

# An Information-Theoretic Measure of Cultural Success

Matthew Cashman

Massachusetts Institute of Technology

Word count: 5,253

## Abstract

Models of genetic evolution are tested empirically by counting alleles: a good model of genetic evolution successfully predicts which genes will be found where. Culture is changing humanity at an astounding rate, but at present we lack a means for measuring the flow of memes through minds in a quantitative, content-agnostic way analogous to counting alleles. I develop a method for measuring the information from a written work that is retained in the minds of those exposed to it, and which is therefore capable of influencing behavior. I estimate the entropy of samples from a target written work using a cloze-completion tasks in a treatment group (those that have read a target work) and a control group (those who have not read the target work). In doing this, we use human minds as encoders-decoders in Shannon's communication model. Difference measures taken between the entropy estimated with the treatment group and that estimated with the control group quantifies the information that the treatment group already knows relative to the control group, in bits. This method can control for shared cultural inheritance naturally, and it is content-agnostic—it does not require strong commitments to what information from the target work is important, nor commitments to what questions are important to ask. The technique can be extended to a variety of domains including evolutionary theory, methods of teaching, and the study of music.

Please address correspondence to Matthew Cashman: [matt@cashman.science](mailto:matt@cashman.science)

# Table of Contents

1 Introduction.....	4
2 Aims.....	7
3 Background: information theory.....	9
4 Measuring entropy in language.....	10
5 Applying entropy measurements to information flows through culture.....	13
6 Discussion.....	15
6.1 Limitations.....	17
6.2 Future directions.....	18
6.2.1 Evolutionary theory.....	18
6.2.2 Measurements across multiple different works.....	19
6.2.3 Flow through social networks.....	19
6.2.4 Education.....	19
6.2.5 Charting cultural tropes and archetypes.....	20
6.2.6 Music.....	20
7 References.....	22

## Table of Figures

Figure 1: A diagram showing the communication system under study.....	9
Figure 2: Example data from Shannon (1951).....	12

# 1 Introduction

Natural selection provides us with a success criterion: success is replication. While the process of genetic evolution has remained relatively stable over human history thus far, there has been a remarkable increase in the bandwidth available for transmission of culture in the past few centuries. This increase is due to genetic adaptations for cultural transmission (as in evolved social learning psychology), cultural adaptations increasing transmission bandwidth and fidelity (such as schools) and most recently cultural artifacts that facilitate information storage, transmission, and interpretation (such as books, libraries, telephones, automatic translators, and the internet).

Dual Inheritance Theory, also referred to as gene-culture coevolution (Boyd & Richerson 1985), describes two mutually influential routes by which variations, either cultural or genetic, can be selected among and passed on—or transmitted to—subsequent generations. The gene-culture feedback loop has been present for as long as there has been transmissible culture, beginning with behaviorally modern humans approximately 80,000 years ago (Fisher & Ridley 2013). In brief, dual-inheritance theory describes the way that cultural traits (defined as socially learned information stored in human brains and capable of affecting behavior) change the environment under which genetic selection operates (Richerson & Boyd, 2005). The canonical example of culture influencing genes is lactase persistence: three groups of humans have separately evolved the ability to produce lactase, an enzyme that enables digesting dairy foods, into adulthood. This happened as a direct result of the rise of dairying, and hence the availability of milk products as food, in these cultures (O'Brien & Laland, 2012). Dairying is culturally-transmitted, and this cultural information has resulted in the genetic adaptation of lactase persistence. Causation in the other direction, genetic influence on the ability to transmit culture, is also widespread, manifesting as psychological adaptations for social learning.

Evolution via natural selection at the genetic level is well-understood. Theories of genetic evolution model empirical data, the data being, at its core, simply counts of alleles: we are interested in the frequency of a particular variation of a gene found in some population being measured. Our models of genetic evolution are grounded in this data, and we evaluate models based on their fit with it.

The study of cultural evolution is a burgeoning field, with many important advances made over the last few decades: scholars have developed mathematical models of cultural evolution starting in the 1970s, (e.g. Cavalli-Sforza & Feldman 1981, Boyd & Richerson 1985, Boyd & Richerson 2005, McElreath & Henrich 2007). However, theories of cultural evolution have thus far lacked empirical grounding analogous to counting alleles. What these theories require is a content-agnostic means of measuring the flow of ideas from one human mind to the next, and for measuring the frequency of these ideas in populations of interest.<sup>1</sup> Up to now we have had limited means for empirically verifying models of cultural evolution. Empirical investigation has thus far relied on qualitative techniques, or on quantitative techniques that require two things: first, strong commitments to what pieces of culture are important and second, measurement of behaviors induced outside of natural settings, as in a lab.

Substantive questions in the study of cultural evolution are often motivated by the flow of whole packages of ideas that are meaningfully connected. Consider, for instance, religion and religious texts: the Bible, the Koran, the Vedas, the Torah, the Pāli Canon, etc. Each is a rich tapestry of interwoven ideas. Probing the minds of informants with specific questions about e.g. the Bible may give some insight into their familiarity with the text or the religion, but it does require a commitment to knowing the right questions to ask beforehand. For instance, we might wish to investigate how Christianity has shaped the mind and behavior of the people making up a small-scale society, and so we might ask something like: “What is the name of the mountain on which Moses received the Ten Commandments?”. And answers to this question will probably produce some signal of Christian-ness. But there are many ways that ideas, beliefs, and

---

<sup>1</sup> See Chvaja (2020) for a discussion of Memetics, a largely qualitative and theoretical effort to articulate how ideas might be quantized and their flow modeled, around the turn of the century.

practices might flow from the Bible and in to our subjects' minds while leaving this particular piece of information behind—or at least making it difficult to recall. Even a whole set of exam-type questions makes for a very strong commitment to a particular kind of interaction with the text being the *right* kind of interaction. People in this hypothetical society may, for instance, have focused on the New Testament, or they may have received the substantive content but not the specific labels associated with the Ten Commandments episode. They may not even know these ideas come from the Bible, and so could be completely unable to answer. In either case, having strong commitments to what questions are the right questions means we will have missed a substantial information flow from the text into their minds.

We have seen increasing empirical investigation with laboratory experiments (see Mesoudi 2016 for a review) that clarifies, among other things, the mechanisms by which we decide what information to retain (e.g., prestige), the demographic structure that is required in a society to maintain transmission (e.g. Henrich 2004, Muthukrishna et al. 2014), and the mechanisms by which the information transmitted changes (e.g. Derex et al. 2019). But empirical investigations of the transmission information via culture have thus far involved artificial settings and deliberately induced transmission, as in laboratory experiments, and limitation to semantically simple ideas, and strong commitments as to what the right questions to ask are in all cases. In the wild, the behaviors that result in the transmission of information via culture happen across many different contexts in response to rich sets of cues about what to transmit, what to receive, and from whom. The ideas that flow from one mind to the next are semantically rich and difficult to bound, making laboratory measurement illuminating but necessarily limited. In addition to working outside natural settings, lab experiments must use very particular, necessarily idiosyncratic ideas: a specific drawing task, transmission of a given sentence, certain software, etc.

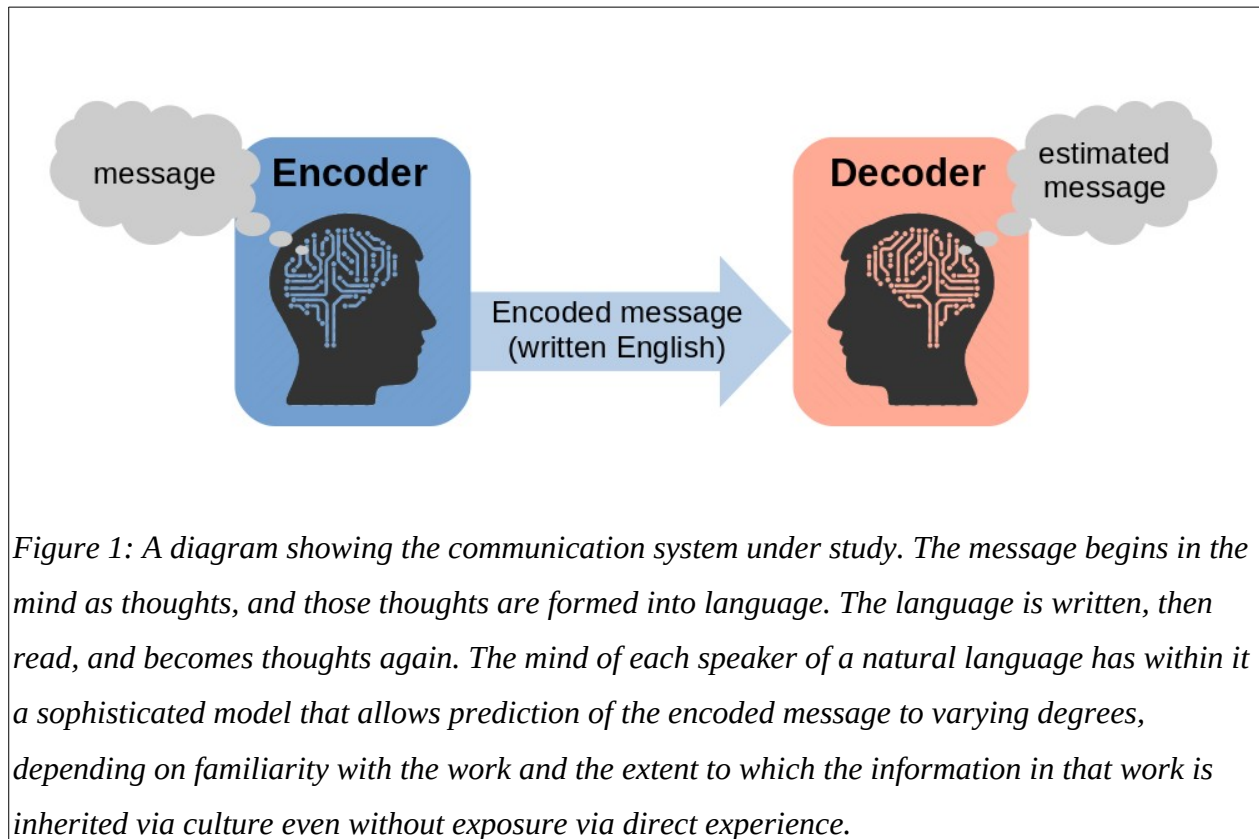
These two classes of limitations highlight the unmet need and potential opportunity developing the empirical underpinnings of theories of cultural evolution. The method set out in this paper is a step towards the kinds of measurement necessary to

make theories of cultural evolution as empirically testable as those of genetic evolution. It is hoped that it is a step towards their further integration as well.

## 2 Aims

The value inherent in modern models of genetic evolution via natural selection is twofold: first, they can predict change in a measurable quantity related to traits (e.g.  $\Delta z$  in the Price equation), and second, their form reflects the underlying data-generating process (albeit greatly simplified). I outline a method for measuring the amount of information from a particular written cultural artifact that is actually present in a human mind, which I will call *retained novel information* (RNI). RNI could allow for a principled measurement of an analog to  $\Delta z$  for culture, as well as allowing for the comparison of information flow via various teaching methods, the mapping of story archetypes across the world, and many other uses.

The method derives from Claude Shannon's 1951 procedure for measuring per-character entropy of written English. Shannon designed this procedure using human minds to predict encoded message content (i.e., written English) because he did not have access to large digitized corpora and because his computers used slide rules and pencils rather than microchips. We all do, in fact, walk around with a detailed knowledge of the statistics of the natural languages we speak held in our minds, and Shannon cleverly took advantage of this fact. Modern computational techniques and corpora remove the need for such a procedure if the quantity of interest is merely the entropy of written language—but what if we were to use human minds to measure entropy *precisely because* we are interested in their properties as codecs (encoders-decoders) when applied to the written word?



The procedure, then, is very simple: define a treatment group (those familiar with a target cultural artifact, perhaps a book they have read) and a control group (matched to the treatment group, but not exposed to the target artifact). Then, each group completes tasks designed to measure the entropy of written language, in our case English, but only with samples from the target artifact. We then compute a difference measure between these two entropy estimates; Kullback-Leiber divergence gives a measure of the number of additional bits it would take to encode the text with the worse (controls') language model, as an example, and simple subtraction of the two estimates gives some measure of storage. The entropy measurement from the treatment group is a property of the coded sequence (in the case of a book, the text) itself *given a particular decoder*, and RNI is the quantity of information from the target that is revealed to be in the decoder—in the mind. This is the information that is both novel relative to the treatment group and the control group's shared experiences and also retained. It is



not the case that all novel information from the target artifact is retained, of course, as memory has evolved to be highly selective (Kuhl & Chun 2014). We can quantify RNI by multiplying the result of the difference measure by the length of the text, denominated in the relevant unit of analysis (character, word, sentence, etc.). This will yield an estimate of the amount of information in a given target artifact for a particular reader.

### 3 Background: information theory

The theory underpinning RNI measurements stems ultimately from *A Mathematical Theory of Communication* (Shannon 1948). In brief, Shannon outlines a measure that describes the information entropy associated with a discrete random variable  $X$ :

$$H(X) = -\sum_{i=1}^n P(x_i) \log_2 P(x_i) \quad (\text{Equation 1})$$

Where  $H$  is information entropy and  $P(x_i)$  is the probability of possible value  $x_i$ . As  $H(X)$  increases, the amount of information a given datum communicates increases. We can take the example of a fair coin and calculate that each flip of that fair coin is one bit:

$$H(X) = -\sum_{i=1}^2 \frac{1}{2} \log_2 \frac{1}{2} \quad H(X) = -\sum_{i=1}^2 \frac{1}{2} \cdot (-1) = 1 \quad (\text{Equation 2})$$

The intuition is that any flip of a fair coin carries the same amount of information—you are no more surprised if it comes up Heads than if it comes up Tails. However, in a situation where all outcomes of an information-generating process are *not* of equal probability, we are more “surprised” by some outcomes than others. Suprisal is formalized as the reciprocal of the probability,  $1/P_x$ , and is perhaps the more intuitive

quantity. In a situation with an unfair coin that has a 99% probability of producing Heads and 1% for tails, we are very surprised when Tails occurs—its probability was low.

$$H(X) = -\left(\frac{99}{100} \log_2 \frac{99}{100} + \frac{1}{100} \log_2 \frac{1}{100}\right) = 0.08 \quad (\text{Equation 3})$$

In this case, each flip of the unfair coin delivers, *on average*, only 0.08 bits of information. This is because we are quite sure that any given flip will result in a Heads. Each flip that results in a Heads delivers much less information than a flip that results in a Tails. A coin with Heads on both sides delivers exactly zero information with each flip; you are completely sure the outcome will be heads. As with a fair coin, a string of letters drawn randomly from the 26 in the English alphabet has the highest possible entropy for that encoding scheme—there is no way to predict what the next character will be on the basis of previous characters. In an optimally compressed data stream it becomes impossible to predict the next datum based on previous data.

This sort of measurement can be applied to any encoding scheme: the bits that are ubiquitous in digital information storage and transmission are a base-2 encoding scheme in that they can take two states (one and zero), language (particularly written language) is amenable to treatment, DNA uses a base-4 rather than base-2 encoding scheme, and many others.

## 4 Measuring entropy in language

We now have the necessary concepts, but we need a way to estimate how much information is transmitted via a particular artifact to a defined set of human minds. Language is made up of words which can be expressed in writing with characters, and characters are limited in number: there are only 26 letters in the English alphabet. Characters as they occur in writing can then be understood as a random variable. If we are using human minds as decoders, entropy per character can be calculated by

asking, “To what degree is a person able to predict the next letter given a sample string?” An optimal encoding scheme for transmission without noise would answer “not at all” (and hence bits per character would be high and uniform). But people are quite able to predict the next letter given the previous  $n-1$  letters of an  $n$ -gram in any natural language. This is due to redundancies built in to language to deal with noise in the transmission channel, as well as higher-level prior expectations about content (as in cultural inheritance; much of what is said is already known to the receiver with high probability). Here we will consider English, but it is expected that all natural languages have very similar properties in this regard (Behr 2006).

In 1951 Shannon published a method for empirically estimating the upper bound of entropy of written English, known as the “Shannon game”. The experiment asked subjects to guess the letters making up a sentence, where the set of symbols is the 26-letter English alphabet plus space, for 27 total symbols. Starting at the first letter, the subject makes a guess. The subject is given feedback on whether her guess is correct, and if it is correct she writes it in the blank and moves on to the next test item. If her guess is wrong, she tries again until it is right. The number of tries it takes to guess correctly is recorded.

(1)	T	H	E	R	E		I	S		N	O		R	E	V	E	R	S	E		O	N		A		M	O	T	O	R	C	Y	C	L	E		A				
(2)	1	1	1	5	1	1	2	1	1	2	1	1	1	5	1	1	7	1	1	1	2	1	3	2	1	2	2	7	1	1	1	1	4	1	1	1	1	3	1		
(1)	F	R	I	E	N	D		O	F		M	I	N	E		F	O	U	N	D		T	H	I	S		O	U	T												
(2)	8	6	1	3	1	1	1	1	1	1	1	1	1	1	6	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
(1)	R	A	T	H	E	R		D	R	A	M	A	T	I	C	A	L	L	Y		T	H	E		O	T	H	E		O	T	H	E		D	A	Y				
(2)	4	1	1	1	1	1	1	1	1	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	(9)

Figure 2: Example data from Shannon (1951). Rows labeled (1) are the target sentence, and rows labeled (2) show the number of guesses necessary to hit upon the correct letter for a given subject. Note the preponderance of ones.

This data allows the estimation of entropy per character of written English conditional on the model used (*i.e.*, on the particular reader’s mind). Shannon estimates  $\hat{q}_i^N$ , the probability that the subject requires  $i$  guesses to land upon the correct letter following a

sequence of  $N-1$  symbols, and shows the following is an upper bound for the entropy of printed English:

$$H(X) \leq - \sum_{i=1}^{27} \hat{q}_i^N \log(\hat{q}_i^N) \quad (\text{Equation 4})$$

More recent work has improved upon that technique. Using a paradigm that involved asking subjects to bet on subsequent characters, Cover and King (1978) estimated about 1.3 bits per character and subsequent estimates have coalesced around approximately 1-1.3 bits per character (Takahira 2016, Ren et al. 2019). An improved estimator such as Cover and King's together with an incentive-compatible task could be used for empirical measurements today.

One may rightly worry that a character-level model may not capture the deeper semantics underpinning our interest in it as a medium for the transmission of culture. After all, interest in the transmission of culture is generally rooted in the ideas that are transmitted<sup>2</sup>. For the sake of clarity I have used examples from character-level models here, though this general technique can be extended to semantically richer word-level models and even further to sentence fragments. There are a variety of enhancements that could allow for accurate estimation of entropy moving up through words and sentence fragments while keeping the task tractable for subjects, such as using language models to eliminate trivial questions (questions where the probability any given respondent is able to get the right answer in one try is quite high), and questions the subject has no hope of getting right. In addition, the sampling of n-grams from the target work would benefit from importance sampling. The specifics remain to be developed in future work.

---

<sup>2</sup> There is some indication that character-level models with long enough n-grams capture higher-level semantic features to some extent. CharRNN character-level models trained on a large English corpus vs. those trained on a large English corpus plus a particular target work do evince a difference in entropy measurements of ~0.2-0.4 bits per character when tested on n-grams (n=64 or 100) sampled from the target work. That these relatively crude models are able to discern a difference is somewhat surprising in itself (Cashman 2018, unpublished).

## 5 Applying entropy measurements to information flows through culture

If  $H(P(x))$  is the entropy measured with the treatment group (those who have read the target work) and  $H(Q(x))$  is the entropy measured with the control group (matched subjects who have not read the target work), then the simplest difference measure that could be used for RNI is subtraction: the difference between the estimates of entropy non-readers and readers produce, given a particular test set. This quantity, when multiplied by the number of symbols in the target artifact,  $n$ , is a measure of the number of bits of the target artifact stored in the reader's mind: RNI.

$$\text{RNI} = n(H(P(x)) - H(Q(x))) \quad (\text{Equation 5})$$

RNI is a representation of the net difference in what subjects in the control group failed to predict but which subjects in the treatment group successfully predicted, so it is a measure of the information from the target work that has been retained in the minds of the treatment group and is therefore capable of influencing behavior. The proportion of RNI will vary among different target artifacts, giving us a window into the extent to which a particular artifact is predictable by anyone regardless of exposure. For example, upon hearing a story that begins with a local youth leaving for parts unknown, many would guess that she will undergo trials and return wiser and stronger. Only a small proportion of a given written work is likely to be novel in the sense of being unpredictable by the naïve, while the remainder can be thought of as shared cultural inheritance because it can be predicted by any member of the culture defined in the experiment by the control group.

To illustrate with a character-level example, we might consider a group that has been quite culturally successful in recent centuries: Mormons. Given a control group (Americans), a treatment group (Mormon Americans), and a target work only the treatment group (presumably) has been directly exposed to (the Book of Mormon), we

can then estimate entropy of the target work in each group. The difference in entropy measurements between the two groups, the RNI, is driven by their minds' differing properties as decoders—differences in their language models themselves. RNI gives us a window into how much information from the target work has actually been transmitted to the treatment group's minds and is capable of influencing behavior. If we were to measure an entropy of 1.3 bits per character in the control group (non-Mormons) and 1.0 bits per character in the treatment group (Mormons), we might see an RNI difference measurement of  $\sim 0.3$  bits per character. This, multiplied by the number of characters in the Book of Mormon, is the number of bits of information from the Book of Mormon that is stored in Mormons' minds vs. those of non-Mormons (about 500Kb). The minimal task is very simple: subjects are given randomly selected n-grams from the target work with the last item covered, and the instruction to guess the missing character. The number of guesses necessary to get the right letter for the blank is recorded, and entropy is calculated from that data.

However, without randomizing assignment to treatment or not the experiment would be confounded by subjects' differing abilities at cloze completion tasks among other things. A difference-in-difference design would allow for meaningful measurements in the field without random assignment. If all subjects complete both items from the target work and items from a huge, diverse corpus we can implement subject-level controls for the influence on RNI of factors other than exposure to the target work. The procedure is simple: at the subject level, calculate  $H$  for the target work and  $H$  for the large corpus control. Take the difference between these two measures to arrive at a measurement of entropy in the target work controlling for the factors that would affect entropy measurements in the large corpus control but which would *not* affect entropy measurements in the target work. The set of things that may influence entropy measurements of the target work but *not* the control are quite small, and probably limited to exposure to the target work itself. Once this is done, one may calculate the difference between the treatment and control groups' entropy estimates to arrive at an RNI measure that controls for ability on cloze tasks, among other things.

## 6 Discussion

I have outlined a general procedure for measuring retained novel information given a control group who are not familiar with a particular book, and a treatment group who are. The ability to perform precise estimations of just how much information from a particular written cultural artifact is present in a living human mind may be novel, but what, really, are we measuring? And how can we use it?

We can consider several different scenarios in order to develop intuitions about what RNI is measuring. First, consider the situation where the target artifact is a book full of truly random gibberish. Entropy of the work is very high, but very little is going to be retained in the minds of humans who read it. If we were to print hundreds of copies and bury them in the desert, we would have a very high-entropy cache of cultural artifacts. Nobody would ever read those books though. For the purposes of investigating cultural evolution, we really only care about information that is in *living* minds and therefore capable of influencing behavior. Random Gibberish, Vols. 1-10 is very high entropy but very low RNI. We can compare this to the Book of Mormon, a religious text written by Joseph Smith in the 1820s, which is noted for its reuse of Abrahamic religious tropes and approximation of the style of the King James Bible. We might expect the Book of Mormon to have approximately average entropy per character for English, while having RNI lower than average. The Pāli Canon, a foundational Theravada Buddhist text, reflects its origins in an oral tradition by being extremely repetitive. It is also, perhaps necessarily, extremely long—it takes up approximately one bookcase. We would expect that a control group, non-readers, would measure below-average entropy (the text is very predictable, perhaps on a character-level model people are guessing whole words or sentences), and that the relative proportion of RNI given the treatment group would be small (there is little novel information to retain per symbol). Contrast these two books with something like a technical manual that outlines how to machine a specific part for an airplane. Though there are the usual regularities, redundancies, and other error-correcting properties of natural language, we would expect that a control group would result in approximately average entropy measurements (perhaps a little

high), while the treatment group would result in an RNI measurement considerably below average for a treatment group: much of the information in such an artifact is very predictable once you have been exposed to it (indeed, you have a model of the machined part in your mind). However, this information would be very difficult to predict *a priori*. There is also a class of target artifacts for which RNI as measured will appear to be very large, but in fact it is merely a few bits. Consider the example of a simple pseudorandom number generating algorithm: such an algorithm could be used to generate a sequence of letters, words, or even sentences with a regularity that only becomes clear once the reader has finished the work. If the reader has inferred the correct algorithm, then prediction of the next symbol from an n-gram becomes trivial—we would observe a very high entropy measurement in the control group and a very low measurement in the treatment group (perhaps zero), but we know that the information that was learned can in fact be expressed with relatively few bits. This sort of problem is not likely to arise in natural language texts, but there may be certain classes of ideas, or groups of ideas, that have something of this property.

## 6.1 Limitations

Perhaps the most important limitation to the method described is its reliance on the written word. Much of culture is transmitted outside of written language, for example via speech, facial expression, changes made to an environment, and observing then imitating others' actions. The proportion of information transmitted via culture that is transmitted via writing has increased exponentially in recent centuries, to be sure, but it is still only a fraction (and perhaps not the most important fraction) of the total flow of information via culture. Speech is amenable to these methods to some extent, since it can be transcribed.

Another concern might be that the method outlined is completely agnostic to content; it does not reflect *what*, though it does accurately reflect *how much*. However, it does depend on the selection of a target artifact from which samples can be drawn for



testing. If one is interested in measuring the flow of novel information from that artifact, then there is no problem. But there may be instances where the ideal measurement is only approximated by a particular work. One of the advantages of using this method to track information flows is that there is no need to make a definitive claim about what is important. More conventional methods might ask questions to test subjects' knowledge of a particular concept, event, vignette, or similar. This requires making a strong claim about what knowledge is important with respect to the subject being researched, namely choosing the questions. In some circumstances we may just trading this problem for the problem of choosing the target artifact well.

Finally, though I give a method for measuring how much information from a given target artifact is present in a living mind, we do not have the theoretical machinery necessary to link this to behavior. There may be a certain number of bits stored in a mind and capable of influencing behavior, but it is far from clear when it would be influential or how influential it would be. We might gain some traction from the assumption that information that is retained has passed through an algorithm optimized to only retain information that increases inclusive fitness. If we make this assumption then perhaps we can make the general claim that more information means more influence on behavior, though it would remain unclear exactly what and when.

## 6.2 Future directions

### 6.2.1 Evolutionary theory

This sort of procedure might also eventually be put to work bringing together models of genetic and cultural evolution. Measuring gene flow, selection among genes, etc. is relatively straightforward with modern technology, but models of cultural evolution have thus far lacked a principled, general-purpose means of empirically measuring the flow of ideas through culture. Information-theoretic measures of cultural success might

provide the basis for that interaction, helping open the way to a more general dual-inheritance theory. Krakauer et al. (2020) describe a method of defining individuals on which selection happens, historically a key puzzle for those attempting to analogize culture to genes (see Chvaja 2020 for a discussion). This sort of principled definition of an individual might allow for the definition of units of culture on which selection is happening. Though it is not clear why one bit of DNA might be comparable to one bit of RNI, it's possible that being able to treat both culture and DNA (a base-4 encoding scheme) in terms of information may be useful in the future as well. Finally, we are on the cusp of having the practical ability to edit our genomes directly. This will allow the transmission of genetic information via cultural means. At first bandwidth will be very small, but it is reasonable to expect the technological capacity will grow. Popular genes may spread like popular books; selection mechanisms and network effects that were once the sole province of clickbait, cat memes, and earworms will soon be applicable to our DNA. There may be a role for the sort of analysis presented here in investigating these changes.

### 6.2.2 Measurements across multiple different works

We need not limit ourselves to a single target work, a treatment, and a control. We can design measurement schemes that look at two or more works. Consider investigating the relationship between two target works, A and B. The treatment group has been exposed to A, while we have taken our test set sampled from B. By testing subjects on material from B after having been exposed to A, in the case of the treatment group, or not, in the case of controls, we can investigate the extent to which A and B are similar relative to each other. We could ask: how much does knowing the Bible mean you already know the Koran? Treatment groups could be exposed to a wide variety of

target works and then all test with samples from B, or we could vary the test sets to come from a wide variety of works.

### 6.2.3 Flow through social networks

We might also deploy this procedure while measuring information flowing from a certain target work through social networks. This would allow the measurement of information flow in a way that is agnostic to content. Attempting to measure the flow of information, beliefs, etc. at present requires knowing which questions to ask. If we can be satisfied with a single work as a target artifact, then we can deploy the framework outlined to trace the information from that work as it is transmitted from person to person via a series of measurements across time and distributed in social networks. This sort of procedure would be amenable to measuring variation within levels of familiarity beyond just a binary familiar/not familiar measure.

### 6.2.4 Education

Principled methods of measuring information flow also lend themselves to the study of teaching and learning. In educational settings, we might compare different teaching methods or materials via measuring RNI pre- and post-treatment, given a target artifact of educational value. For instance, three matched groups of students might be taught using the same history textbook, but each group is randomized to a different teaching method. We can then measure RNI from the history textbook at the end of the class to assess how much information was retained. This measurement scheme could shed light on how much pupils are learning from a target work, and critically without the assumptions about what is important inherent in traditional testing regimes: a student may not be able to recall the year of the Battle of Hastings, but the same student may have in fact learned many other things from the book and instruction.

## 6.2.5 Charting cultural tropes and archetypes

We might also chart how story archetypes are distributed across the world. One way this could be accomplished is to carefully select target stories that are then coded as particular archetypes and used the standard procedure, testing for RNI within many different cultures relative to a single control. Another related technique might use specific, well-known works. We might investigate how the predictability of the Bible or the Koran varies across cultures, giving us a new window into the influence the information in those books has across the globe. Given that these are long and varied texts, it would also be possible to partition them and look at the RNI for subsections such as the Old Testament versus the New Testament.

## 6.2.6 Music

The examples above all use written language to comprise the set of symbols, but these methods can be applied without change to music as well. Music can be coded as a set of symbols (notes on a staff or scientific pitch notation), and we can measure expectation in exactly the same way. There is a small literature on expectation in music (see Large & Kim 2019 for a review) including some that incorporates Shannon's methods of estimating entropy (Manzara et al. 1992, Pearce & Wiggins 2012). We could use a very similar technique as for language to investigate the distribution and spread of musical archetypes and tropes throughout the world.

## 7 References

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

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

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

Chvaja, R. (2020). Why Did Memetics Fail? Comparative Case Study. *Perspectives on Science*, 28(4), 542–570. [https://doi.org/10.1162/posc\\_a\\_00350](https://doi.org/10.1162/posc_a_00350)

Cover, T., & King, R. (1978). A convergent gambling estimate of the entropy of English. *IEEE Transactions on Information Theory*, 24(4), 413–421. <https://doi.org/10/bfhsp5>

Derex, M., Bonnefon, J.-F., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature Human Behaviour*, 3(5), 446–452. <https://doi.org/10.1038/s41562-019-0567-9>

Fisher, S. E., & Ridley, M. (2013). Culture, Genes, and the Human Revolution. *Science*, 340(6135), 929–930. <https://doi.org/10.1126/science.1236171>

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., ... Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3), 369–380. <https://doi.org/10.1038/s41593-022-01026-4>

Henrich, J. (2004). Demography and Cultural Evolution: How Adaptive Cultural Processes can Produce Maladaptive Losses: The Tasmanian Case. *American Antiquity*, 69(2), 197. <https://doi.org/10.2307/4128416>

Krakauer, D., Bertschinger, N., Olbrich, E., Flack, J. C., & Ay, N. (2020). The information theory of individuality. *Theory in Biosciences*, 139(2), 209–223. <https://doi.org/10.1007/s12064-020-00313-7>

Kuhl, B. A., & Chun, M. (2014). *Memory and Attention* (A. C. (Kia) Nobre & S. Kastner,

Eds.; Vol. 1). Oxford University Press.

<https://doi.org/10.1093/oxfordhb/9780199675111.013.034>

Large, E. W., & Kim, J. C. (2019). Musical expectancy. In *Foundations in music psychology: Theory and research* (pp. 221–263). The MIT Press.

Manzara, L. C., Witten, I. H., & James, M. (1992). On the Entropy of Music: An Experiment with Bach Chorale Melodies. *Leonardo Music Journal*, 2(1), 81–88.

McElreath, R., & Henrich, J. (2007). *Modelling cultural evolution*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198568308.013.0039>

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

Muthukrishna, M., Shulman, B. W., Vasilescu, V., & Henrich, J. (2014). Sociality influences cultural complexity. *Proceedings of the Royal Society B: Biological Sciences*, 281(1774), 20132511. <https://doi.org/10.1098/rspb.2013.2511>

O'Brien, M. J., & Laland, K. N. (2012). Genes, Culture, and Agriculture: An Example of Human Niche Construction. *Current Anthropology*, 53(4), 434–470. <https://doi.org/10.1086/666585>

Pearce, M. T., & Wiggins, G. A. (2012). Auditory Expectation: The Information Dynamics of Music Perception and Cognition. *Topics in Cognitive Science*, 4(4), 625–652. <https://doi.org/10.1111/j.1756-8765.2012.01214.x>

Ren, G., Takahashi, S., & Tanaka-Ishii, K. (2019). Entropy Rate Estimation for English via a Large Cognitive Experiment Using Mechanical Turk. *Entropy*, 21(12), 1201. <https://doi.org/10.3390/e21121201>

Shannon, C. (1951). Prediction and Entropy of Printed English. *The Bell System Technical Journal*, 50–64. <https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

Takahira, R., & Tanaka-Ishii, K. (2016). Upper Bound of Entropy Rate Revisited; A New Extrapolation of Compressed Large-Scale Corpora. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity, Osaka, Japan*, 213–221.