasons
Hans	Subset of (already comprised in) Hani
Hant	Subset of (already comprised in) Hani
Hatr	Variant of Aramaic
Hira	Included in Hrkt
Inds	Undeciphered
Jamo	Subset of Hangul. Fused with Hang
Jpan	Super-category including Kana and Kanji
Jurc	Only partially deciphered
Kana	Included in Hrkt
Kitl	Not fully deciphered
Kits	Not fully deciphered
Kore	Super-category including Hangul and Chinese characters
Latf	Typeface of Latin
Latg	Typeface of Latin
LinA	Undeciphered
Maya	Not fully deciphered (only "substantially")
Moon	Not a static visual writing system
Nkgb	Phonetic alphabet
Nkdb	Shamanic pictographs
Piqd	Fictional writing system.
Roro	Undeciphered
Sara	Fictional writing system.
Sgnw	Not a static visual writing system
Syre	Estrangelo Syriac - typographic variant. Fused with Syrc
Syrj	Western Syriac - typographic variant. Fused with Syrc
Syrn	Eastern Syriac - typographic variant. Fused with Syrc
Teng	Fictional writing system.
Visp	Not a static visual writing system

Diwani Siyaq	Numbers only
Persian Siyaq	Numbers only
Indic Siyaq	Numbers only
Ottoman Siyaq	Numbers only

Exclusions from problems encountered during data collection

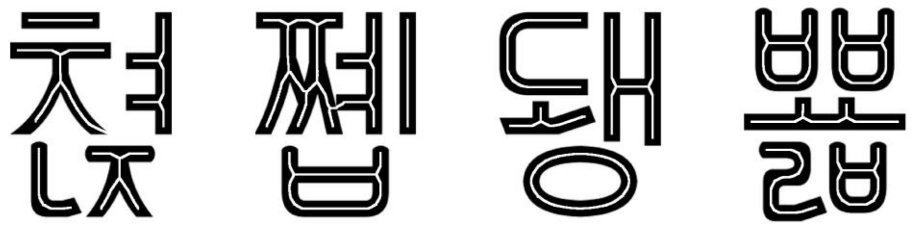
Balti A	No available font
Dhives Akuru	No available font
Dogr	No available font
Elymaic	No available font
Eskaya	No available font
Garay	No available font
Geok	No available font
Gong	No available font
Gonm	No available font
Kawi	No available font
Khema (Gurung)	No available font
Kirat Rai	No available font
Kpel	Too many symbols missing
Leke	No available font
Loma	Couldn't be remapped
Maka	No available font
Medf	No available font
Mwangwego	No available font
Nandinagari	No available font
Nshu	No available font
Pau Cin Hau syllabary	No available font
Phlv	No available font
Pyu	No available font
Shui	No available font
Sogd	No available font
Sogo	No available font
Tangsa (Mossang)	No available font
Tangsa (Khimhun)	No available font
Wcho	No available font
Wole	No available font
Xsux	Problems with both available fonts (missing symbols, scrambled output)

Examples of the Treatment Applied to characters' pictures

In black is the outline of the (resized) original character, with the result from our process in white superimposed on top of it. Characters were randomly chosen from all characters from each script.

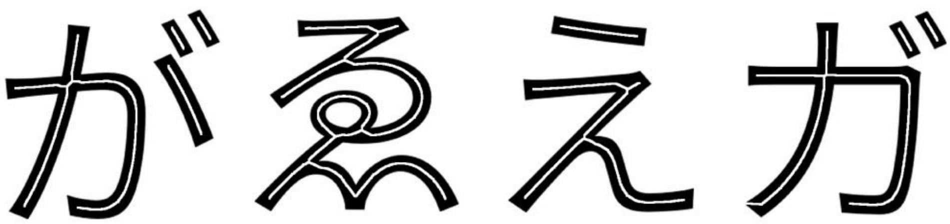
East Asian

HANG {0CCD5.pnm, 0CAB1.pnm, 0B411.pnm, 0BFBB.pnm}



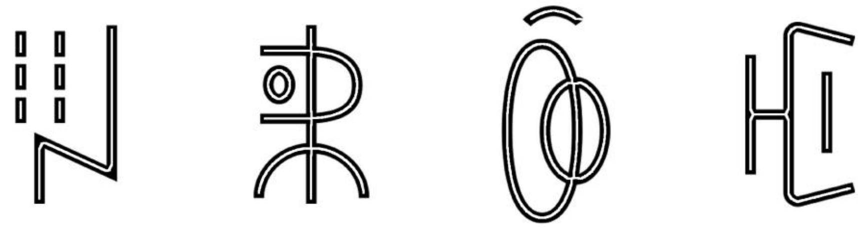
HANI {07CF6.pnm, 09E4A.pnm, 0616C.pnm, 061D7.pnm}

HRKT {0304C.pnm, 03091.pnm, 03048.pnm, 030AC.pnm}



TANG {17554.pnm, 1727F.pnm, 1724E.pnm, 17552.pnm}

YIII {0A0AB.pnm, 0A278.pnm, 0A3CE.pnm, 0A1FF.pnm}



Indian

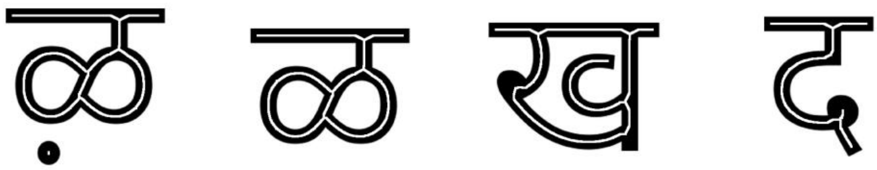
AHOM {11705.pnm, 11713.pnm, 11705.pnm, 11719.pnm}

BENG {009A2.pnm, 00998.pnm, 00995.pnm, 00989.pnm}

BHKS {11C08.pnm, 11C04.pnm, 11C23.pnm, 11C2C.pnm}

BRAH {11031.pnm, 1100A.pnm, 1102D.pnm, 11028.pnm}

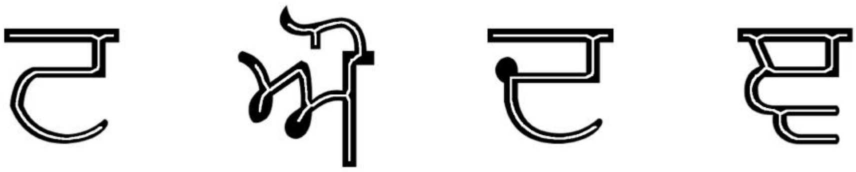
DEVA {00934.pnm, 00933.pnm, 00916.pnm, 00926.pnm}



GRAN {11328.pnm, 11314.pnm, 11321.pnm, 11310.pnm}

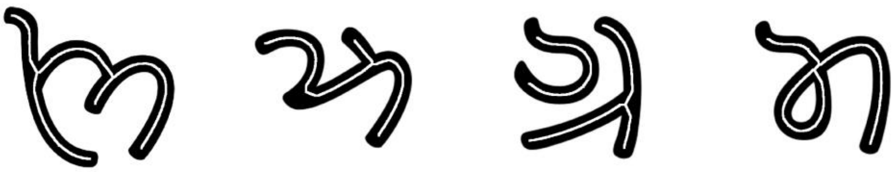
GUJR {00AA0.pnm, 00A9F.pnm, 00A96.pnm, 00A93.pnm}

GURU {00A1F.pnm, 00A14.pnm, 00A26.pnm, 00A1E.pnm}

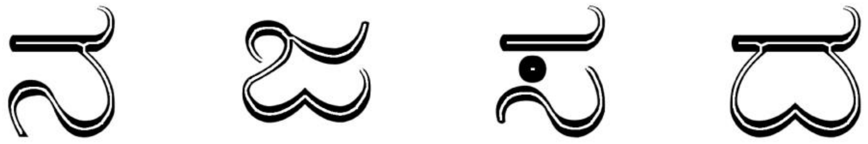


KHAR {10A00.pnm, 10A2C.pnm, 10A1D.pnm, 10A12.pnm}

KHOJ {11218.pnm, 11211.pnm, 1120B.pnm, 11224.pnm}



KNDA {00CA8.pnm, 00C9C.pnm, 00CB8.pnm, 00CA6.pnm}

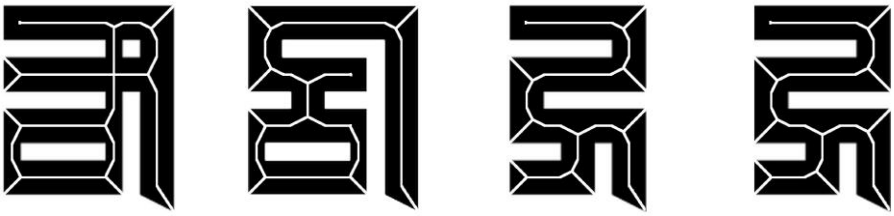


KTHI {110A6.pnm, 1109F.pnm, 11087.pnm, 110AC.pnm}

LIMB {01915.pnm, 01915.pnm, 01907.pnm, 0190C.pnm}

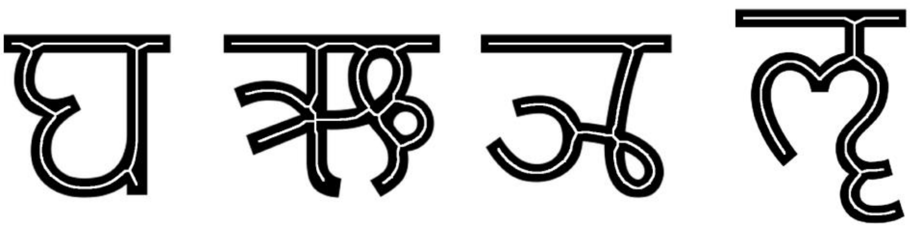
MAHJ {11169.pnm, 11152.pnm, 1115A.pnm, 11156.pnm}

MARC {11C85.pnm, 11C80.pnm, 11C89.pnm, 11C89.pnm}



MLYM {00D0E.pnm, 00D12.pnm, 00D37.pnm, 00D2B.pnm}

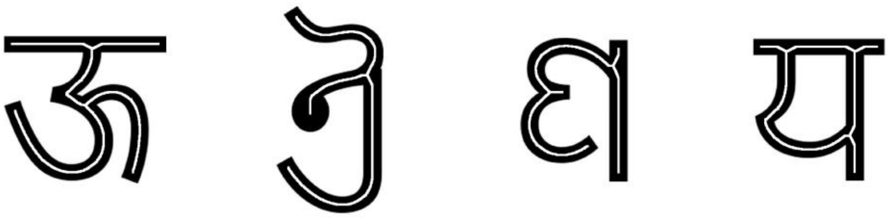
MODI {11624.pnm, 11607.pnm, 11617.pnm, 11609.pnm}



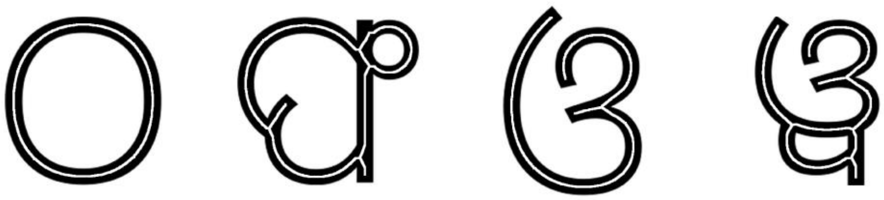
MTEI {0AAE7.pnm, 0ABCB.pnm, 0ABC1.pnm, 0ABE0.pnm}

MULT {11297.pnm, 11280.pnm, 1128A.pnm, 11285.pnm}

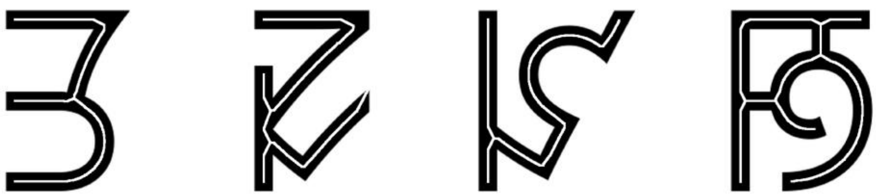
NEWA {11405.pnm, 1140B.pnm, 1141E.pnm, 1142B.pnm}



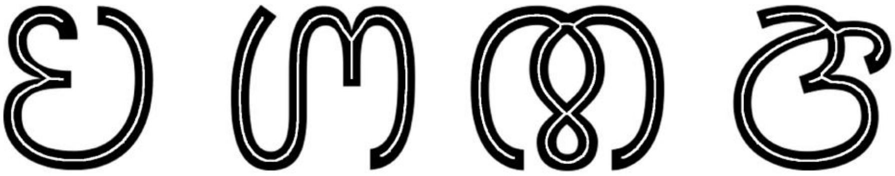
ORYA {00B20.pnm, 00B2B.pnm, 00B13.pnm, 00B71.pnm}



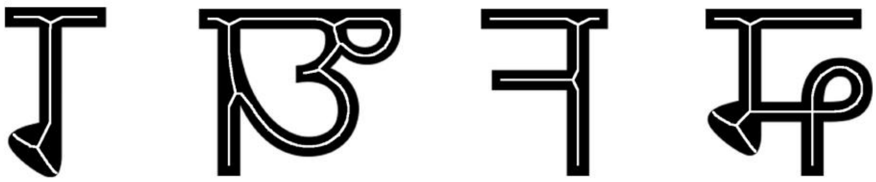
PHAG {0A857.pnm, 0A85B.pnm, 0A85C.pnm, 0A850.pnm}



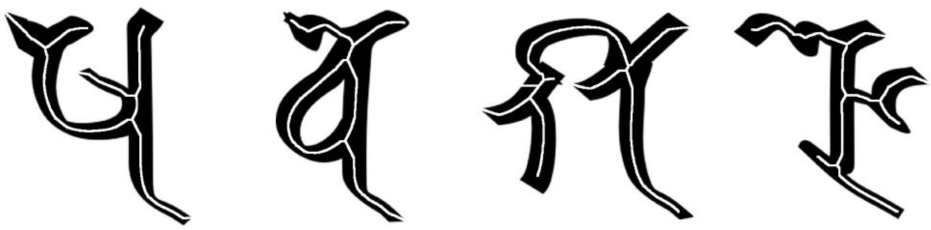
SAUR {0A8A4.pnm, 0A8A9.pnm, 0A89A.pnm, 0A886.pnm}



SHRD {111AB.pnm, 11190.pnm, 111A4.pnm, 1119E.pnm}



SIDD {115AC.pnm, 115A4.pnm, 115AB.pnm, 11597.pnm}

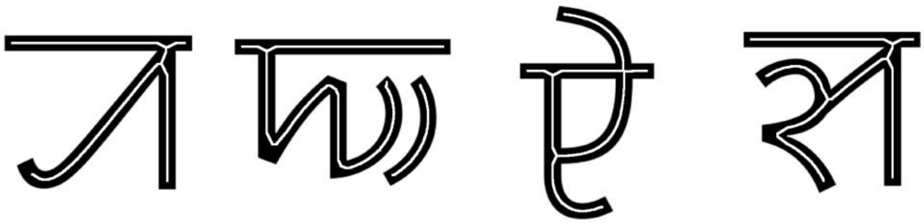


SIND {112C1.pnm, 112B0.pnm, 112D0.pnm, 112D8.pnm}

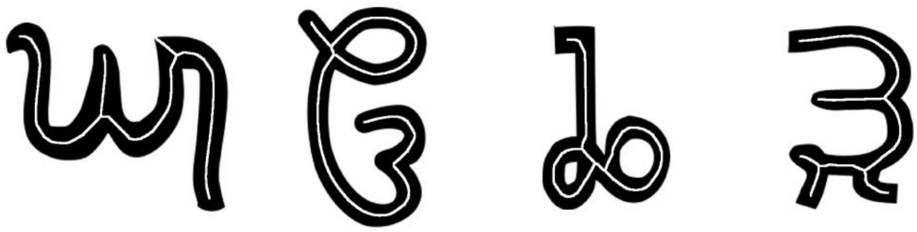
SINH {00DBB.pnm, 00D92.pnm, 00DB4.pnm, 00DC1.pnm}

SOYO {11A71.pnm, 11A62.pnm, 11A60.pnm, 11A6D.pnm}

SYLO {0A800.pnm, 0A80F.pnm, 0A804.pnm, 0A821.pnm}

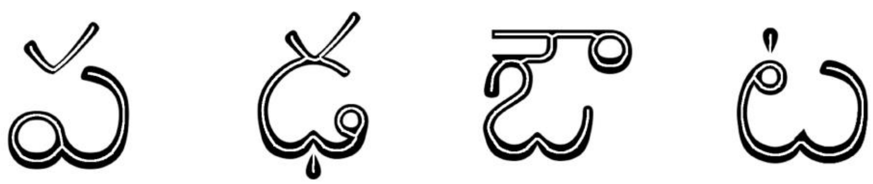


TAKR {1168D.pnm, 11685.pnm, 11697.pnm, 116AA.pnm}



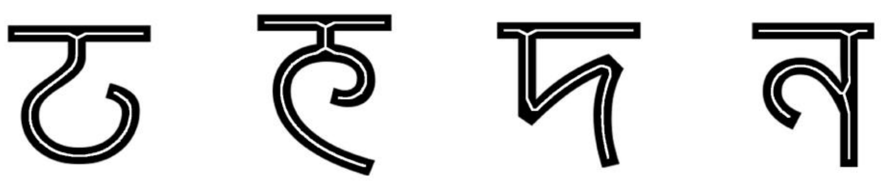
TAML {00BB3.pnm, 00B89.pnm, 00B85.pnm, 00BB6.pnm}

TELU {00C2A.pnm, 00C22.pnm, 00C14.pnm, 00C1F.pnm}

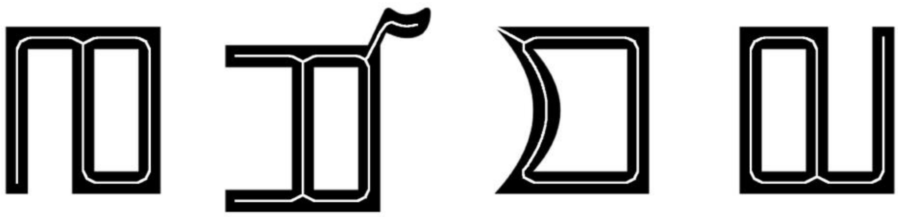


TIBT {00F5F.pnm, 00F68.pnm, 00F56.pnm, 00F4E.pnm}

TIRH {1149C.pnm, 114AF.pnm, 114A0.pnm, 114AA.pnm}

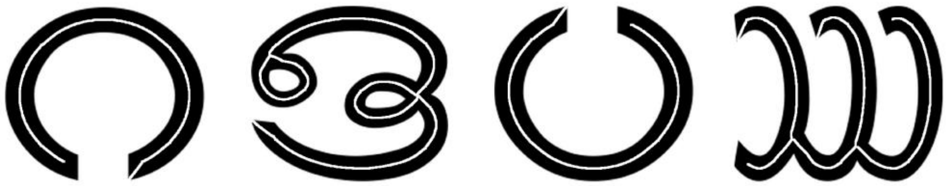


ZANB {11A0C.pnm, 11A23.pnm, 11A21.pnm, 11A22.pnm}



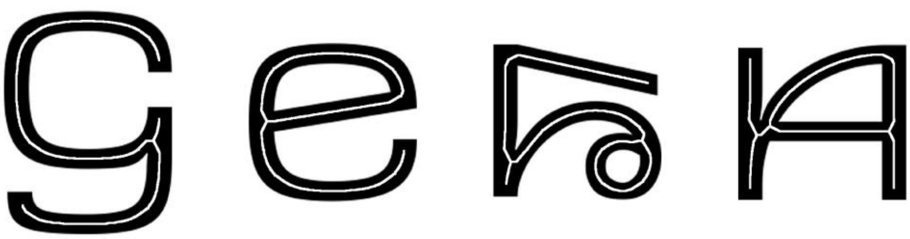
Mainland South East Asia

CAKM {11109.pnm, 1110E.pnm, 1111B.pnm, 1110A.pnm}

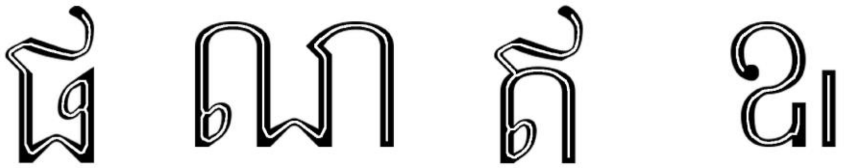


CHAM {0AA14.pnm, 0AA1A.pnm, 0AA12.pnm, 0AA1A.pnm}

KALI {0A91E.pnm, 0A915.pnm, 0A912.pnm, 0A90C.pnm}

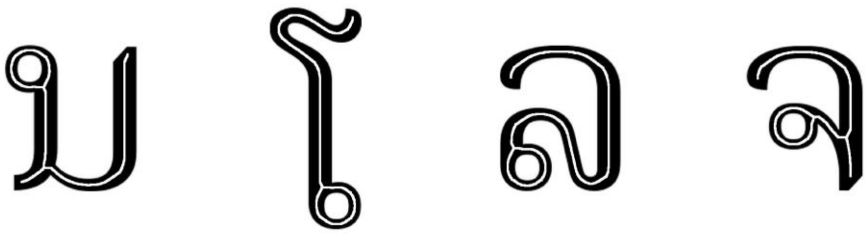


KHMR {01795.pnm, 0178E.pnm, 017A5.pnm, 017A9.pnm}



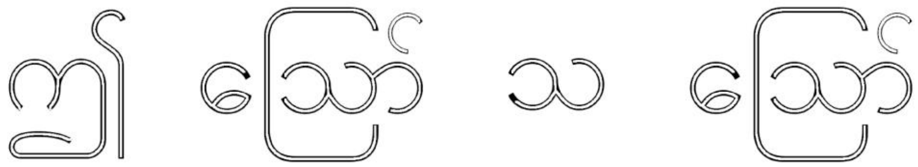
LANA {01A37.pnm, 01A38.pnm, 01A22.pnm, 01A32.pnm}

LAOO {00E99.pnm, 00EC2.pnm, 00EA5.pnm, 00E88.pnm}



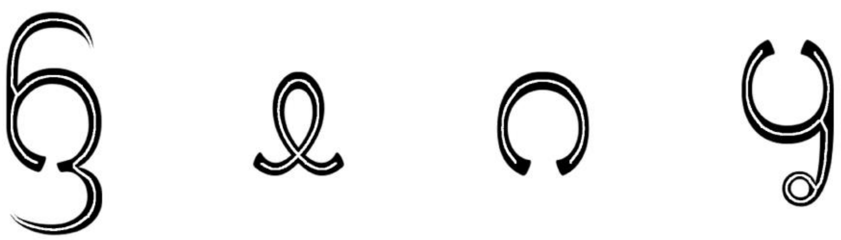
LEPC {01C11.pnm, 01C21.pnm, 01C0C.pnm, 01C1F.pnm}

MYMR {01024.pnm, 0102A.pnm, 0101E.pnm, 0102A.pnm}

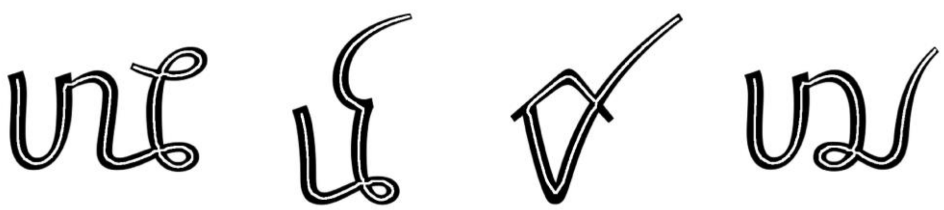


TALE {01963.pnm, 01963.pnm, 01960.pnm, 01961.pnm}

TALU {0198C.pnm, 019B8.pnm, 01985.pnm, 019C7.pnm}



TAVT {0AA88.pnm, 0AA8C.pnm, 0AAA4.pnm, 0AAA2.pnm}



THAI {00E12.pnm, 00E0D.pnm, 00E40.pnm, 00E03.pnm}

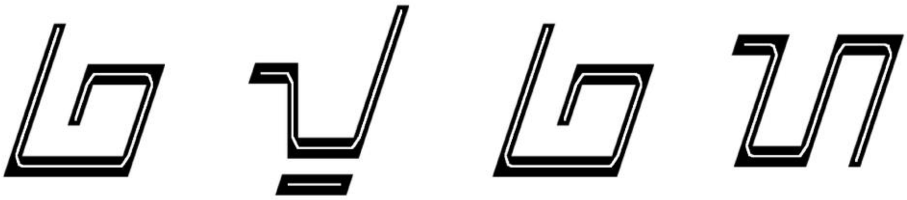
Insular South East Asia

BALI {01B2F.pnm, 01B27.pnm, 01B29.pnm, 01B16.pnm}

BATK {01BD4.pnm, 01BC0.pnm, 01BC8.pnm, 01BE1.pnm}

BUGI {01A05.pnm, 01A00.pnm, 01A04.pnm, 01A0F.pnm}

BUHD {0174F.pnm, 01741.pnm, 0174F.pnm, 0174C.pnm}



HANO {01725.pnm, 0172C.pnm, 0172A.pnm, 0172A.pnm}

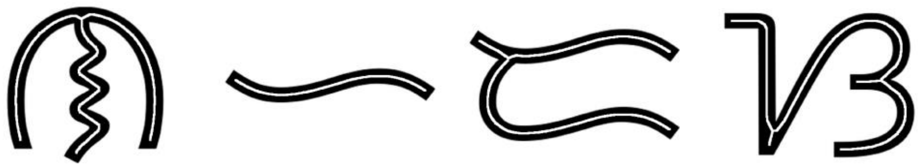
JAVA {0A993.pnm, 0A985.pnm, 0A997.pnm, 0A991.pnm}

RJNG {0A93D.pnm, 0A93E.pnm, 0A944.pnm, 0A935.pnm}

SUND {01B8F.pnm, 01BBD.pnm, 01B9E.pnm, 01BA0.pnm}

TAGB {01767.pnm, 01764.pnm, 01762.pnm, 01762.pnm}

TGLG {01708.pnm, 01711.pnm, 01707.pnm, 01710.pnm}

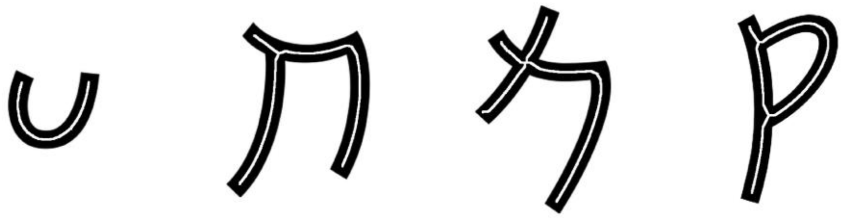


THAA {00793.pnm, 0078F.pnm, 00784.pnm, 00796.pnm}

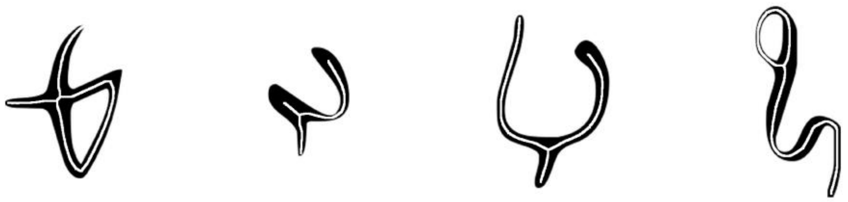
Middle East

ARAB {00634.pnm, 00621.pnm, 00648.pnm, 00649.pnm}

ARMI {1084F.pnm, 10847.pnm, 1084C.pnm, 10852.pnm}



AVST {10B28.pnm, 10B0F.pnm, 10B2C.pnm, 10B16.pnm}

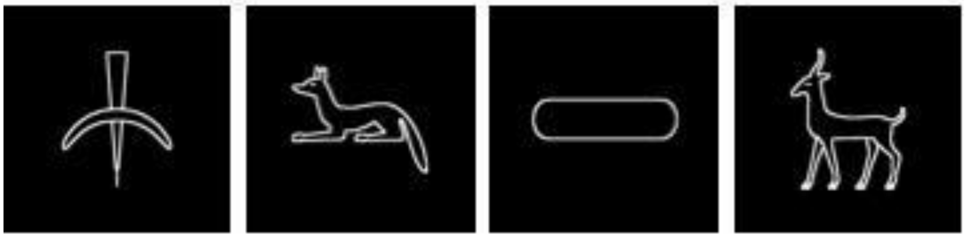


EGYP {13427.pnm, 130E2.pnm, 13200.pnm, 130F4.pnm}

Because line thickness of the resized characters and of the fully processed characters were almost the same, the difference between both is hard to see. The next pictures allow for better comparison:

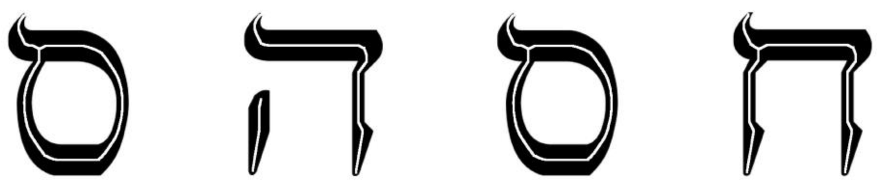
- resized characters

- finally processed characters

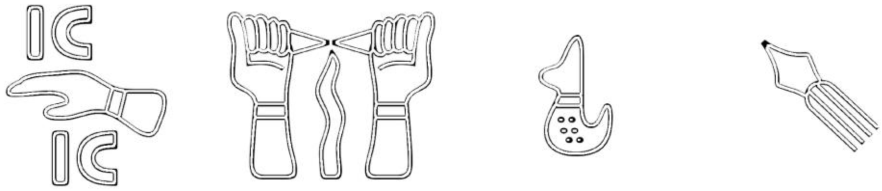


ETHI {012D0.pnm, 0126B.pnm, 01309.pnm, 012A5.pnm}

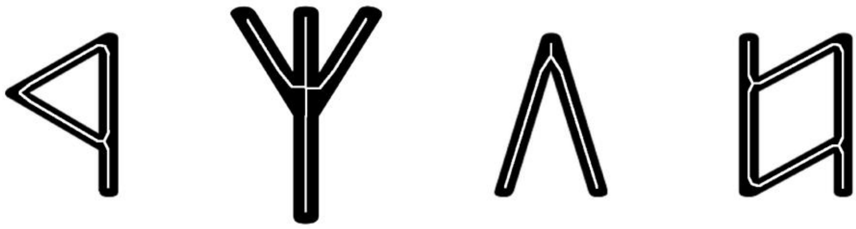
HEBR {005E1.pnm, 005D4.pnm, 005E1.pnm, 005D7.pnm}



HLUW {14430.pnm, 1441F.pnm, 14533.pnm, 14544.pnm}



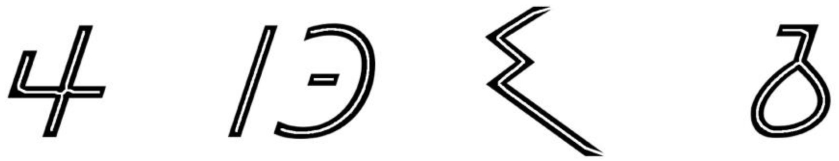
HUNG {10CC1.pnm, 10CB0.pnm, 10CE4.pnm, 10CC6.pnm}



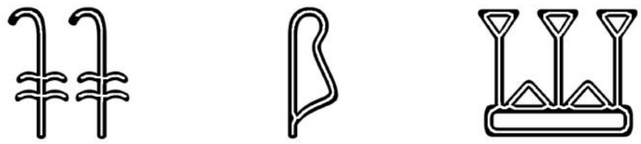
MAND {0084D.pnm, 00849.pnm, 00843.pnm, 00855.pnm}

MANI {10ADB.pnm, 10ACA.pnm, 10ACB.pnm, 10AC5.pnm}

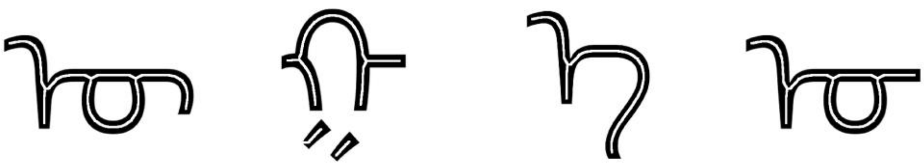
MERC {109A2.pnm, 109B3.pnm, 109A7.pnm, 109A5.pnm}



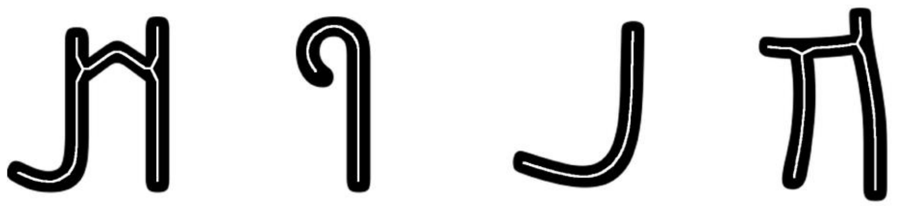
MERO {1098C.pnm, 10981.pnm, 10993.pnm, 10981.pnm}



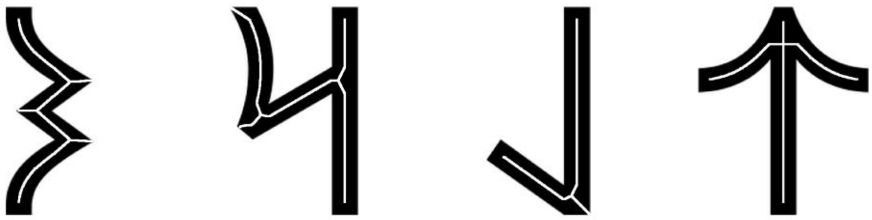
MONG {01825.pnm, 0182D.pnm, 01821.pnm, 01824.pnm}



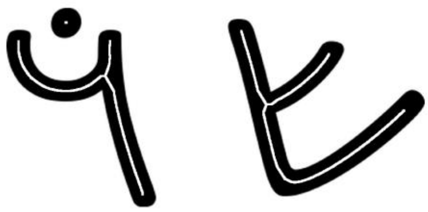
NBAT {1088A.pnm, 10888.pnm, 10883.pnm, 10887.pnm}



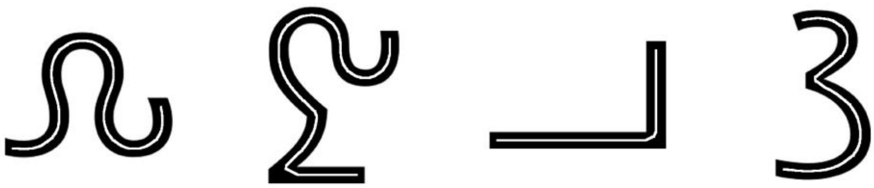
ORKH {10C28.pnm, 10C3A.pnm, 10C1E.pnm, 10C39.pnm}



PALM {10874.pnm, 10875.pnm, 10863.pnm, 1086A.pnm}

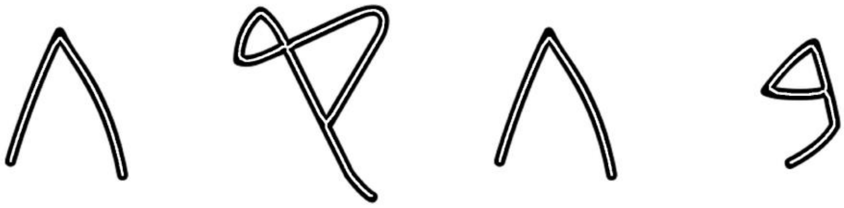


PHLI {10B67.pnm, 10B70.pnm, 10B61.pnm, 10B63.pnm}



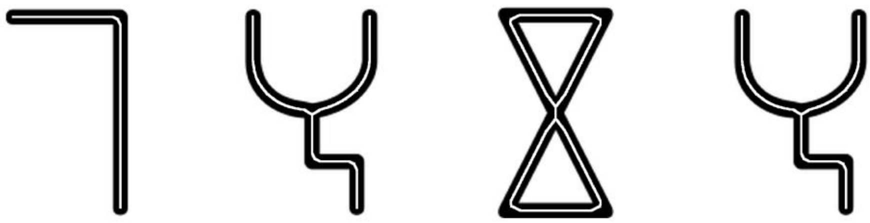
PHLP {10B88.pnm, 10B87.pnm, 10B8D.pnm, 10B80.pnm}

PHNX {10902.pnm, 10912.pnm, 10902.pnm, 10901.pnm}



PRTI {10B51.pnm, 10B54.pnm, 10B43.pnm, 10B44.pnm}

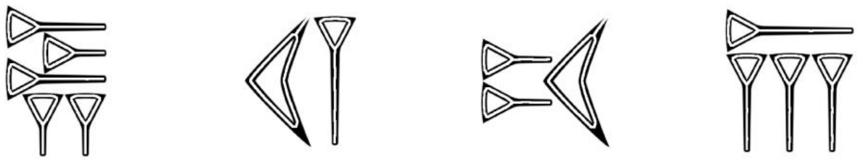
SARB {10A74.pnm, 10A6D.pnm, 10A78.pnm, 10A6D.pnm}



SYRC {00715.pnm, 0072A.pnm, 0072C.pnm, 0072B.pnm}

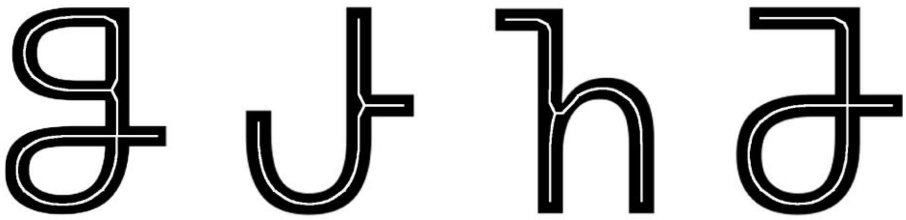
UGAR {1038D.pnm, 10392.pnm, 10382.pnm, 10393.pnm}

XPEO {103B1.pnm, 103A4.pnm, 103B4.pnm, 103A0.pnm}

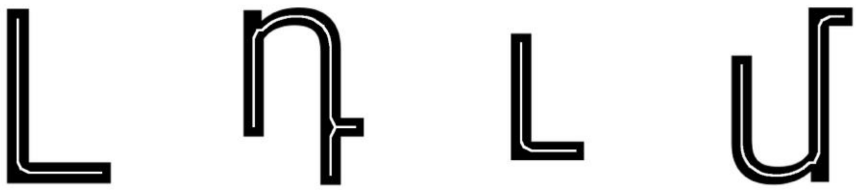


Phoenician

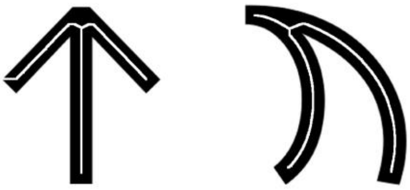
AGHB {1054A.pnm, 1054F.pnm, 1055A.pnm, 10550.pnm}



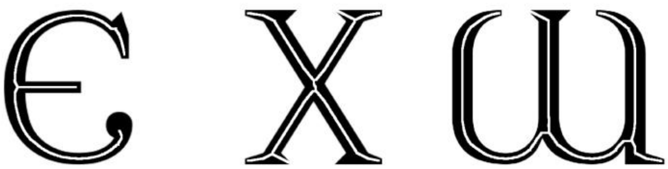
ARMN {0053C.pnm, 00564.pnm, 00582.pnm, 00574.pnm}



CARI {102C7.pnm, 102CB.pnm, 102CA.pnm, 102B2.pnm}

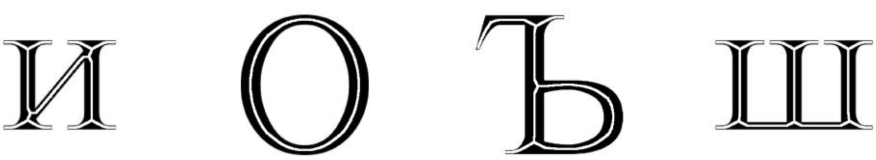


COPT {02C88.pnm, 02CAC.pnm, 02CB0.pnm, 02C97.pnm}



CPRT {10829.pnm, 1083F.pnm, 10820.pnm, 10819.pnm}

CYRL {00438.pnm, 0041E.pnm, 0042A.pnm, 00448.pnm}



CYRS {00465.pnm, 00467.pnm, 0047E.pnm, 00461.pnm}

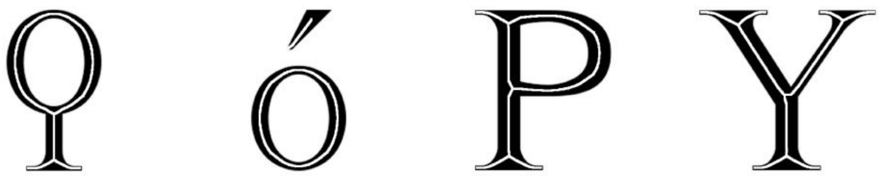
ELBA {1051F.pnm, 10513.pnm, 10519.pnm, 10515.pnm}

GEOR {010E7.pnm, 010DD.pnm, 010DB.pnm, 010DC.pnm}

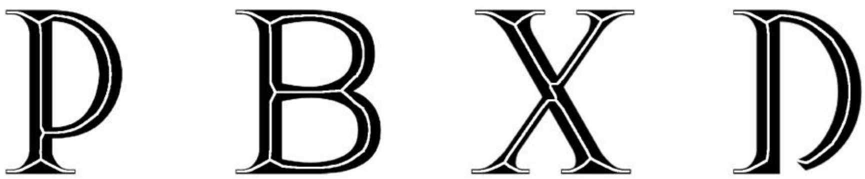
GLAG {02C37.pnm, 02C5A.pnm, 02C32.pnm, 02C5D.pnm}

GOTH {1033D.pnm, 10343.pnm, 1033F.pnm, 10340.pnm}

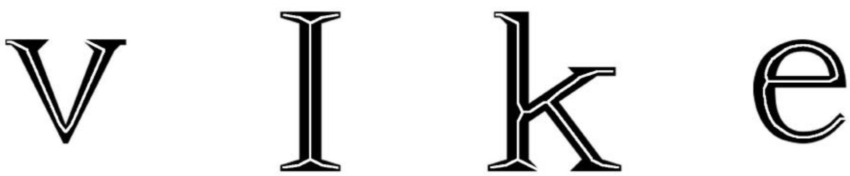
GREK {003D9.pnm, 003CC.pnm, 003A1.pnm, 003A5.pnm}



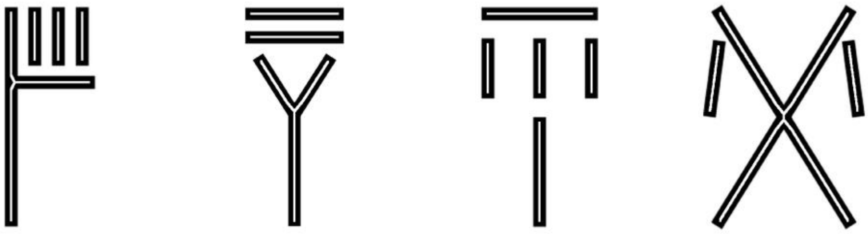
ITAL {10313.pnm, 10301.pnm, 10317.pnm, 10303.pnm}



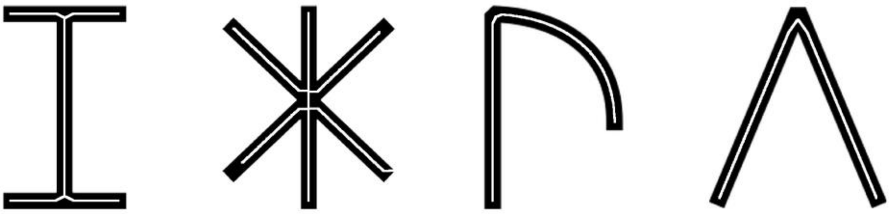
LATN {00076.pnm, 00049.pnm, 0006B.pnm, 00065.pnm}



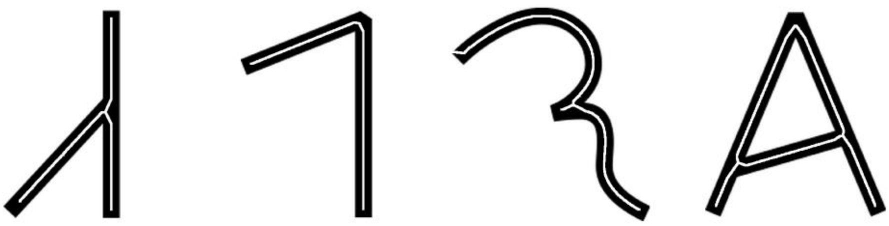
LINB {1002E.pnm, 10019.pnm, 10007.pnm, 1000B.pnm}



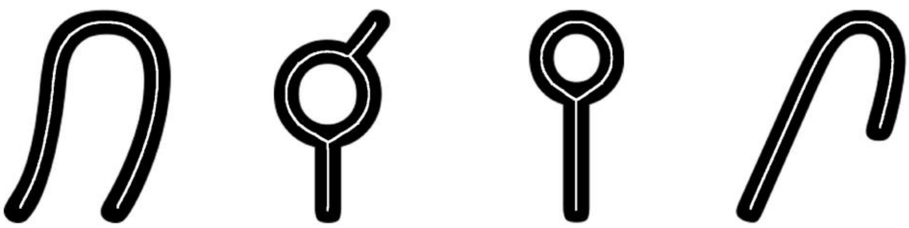
LYCI {10288.pnm, 1028C.pnm, 10293.pnm, 1028D.pnm}



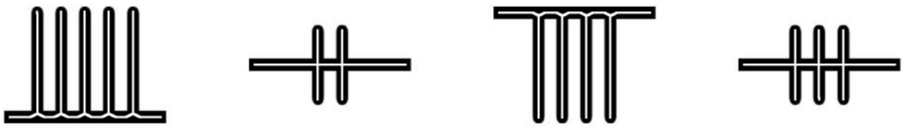
LYDI {10923.pnm, 10929.pnm, 1092E.pnm, 10920.pnm}



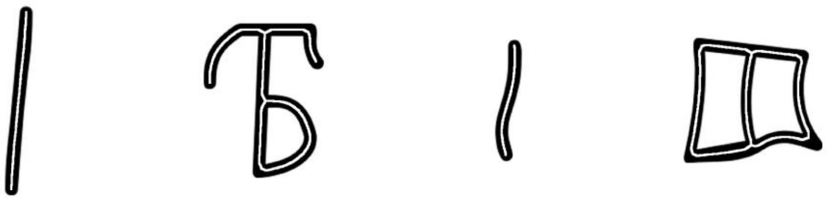
NARB {10A83.pnm, 10A84.pnm, 10A9A.pnm, 10A81.pnm}



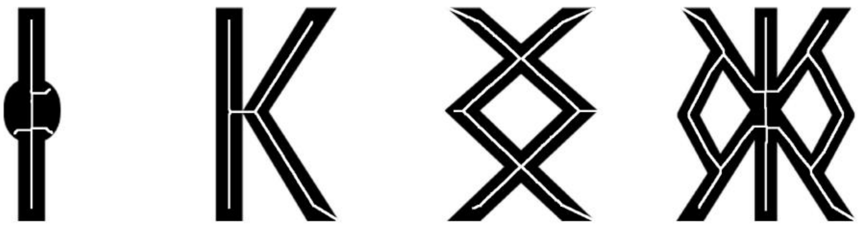
OGAM {0168A.pnm, 01691.pnm, 01684.pnm, 01692.pnm}



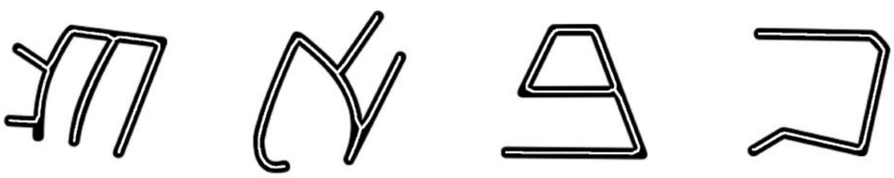
PERM {1036D.pnm, 10371.pnm, 10358.pnm, 10356.pnm}



RUNR {016C2.pnm, 016D5.pnm, 016DD.pnm, 016E4.pnm}



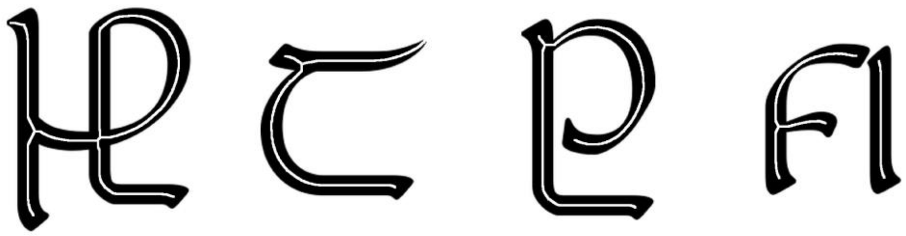
SAMR {00811.pnm, 00800.pnm, 00801.pnm, 00810.pnm}



TFNG {02D3E.pnm, 02D53.pnm, 02D30.pnm, 02D4C.pnm}

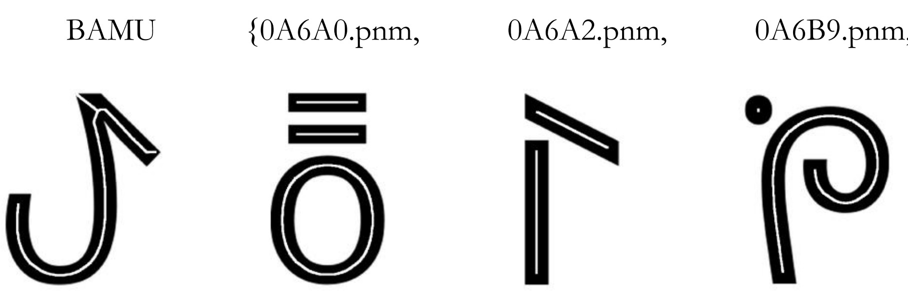
Recent Inventions

ADLM {1E904.pnm, 1E930.pnm, 1E920.pnm, 1E931.pnm}



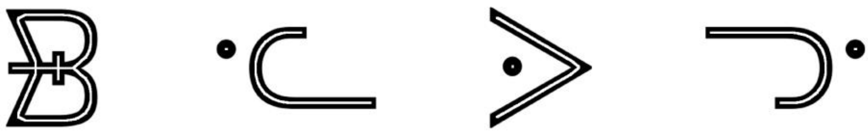
AFAK {16C92.pnm, 16C82.pnm, 16CAB.pnm, 16C97.pnm}

BAMU {0A6A0.pnm, 0A6A2.pnm, 0A6B9.pnm, 0A6AC.pnm}

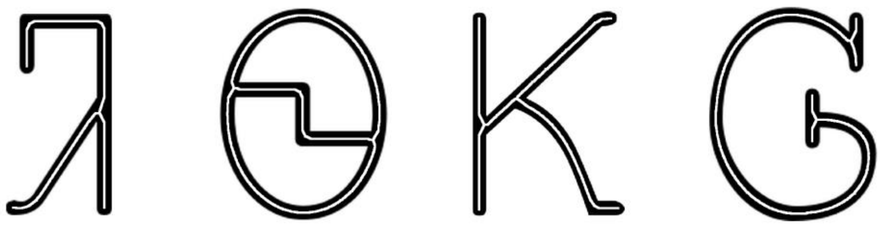


BASS {16AE7.pnm, 16AED.pnm, 16AD4.pnm, 16AEB.pnm}

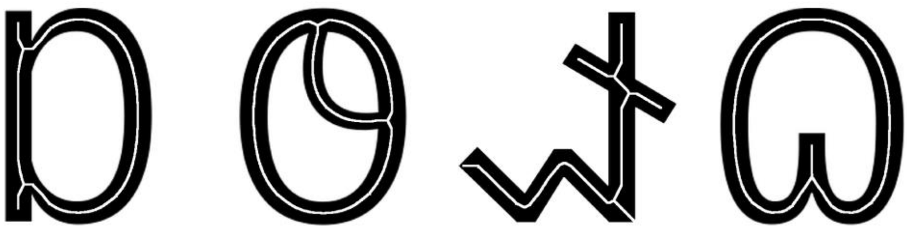
CANS {015FE.pnm, 014E6.pnm, 01437.pnm, 014DD.pnm}



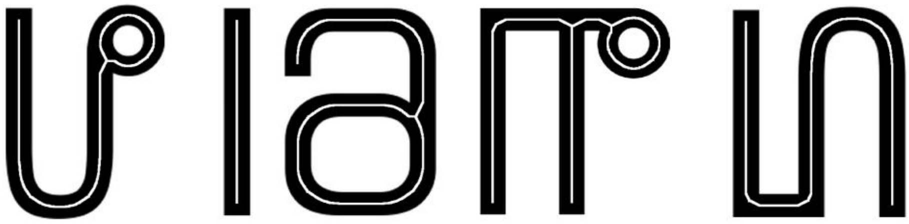
CHER {013D8.pnm, 013EB.pnm, 013E6.pnm, 013B6.pnm}



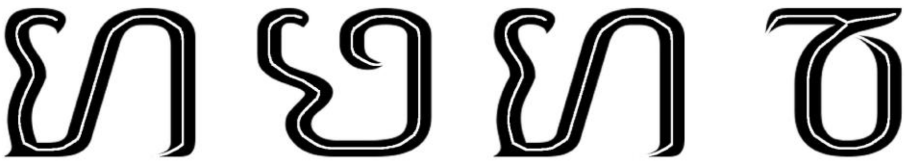
DSRT {1041F.pnm, 10403.pnm, 10426.pnm, 10417.pnm}



HMNG {16B17.pnm, 16B1A.pnm, 16B1F.pnm, 16B1E.pnm}



HMNP {1E105.pnm, 1E116.pnm, 1E105.pnm, 1E12B.pnm}

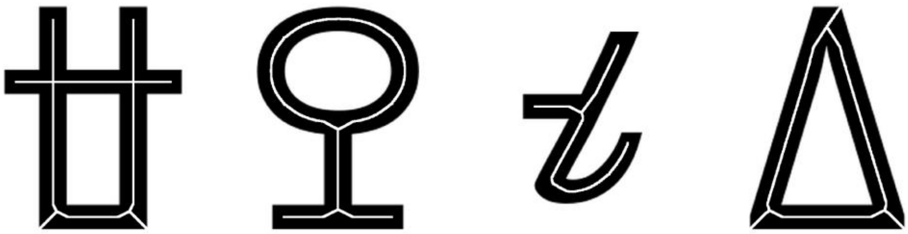


LISU {0A4F3.pnm, 0A4D2.pnm, 0A4DD.pnm, 0A4

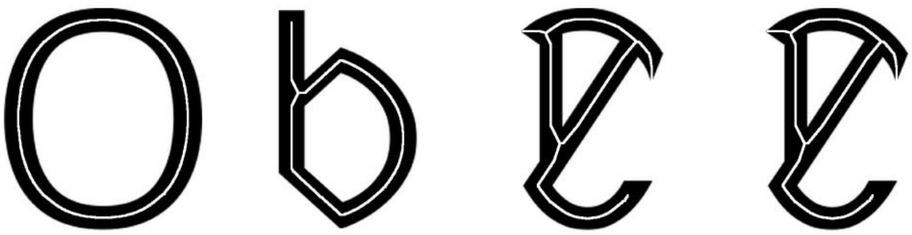
MEND {1E87D.pnm, 1E817.pnm, 1E866.pnm, 1E812.pnm}

MROO {16A54.pnm, 16A4A.pnm, 16A4E.pnm, 16A4A.pnm}

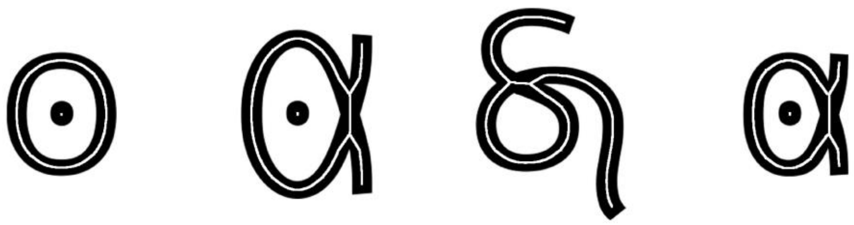
NKOO {007DA.pnm, 007D0.pnm, 007E8.pnm, 007E1.pnm}



OLCK {01C5B.pnm, 01C60.pnm, 01C71.pnm, 01C71.pnm}



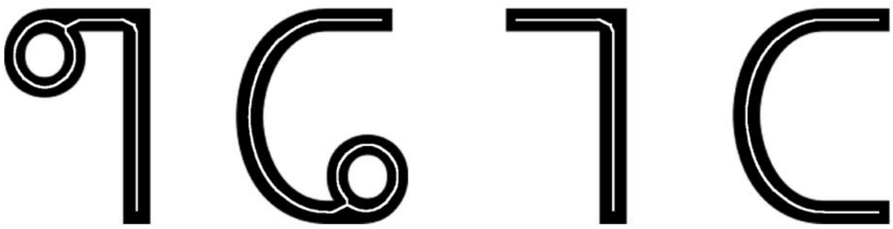
OSGE {104EB.pnm, 104B8.pnm, 104C7.pnm, 104E0.pnm}



OSMA {1049D.pnm, 10495.pnm, 1048F.pnm, 10493.pnm}

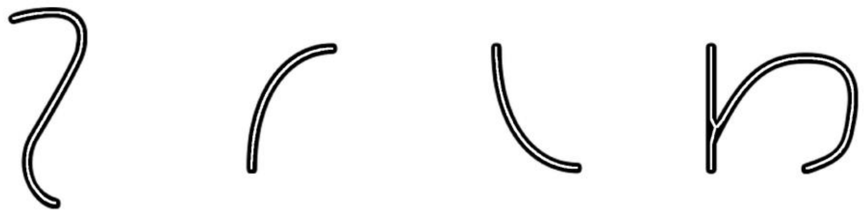
PAUC {11AC4.pnm, 11AE0.pnm, 11ACE.pnm, 11AC9.pnm}

PLRD {16F27.pnm, 16F23.pnm, 16F26.pnm, 16F10.pnm}

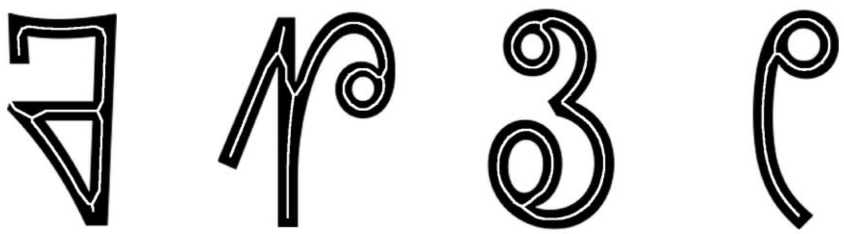


ROHG {10D0A.pnm, 10D0D.pnm, 10D20.pnm, 10D0E.pnm}

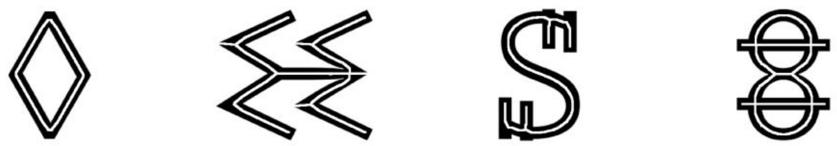
SHAW {1045F.pnm, 10469.pnm, 10467.pnm, 1047D.pnm}



SORA {110D2.pnm, 110D0.pnm, 110D1.pnm, 110E8.pnm}



VAII {0A57A.pnm, 0A612.pnm, 0A5DF.pnm, 0A5CA.pnm}



WARA {118B3.pnm, 118B4.pnm, 118AB.pnm, 118C7.pnm}

Appendix D - For Chapter 4

Additional figures

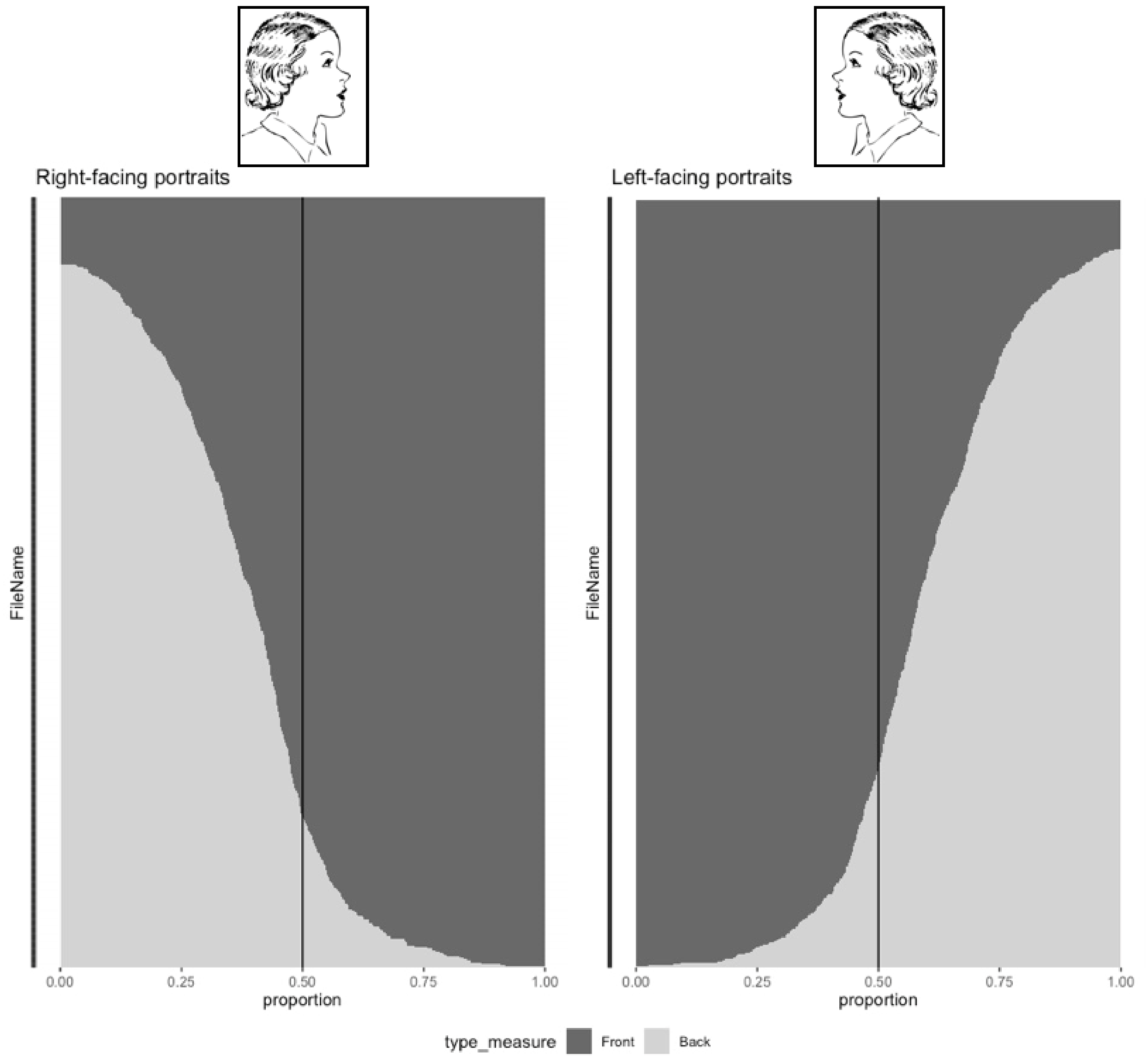


Figure 1. All portraits from our dataset (each horizontal line represents one portrait), with their proportion of free space in front and behind the sitter's head.

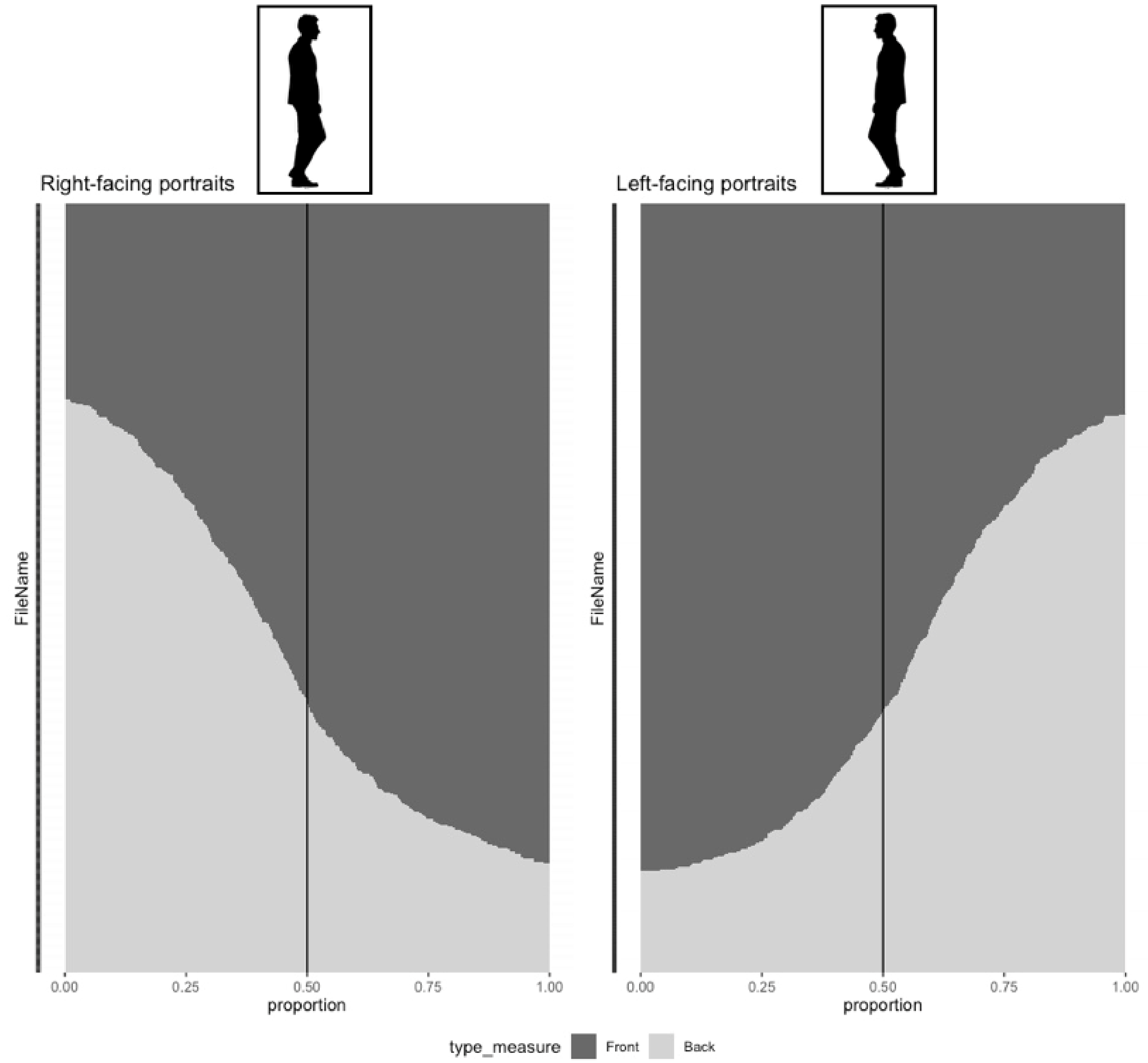


Figure 2. All portraits from our dataset (each horizontal line represents one portrait), with their proportion of free space in front and behind the sitter's body.

Additional details on methodological decisions

Historical and painter information

Whenever the date was a bracket (e.g., 1452-1466), we recoded the date as the upper bracket, i.e., more recent year (in this example, 1466). Whenever the date was indicating a decade, it was recoded as the middle year of the decade (e.g., 1870s was recoded as 1875). Whenever the date was an imprecise indication, e.g., « before 1906 », « after 1474 », or « 19th C », it was recoded as NA.

Treatments of frames and backgrounds

Part of the paintings included in our dataset had frames, and such frames could take various forms, to which we adjusted our measures for distances in light of the type of frame.
External frames are frames that are in another material, typically added around the canvas, after the canvas had been painted on. For such frames, all measures excluded the frame: they were taken to and from the inside limit of the frame, see **(a) in Figure 3**.

Other types of frames included drawn or painted frames. Whenever that frame determined a limit of the portrait – i.e., there is 'blank' space around the frame which is not part of the drawing or painting. In such cases, we used the drawn or painted frame as the reference for our measures, see **(b) in Figure 3.** In some (rare) cases, the portraits depicted more than one (rectangular) frame – in such cases, the smallest rectangular frame was used.

Some other drawn frames are part of the portrait, i.e., they do not delimitate the portrait's frame per se, but are an integral part of the portraits' composition. In such cases, the painted or drawn frames are considered part of the portraits, and the measures are taken as if the drawn or painted frame was part of the portrait itself, see **(c) in Figure 3**.

Whenever a background was drawn with some blank space around it, the blank space was not considered part of the portrait, even if no frame per se was drawn, see **(d) in Figure 3**. This happened in particular whenever the portraits were reproduced on a page or leaflet.

(a) **(b)** **(c)** **(d)**

Figure 3. Examples of the different cases of frames and backgrounds encountered: (a) external hard frame in another material (*Spring* by Franz Stuck), (b) drawn frame that delimitates the portrait (*A Decorative Stain In Red And Green 1897* by Koloman Moser), (c) drawn or painted frames that are part of the portrait (*Portrait of an Unknown 'Schepen'* by Anthony van Dyck), and (d) the background delimitates what is part of the portrait, even if there is no frame drawn per se (*By Volga* by Mikhail Nesterov). The red dashed lines indicate what was considered as the portraits' relevant limits (i.e., the one used for measures), and the black thin lines indicate the limits of the picture we had for that artwork.

Separate results on WikiArt and ArtUK datasets

More analyses, including parametric versions of statistical tests and robustness of the effect to gender of the sitter, are available at:
https://osf.io/6dsnx/?view_only=9835c19958754d5ab0672e7e54a9edae

Hypothesis 1: Prevalence of a forward bias

Both datasets showed that painters had a tendency to put more free space in front, rather than behind the sitter, and so, whether measures were taken from the body or the head of the sitter.

ArtUK

When using head measures, 173 out of 221 of the paintings showed the bias, which is significantly different from chance level of 50% as tested by a Fischer exact test (OR = 12.61, 95%CI [6.44, 25.67], $p < .001$). On average 63.04% of the free space was located in front of the sitter's head, which was significantly more than expected by chance by a one-sample Wilcoxon test ($V = 21273$, $p < .001$, $r = 0.637$). Similarly, when using body measures, 117 out of 196 of the paintings showed the bias, which is significantly different from chance as tested by a Fischer exact test (OR = 4.99, 95%CI [2.49,10.23], $p < .001$). On average 60.59% of the free space was located in front of the sitter's body, which was higher than the 50% expected by chance, by a Wilcoxon test ($V = 9672$, $p < .001$, $r = 0.266$).

This forward bias was present in both left- and right- facing portraits. For head measures, the proportion of the portrait's free space in front of the sitter was not significantly different between left-facing portraits left-facing portraits ($Med = 60.62$) did and right-facing portraits ($Med = 58.99$), by a Mann-Whitney U test - $U = 6674$, $p = 0.187$. For body measures, left-facing portraits ($Med = 62.07$) had not significantly more of their free space in front of the sitter, than right-facing portraits ($Med = 63.45$), by a Mann-Whitney U test, $U = 3520$, $p = 0.791$.

WikiArt

When using head measures, 1222 out of 1610 of the paintings showed the bias, which is significantly different from chance as tested by a Fischer exact test ($OR = 11.88$, 95%CI [9.26, 15.34], $p < .001$). On average 62.22% of the free space was located in front of the sitter's head, which was higher than expected by chance, by Wilcoxon one-sample signed rank test (($V = 1058400$, $p < .001$, $r = 0.566$). Similarly, when using body measures, 809 out of 1423 of the paintings showed the bias, which is significantly different from chance as tested by a Fischer exact test ($OR = 3.43$, 95%CI[2.68, 4.39] $p < .001$). On average 60.55% of the free space was located in front of the sitter, which was higher than expected by chance, by a Wilcoxon one-sample signed rank test, $V = 518440$, $p < .001$, $r = 0.259$.

Again, this forward bias was present in both left- and right- facing portraits. For head measures, right-facing portraits ($Med = 62.32$) has a larger proportion of free space in front of their sitters than left-facing portraits ($Med = 58.86$), Mann-Whitney U test, $U =341460$, $Z = -3.46$, $p < .001$, $r = 0.863$. For body measures, a Mann-Whitney U test ($U = 191370$, $Z = 0.965$, $p = 0.33$) showed that right-facing ($Med = 62.50$) and left-facing portraits ($Med = 64.16$) did not differ significantly in their proportion of free space in front of the sitter.

Hypothesis 2: Historical emergence of the forward bias

Hypothesis 2a: Increase in the prevalence of the portraits with the forward bias

The distribution of dates in our datasets was strongly negatively (left-) skewed, and did not follow a normal distribution. In order to have reliable results on regressions, the date variable was mirrored and log-transformed and finally mirrored back. We ran a binary logistic regression with portrait's date (mirrored, log-transformed and mirrored back) as independent variable and showing (coded as 1) or not showing (coded as 0) the forward bias to determine whether date impacted how likely a portrait was to exhibit the bias. On the ArtUK dataset, this model did show that more recent paintings were slightly more likely to show a forward bias when measured from the sitters' head (OR = 38251, 95%CI [487, 438837], $p < .001$ – Wald χ^2 (2) = 57.8, $p < .001$), but not when measured from her body (OR = 0.21, 95%CI [0.003, 12.9], p = 0.476 – Wald $\chi^2(2)$= 7.0, p = 0.03). Similar results were obtained on the WikiArt dataset : it is was more likely for more recent paintings to show the bias when measured from the head of the sitter (OR = 24.38, 95%CI [4.03, 142.11], $p < .001$ – Wald $\chi^2(2)$= 319.3, $p <.001$), but not when measured from the body (OR = 1.34, 95%CI [0.21, 8.06], p = .748 - Wald – $\chi^2(2)$= 19.8, $p < .001$).

Hypothesis 2b: Increase in the amplitude of the forward bias

Overall ex-centering, i.e., the asymmetry between the spaces on both sides of the sitter (either in the direction of a forward bias or opposite to it), increased over time in both our datasets. There was a positive correlation between date and overall excentricity : the more recent the portrait, the less centered its sitter, on both our datasets and types of measures (ArtUK = : r_τ = .26, $p < .001$, 95% CI [0.156, 0.362] for head, r_τ = .14, p = .012, 95% CI [0.036, 0.25] for body measures ; WikiArt, head, r_τ = .11, $p < .001$, 95%CI [0.077, 0.147], body = r_τ = .08, $p < .001$, 95%CI[0.042, 0.118]).

The more recent a portrait was, the more pronounced was the forward bias, in both our datasets. For the ArtUK dataset, when considering only portraits exhibiting a forward bias (i.e., have more free space in front than behind a sitter), the more recent the portrait, the stronger the forward bias, (r_τ = .24, $p < .001$, 95%CI [0.128, 0.352] for head measures; r_τ = .21, p = .004, 95%CI [0.068, 0.344] for body measures). The WikiArt dataset had similar results: as Date increases, so does the proportion of free space put in front of the sitter (r_τ = .13, $p < .001$, 95%CI [0.089, 0.173]; r_τ = .09, p = .001, 95%CI [0.036, 0.141] for body measures).

Appendix E – For chapter 5

Results presented in the chapter focused on the central part of the rhythmical sequences produced by participants. Here are presented the same analyses, but on the whole sets of 13 taps produced by participants.

Hypothesis 1: All movements equal vs. not all movements equal

Visual inspection confirms that distributions of IOI in the conditions with two types of movements tended to become bimodal, whereas it wasn't the case for conditions in which all movements were of the same type (Figure 1).

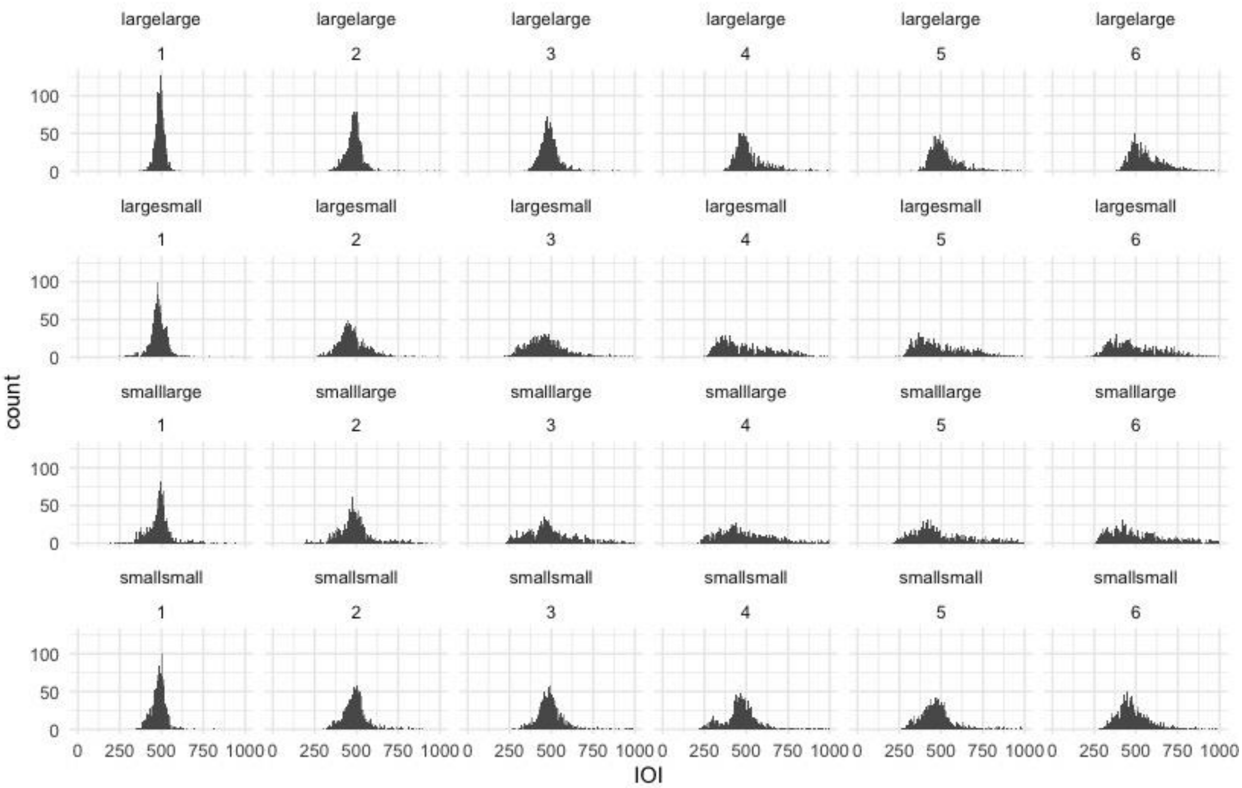


Figure 1. Histogram of IOIs per condition and generation.

Rhythmical structure was assessed using normalized pairwise calculations (nPC)(Condit-Schultz, 2019; Toussaint, 2012). Visual inspection suggests that the distribution of nPC became bimodal for both conditions that included movements of both amplitude (i.e., LARGE SMALL and

SMALL LARGE), but that this was not the case for conditions that included only movements of the same amplitude (LARGE and SMALL SMALL) – see Figure 2.

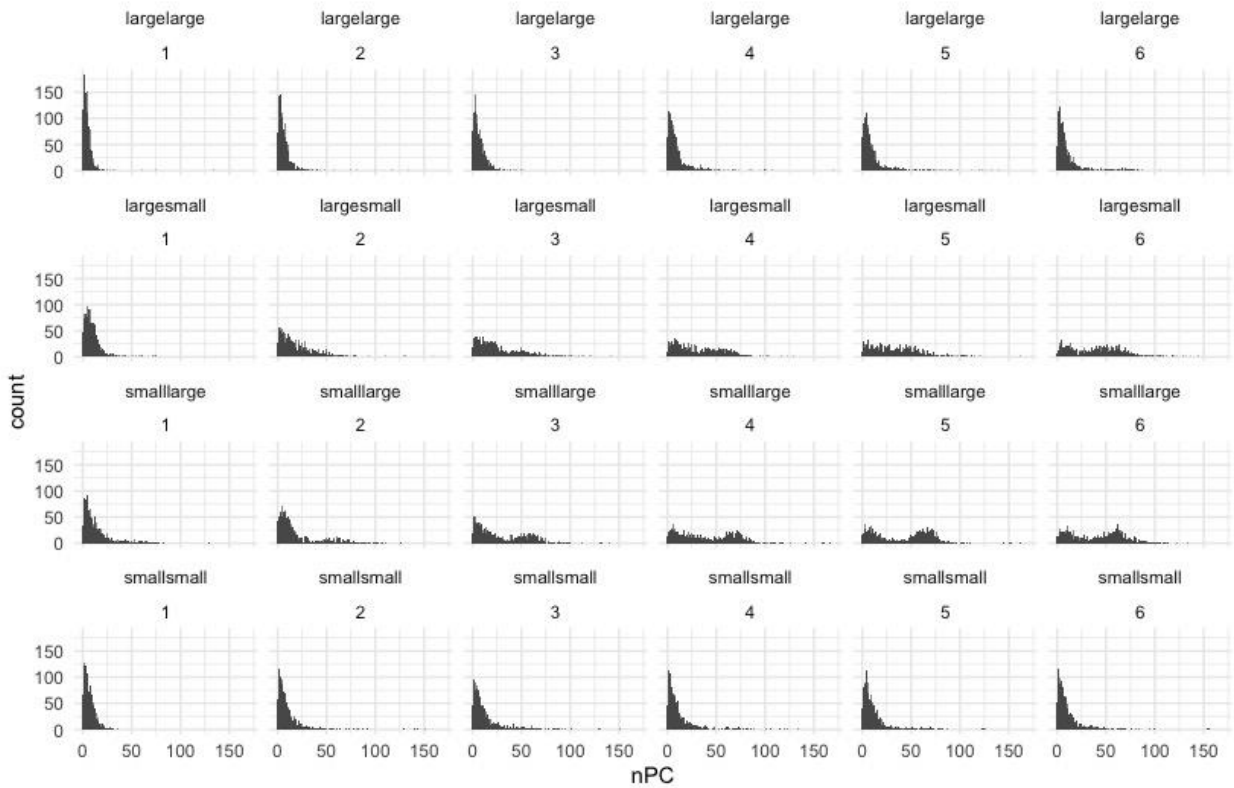


Figure 2. Histogram of nPCs per condition and generation.

In order to test for a difference in the types of rhythm, we computed the normalized pairwise variability index (nPVI, see below) for each sequence produced by participants.

The normalized pairwise variability index is a measure that allows for a minimal value of 0 when all IOIs are equal, and increases as a sequence gets more unequal IOIs. This distribution of nPVIs is used to test whether there is a change from the seed (i.e., metronome sequence): any divergence from this rhythm translates to an increase of the nPVIs. Overall, nPVI increased for both conditions with only one type of movement amplitude ($L = 813$, $k = 6$, $N = 10$, $p < .001$ including the first generation, but not when excluding the first generation, $L = 466$, $k = 5$, $N = 10$, $p = 0.1654$) and unequal conditions ($L = 879$, $k = 6$, $N = 10$, $p < .001$ including the first generation, $L = 520$, $k = 5$, $N = 10$, $p < .001$ excluding the first generation).

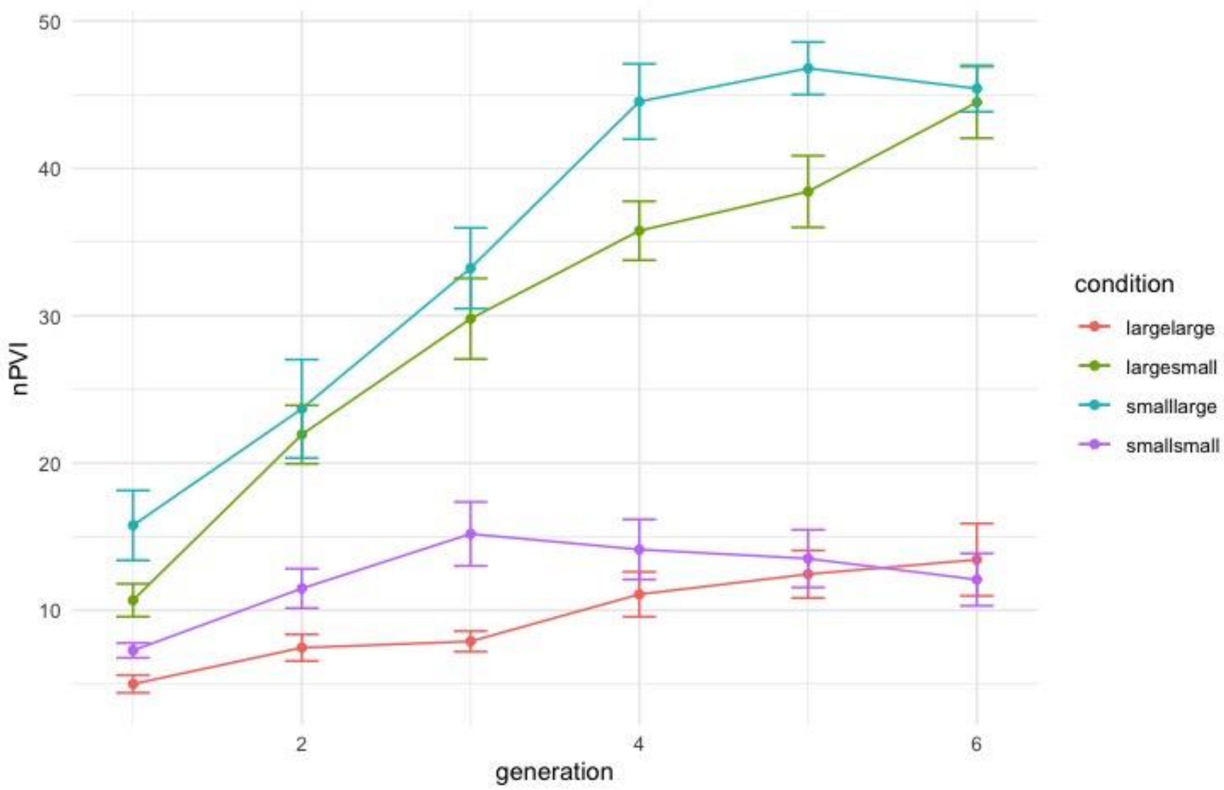


Figure 3. Normalized pairwise variability index (nPVI) by generation, colour represents the different conditions. Error bars represent standard 95% confidence intervals.

A Kolmogorov-Smirnov distance on distribution of nPVI at the last generation confirmed that the equal movement conditions (Large large and Small small) had a different nPVI from the unequal movement conditions (Large Small and Small Large), $D = 0.86$, $p < .001$. A t-test at the final generation suggested that unequal movement conditions ($M = 44.95$, $SD = 11.07$) had higher nPVIs than equal movement conditions ($M = 12.76$, $SD = 11.50$), $t(454.58) = 30.52$, $p < .001$, $d = 2.85$).

Hypothesis 2: Small versus large movements

We predicted that both conditions with all movements equal (Large large and Small small) would show isochronous rhythms, but with different IOIs (Small small should have shorter IOIs than Large large). A t-test at the final generation indicated that the Small Small condition ($M = 496$ ms, $SD = 175$ ms) had shorter ITIs than the Large Large condition ($M = 580$ ms, $SD = 142$ ms; $t(2617.79) = 13.83$, $p < .001$, $d = 0.53$), see Figure 4. IOIs were not normally distributed (Shapiro Wilk: $W = 0.692$, $p < .001$), but the difference between the IOI produced in both conditions were also confirmed by a Mann Whitney U test (U = 1353100, $p < .001$): IOI produced in the Large large condition (Med = 542 ms) were larger than the ones produced in the Small small condition (Med = 467 ms).

A mixed-effects model, including condition and generation as main effects, and participants nested by chain as a random effect, showed that this pattern emerged over time. There was a significant interaction effect between condition and generation ($\beta = -18.859$, SE = 9.007, $t(55.979) = -2.094$, $p = .0408$), indicating that as generations passed, the difference in IOI between the Large large and the Small Small conditions increased. There was also a significant effect of generation ($\beta = 19.696$, SE = 6.37, $t(55.989) = 3.092$, $p = .0031$), but not of condition ($p = .29$) - see Figure 4.

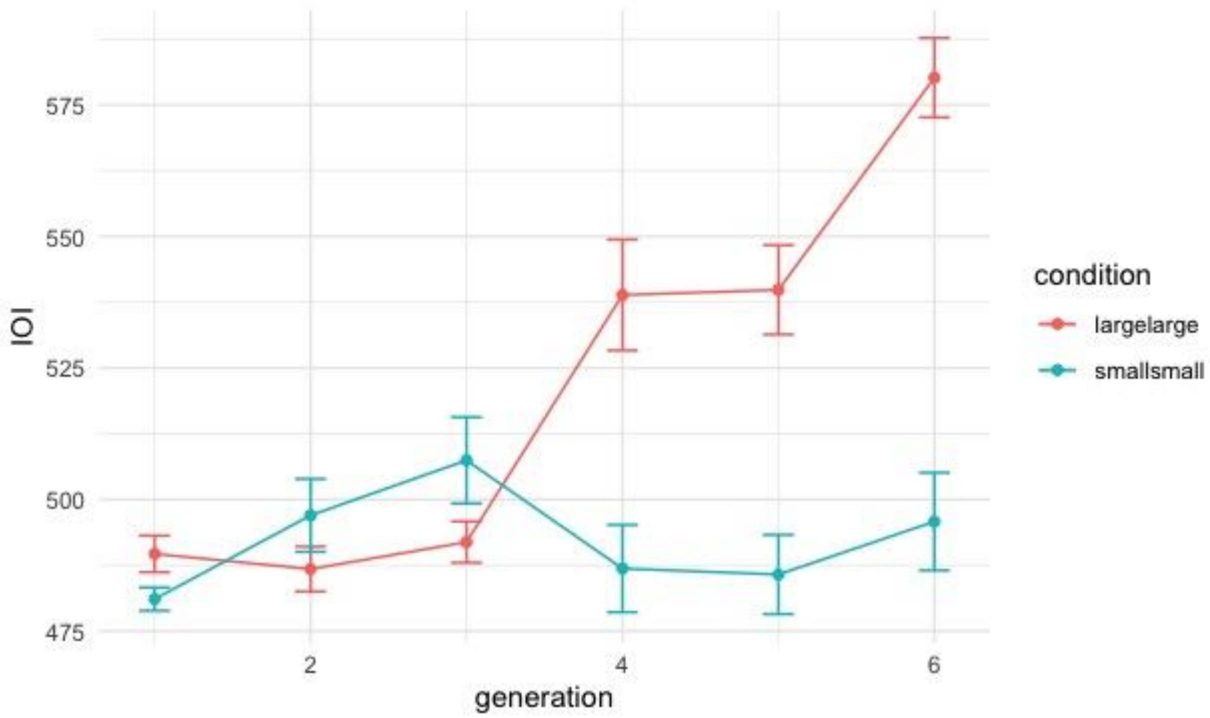


Figure 4. Mean InterOnset Intervals (IOIs) by condition (Large Large or Small Small) and generation (first to sixth). Error bars represent 95% confidence intervals.

Hypothesis 3: Large movement as the first or the third of the sequence

Condition and order in the sequence (MapIOI_Index) were used as fixed effects, and participant nested by chain were used as random effects. In the last generation, the mixed effects model revealed significant effects of both order in sequence ($\beta = -72.108$, SE = 5.584, $t(2760) =$ -

12.913, $p < .001$), and condition (i.e., Small large, compared to Large Small – $\beta = -460.950$, SE = 51.011, $t(9.78) = -9.036$, $p < .001$). An interaction effect between condition and position of the Tap confirmed our prediction ($\beta = 246.25$, SE = 7.88, $t(2760) = 31.25$, $p < .001$), see Figure 5.

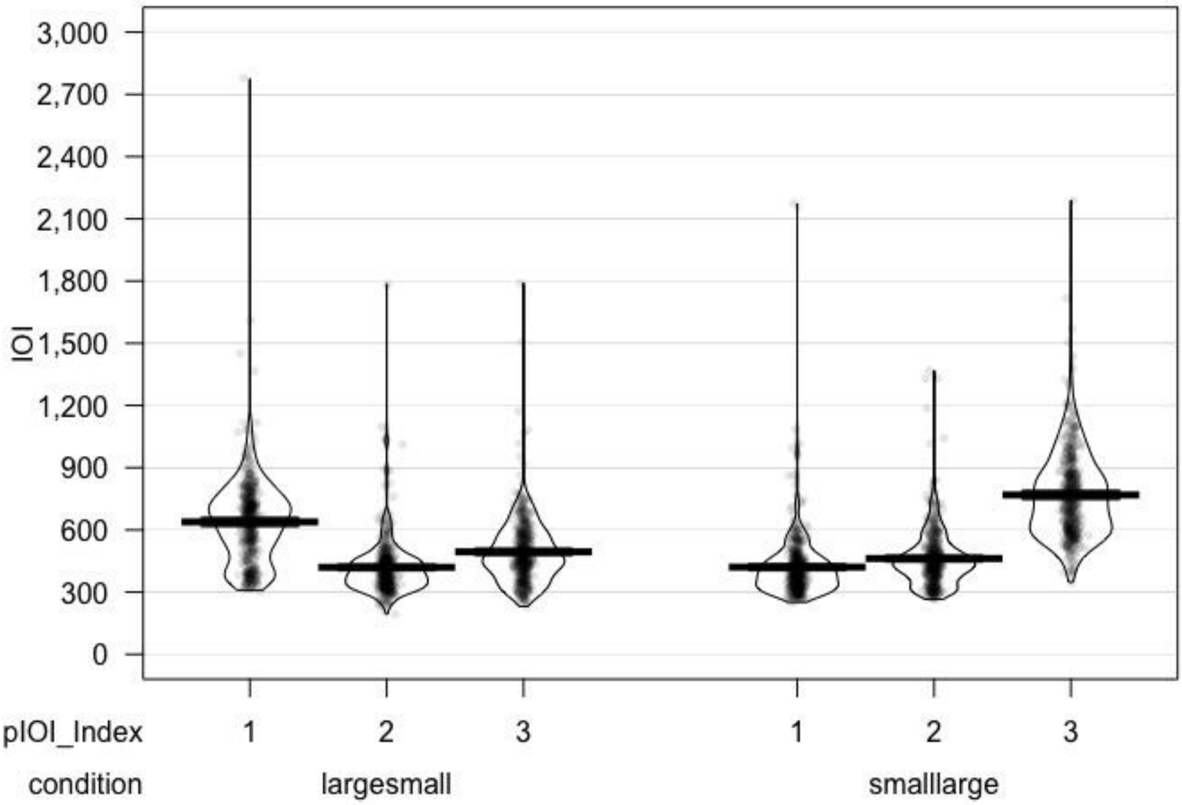


Figure 5. IOI by their position in the sequence, by condition, at the last generation. The coloured bands represent the 95% confidence intervals, points the raw data, and the external curve is the distribution's density.

A similar mixed effects model was run on data including all six generations, with generation as a main effect. Results of this model suggested that the difference emerged over the course of the experiment, as we observed a three-way interaction effect between condition, generation and order in the sequence ($\beta = 38.090$, SE = 1.682, $t(16568) = 22.641$, $p < .001$). The mixed effects model also included significant effects of condition ($\beta = -98.026$, SE = 40.251, $t(70.087) = -2.435$, $p = .017$), generation ($\beta = 27.954$, SE = 7.31, $t(70.163) = 3.824$, $p < .001$), and order in sequence ($\beta = -29.263$, SE = 4.643, $t(16568) = -6.303$, $p < .001$), as well as interaction effects between condition and generation ($\beta = -73.089$, SE = 10.335, $t(70.087) = -7.072$, $p < .001$), between order in sequence and condition (β = 55.089, SE = 6.552, t(16568) = 8.408, p < .001), and finally order in sequence and generation (β = -9.482, SE = 1.192, t(16568) = -7.953, p < .001).

Hypothesis 4: Divergence

The JSD was calculated between each trial to each trial from other conditions, in each generation. The average distance of a chain to other chains that aren't from the same condition (the divergence between conditions) increased over time, as confirmed by a Page trend test confirmed that the JSD between conditions increased over generations, whether we included the first generation (L = 1769, k = 6, N = 20, p < .001) or not (L = 1049, k = 5, N = 20, p < .001), see Figure 6.

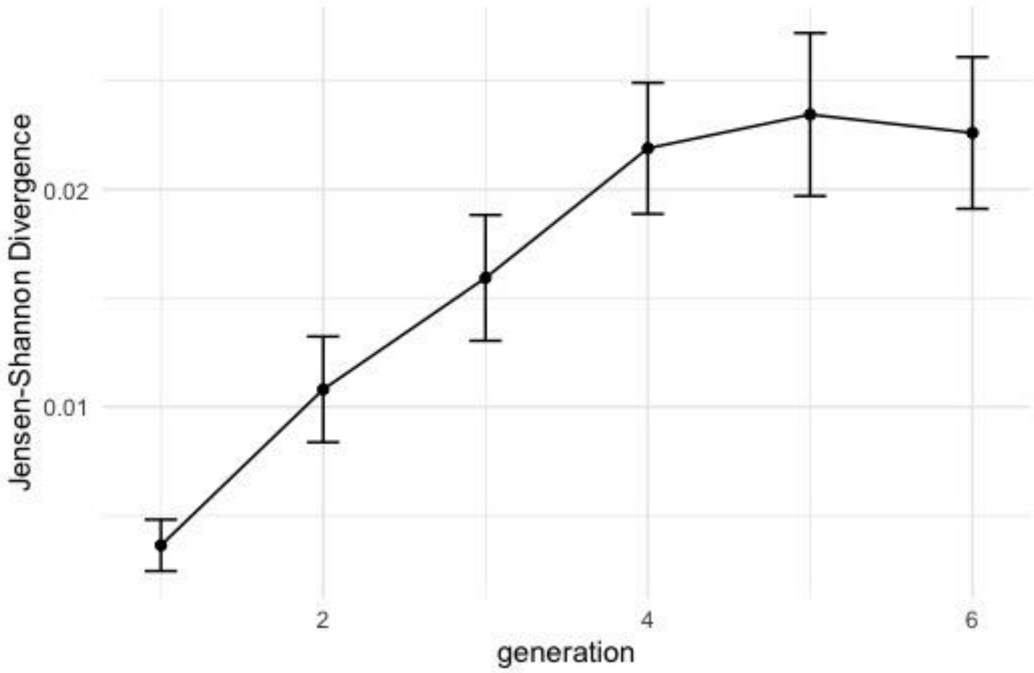


Figure 6. JSD calculated between each trial and all trials from different conditions at each generation, by generation. Error bars represent the 95% confidence intervals.

Another way to visualize such differences is to use ternary plots. We plotted the IOIs on a triangular simplex, such that each side of the simplex represents either the first IOI of the sequence, the second one, or the third one. As our design includes a cycle of three movements (3 IOIs, produced from 4 taps), this is particularly fitting and allows us to have a quick, visualization-based idea of how the rhythms evolved in the different conditions (Figure 7).

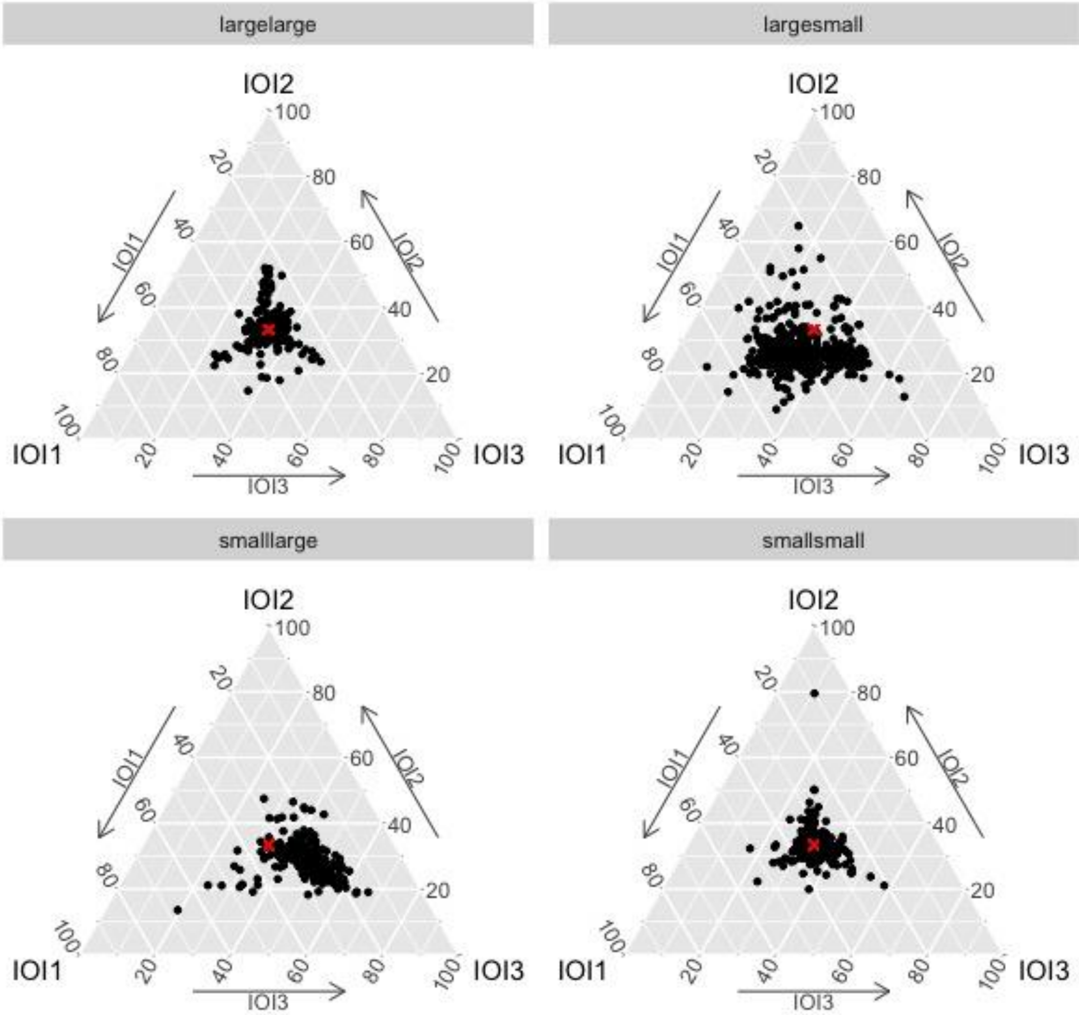


Figure 7. Ternary plots of the distribution of IOIs at the last generation, for each condition. The red cross indicates where is the 1:1:1 integer ratio (i.e., the metronome-like sequence with which the chains were seeded).

Hypothesis 5: Stability

Edit Time distance

This time distance increased over time, as confirmed by a Page Trend test, both when including the first generation ($L = 1694$, $k = 6$, $N = 20$, $p < .001$), or excluding the first generation ($L = 990$, $k = 5$, $N = 20$, $p < .001$), see Figure 8.

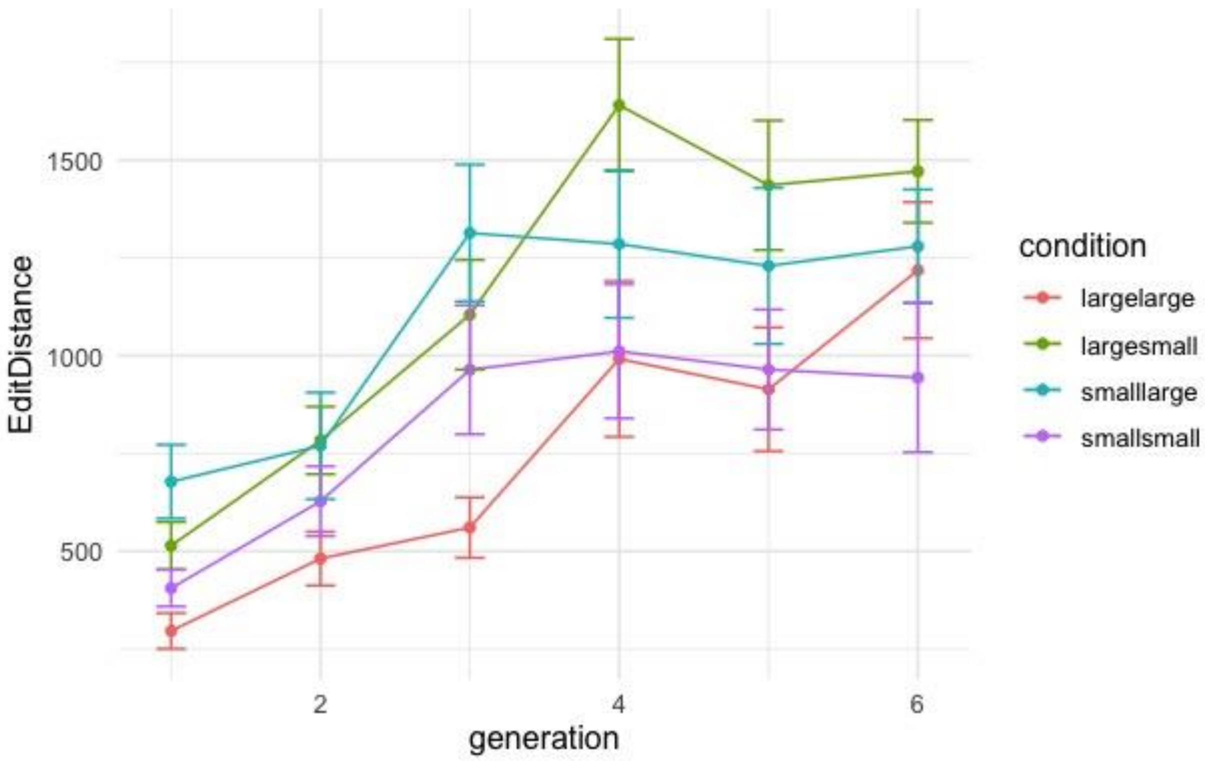


Figure 8. Edit time distance, by generation and condition. Error bars represent 95% confidence intervals.

Stability of the rhythmical structure: Edit nPC distance

Here too, the distance increased over time, as confirmed by a Page Trend test, both when including the first generation ($L = 1687$, $k = 6$, $N = 20$, $p <.001$) or excluding it ($L = 992$, $k = 5$, $N = 20$, $p < .001$), see Figure 9.

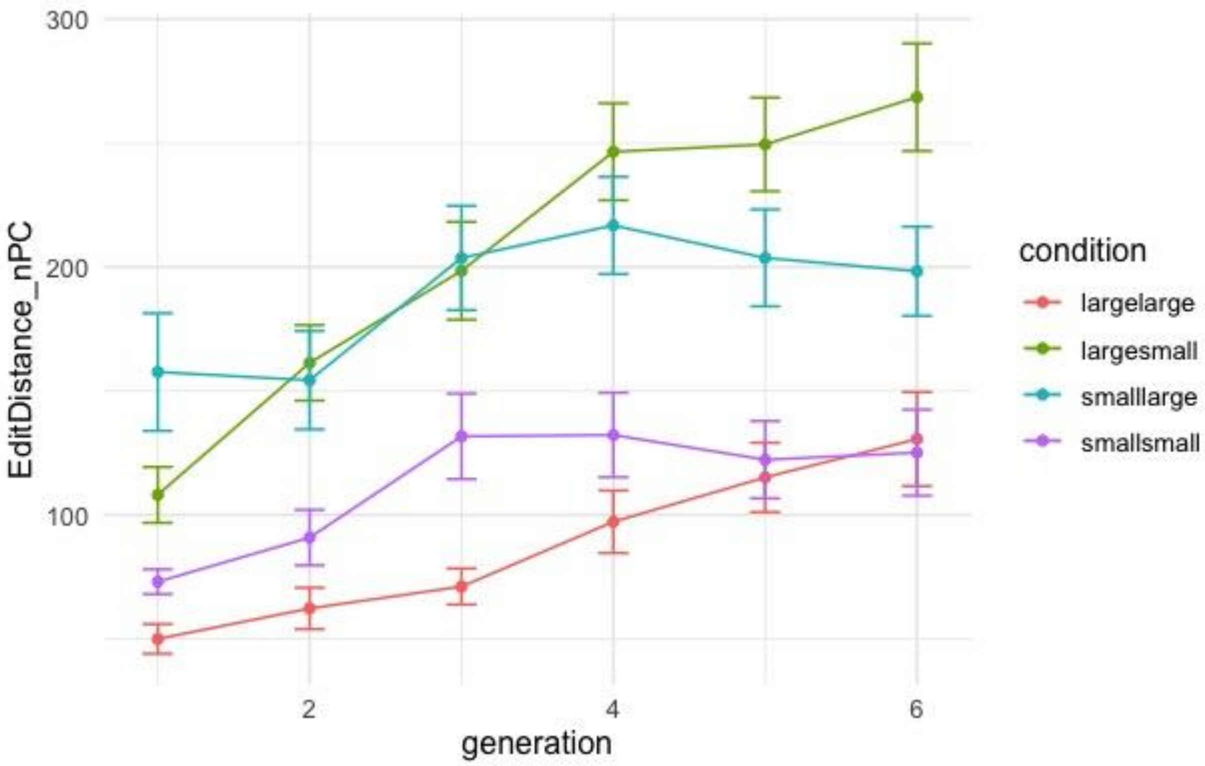


Figure 9. Edit distance based on nPC, by generation and condition. Error bars represent 95% confidence interval.

Relation between conditions and stability

As this pattern was rather unexpected – we predicted an increase in learnability, not an increase in the amount of change - we explored whether this effect was also driven by our conditions.

We ran a mixed effects model with edit distances as the dependent variable, condition and generation as fixed effects (independent variables), and participant and chain as nested random effects.

On the edit time distance, this mixed effects model confirmed that the edit time distance increased with generation ($\beta = 185.76$, SE = 45.67, $t(112.30) = 4.068$, $p < .001$). The only condition to significantly depart from Large Large (our baseline) was Small Large ($\beta = 556.36$, SE = 251.33, $t(111.95) = 2.214$, $p = 0.029$). No other main or interaction effect was significant (all *p*s > .176).

On the edit nPC distance, this mixed effects model revealed a significant effect of generation ($\beta = 17.085$, SE = 5.907, $t(112.349) = 2.893$, $p < .01$), and of both Small Large ($\beta = 124.299$, SE = 32.512, $t(112.095) = 3.823$, $p <.001$) and Large Small ($\beta = 65.60$, SE = 32.519, t(112.181) = 2.017, p = 0.046) conditions, but not of SMALL SMALL condition (p >. 139). These results suggest that conditions including both amplitudes of movements led to higher edit nPC distances, i.e., the difference between what participants heard and produced was higher in those conditions than in conditions including only small or only large movements. There was a trend for an additional interaction effect between generation and the Large Small condition (β = 14.948, SE = 8.35, t(112.181) = 1.790, p = .076), but no effect was significant (all *p*s >.139). Edit distances depended on both generation and, especially when considering rhythmical stability, condition.

ACKNOWLEDGEMENT

TO EXTERNAL FUNDING AGENCIES
CONTRIBUTING TO PHD DISSERTATIONS

Name of doctoral candidate: Helena Miton

Title of dissertation: Explorations in cultural evolution: methodological challenges and case studies

Name of supervisor(s): Christophe Heintz, Dan Sperber

External funding agency: **European Research Council**

Acknowledgement:

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) under **grant agreement No 609819 – SOMICS.**