A conceptual and empirical framework for the study of online cultural evolution

Pietro Nickl[1,2,3*] **and Mehdi Moussaïd**[1,4*]

[1] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany.

[2] Humboldt-Universität zu Berlin, Faculty of Life Sciences, 10099 Berlin, Germany

[3] Max Planck School of Cognition, Stephanstrasse 1a, Leipzig, Germany

[4] School of Collective Intelligence, Mohammed VI Polytechnic University, Rabat, Morocco

*Corresponding author: nickl@mpib-berlin.mpg.de

Abstract

More and more cultural content is being produced and transmitted online while increasing competition for limited attention likely intensifies existing selective pressures on content. Therefore, it becomes important to understand the driving forces of cultural evolution in this environment. In this paper, we present an agent-based model to capture cultural transmission dynamics online, considering characteristics of digital online environments (low-cost production, high-fidelity transmission, high individual reach) that set them apart from offline scenarios. As a case study, we report preliminary findings of a rise in clickbait style in titles across online platforms, and use our model to explain which mechanisms could give rise to this pattern. Generally, the model serves to formulate predictions against which to compare observational data and we show how online trends can be interpreted in terms of cultural evolution. This work outlines an interdisciplinary research agenda for studying and understanding cultural evolution online, with possible implications for regulating online content and designing digital environments.

Keywords: cultural evolution, contagion, social influence, content bias, clickbait, agent-based model

Introduction

Online environments constitute a novel ecosystem for culture to evolve in. After the inventions of writing and the printing press, the digital age brought high-fidelity transmission and production of content at low cost, enabling both the accumulation and easy access of available knowledge (e.g., Wikipedia), as well as problematic phenomena such an abundance of clickbait content and fake news. The internet is therefore characterized by information proliferation [1] and an attention economy [2]: content abounds, creating fierce competition for limited attention. To understand how digital online environments affect the trajectory of cultural change, we present a conceptual and empirical framework to study online cultural evolution.

Mechanisms of cultural evolution

We adopt the definition of culture as socially transmitted information [3]. The units of transmission – cultural traits – diffuse, recombine and give rise to traditions [4]. Tools, norms, stories are all examples of cultural traits. Cultural evolution models cultural change as an evolutionary process, generally assuming a Darwinian process in analogy to biological evolution [3] (for a complementary view see [5,6]). Cultural traits are copied with variation, and are subject to selective pressures (e.g. memory limitations), giving rise to an evolutionary process. This may be a cumulative process, where every trait presents some form of improvement compared to its predecessor [7]. For example, a story may become more structured and memorable through many retellings, or online content could evolve to become more attention-grabbing. The main differences to genetic evolution are (1) transmission does not have to be vertical (parent to offspring) but can take place e.g. from peer to peer, and (2) variation need not stem from random mutation, but also *guided* variation [8]. Guided variation is not a population-level process of selection among faithfully copied traits. Instead, it reflects a tendency of individual learners to change a learnt trait in a non-random way: For example, in copying a knife, learners can decide to make the blade sharper, or a negative story could be retold as even more negative [6]. Whereas in genetic transmission, random mutations are rare and often maladaptive, some take adaptive guided variation as the norm in cultural transmission [9,10].

In addition, the transmission of cultural content exhibits specific biases, as social learners are selective about when, who and what to copy for an overview of social learning strategies, see [11]. More memorable stories may be more likely to be retold (the transmission is biased towards memorability), and a story may become more memorable as it is being passed on (the *mutation* is biased towards memorability). Here, memorability is a so-called *content bias* (or direct bias), where the trait derives its success from an intrinsic property. In contrast, c*ontext biases* (or indirect biases) result from a social learning rule such as conformity bias (a disproportionate tendency to copy a majority trait; [8]), or prestige bias (copy from models with high social status, e.g. [12]). The process of cultural transmission can be broken down into three distinct stages: (1) choose-to-receive, (2) encode and retrieve, and (3) choose-to-transmit [13]. That is, we may choose to listen to a story (1), and if we remember it (2) we can choose to retell it to our peers (3). Note the crucial role of memory here, constituting a bottleneck for transmission and the main locus of transformation in offline cultural evolution.

Online vs. offline cultural evolution

Though traditionally, the mathematical models and experimental paradigms to study cultural evolution link to anthropological data from field studies or archaeological records, the same tools and concepts apply to the online realm [14,15]. Nevertheless, online environments exhibit specific features that importantly distinguish them from offline environments. Following Acerbi, we take a) low-cost production, b) high-fidelity transmission, and c) high reach/availability as representative characteristics of online transmission [15], explicitly adding d) sharing cascades and e) decentralized production of items. First, online transmission is near-instantaneous, while (re)production costs are low and fidelity is high. While information will be distorted along an oral transmission chain, the same tweet or video can be spread in a deep sharing cascade and, in the extreme case, "go viral", i.e. spread very far. Second, Internet users have far reach and high availability. At the time of writing this article, the official Barack Obama Twitter account had 133.3 million followers, an example of very wide reach. Unlike learners in offline contexts, internet users are not restricted to their local community in tapping a large body of cultural information.

To our knowledge, no offline scenario combines all the above characteristics of online environments. Consider **gossip cascades**: Millenia before social media posts, a piece of gossip could go viral, being retold and retold. However, individuals retelling a story from memory introduce systematic variations at each transmission (e.g., [16]), whereas online sharing cascades can perfectly preserve the original content. Another offline scenario is **tool emulation** (c.f. [17]): A well-manufactured knife is being passed around and those seeing this product can copy its overall design, material or sharpness. This captures the process of emulation: individuals take an end product as a model for their own productions, similar to digital items on Youtube or Twitter. However, the same knife will not reach as many sets of eyes as a YouTube video might. High reach is a characteristic of traditional **broadcast media**. TV, radio, and newspapers can disseminate one and the same message to the public in a one-step transmission. However, mass media are centralized, pertaining to a nation, company or individual who can afford the means of production, meaning most people cannot become producers. **Mass-produced items** such as leaflets or books are a similar case: While these productions may influence future productions (a novelist has read previous novels), again, producers are scarce. Table 1 summarizes the specific features of the online environment.

	offline scenarios				online
	gossip cascade	tool emulation	broadcast media (TV, radio)	mass-produced items	YouTube videos, tweets etc.
low-cost production	Yes	depending on tool complexity	No	No	Yes
high-fidelity transmission	No	not when copied	Yes	Yes	Yes
high individual reach	No	No	Yes	Yes	Yes
deep sharing cascades	Yes	Yes	No	No	Yes
decentralized production	Yes	Yes	No	No	Yes

Table 1: Summary of the dynamics distinguishing cultural transmission in online contexts from offline scenarios. The rows correspond to five characteristics of online environments, which do not combine in any of the offline scenarios we consider. For example, the transmission of gossip offline is low-cost, decentralized and may exhibit deep sharing cascades. At the same time, transmission does not ensure high fidelity (gossip changes as it is being passed on), and individual reach is much narrower offline (individuals do not pass on gossip to hundreds or thousands of people).

Both human psychology and the online environment exert pressure on cultural content online – and the phenomena we observe, such as false news and clickbait, may be adaptations to these conditions.
On the one hand, cultural success can be a function of human psychology: For example, emotional, negative, threat-relating, disgusting or social content may enjoy a transmission advantage (e.g., [16,13,18–22]). Coupled with information proliferation and the fierce competition of content online, cognitive content biases may explain the spread of false news [23]: As false news are not constrained by truth, they can be engineered to maximize psychological appeal [24,25].
On the other hand, the structure of online environments also shapes online culture. Clickbait-style headlines exploit human psychology, specifically curiosity [26], in order to maximize clicks. Still, their success is determined by algorithms and which metrics they track: Clickbait is only effective, if e.g. a newsfeed algorithm favors most-clicked items over other measures of success. This shows how cognitive biases and algorithmic selection work in tandem.

To investigate the interplay of different mechanisms underlying online culture, we first present a conceptual framework. The goal is to formally describe the key components involved in the diffusion and the evolution of cultural traits in digital environments, using agent-based simulations. We then describe how empirical data informs the model's assumptions, using the example of clickbait style in online titles. We report preliminary findings suggesting a rise of clickbait features in online titles over the last decade. Finally, we use our model to propose mechanistic explanations for this trend. We end by discussing some implications and future steps.

Conceptual framework

Agent-based models start with simple assumptions about individual behavior to simulate population-level outcomes, often resulting in unforeseen emergent phenomena [27]. The general architecture of the following model is inspired by Axelrod's model of the dissemination of culture [28] as well as classic models of contagion and social influence [29,30]. It builds on the features of online cultural evolution introduced in Table 1 and consists of three main components: (1) how digital items are created by the agents, (2) how digital items spread in the population, and (3) how individuals are influenced by the existing items.

Item creation

We consider a large population of individuals who continuously create *digital items*. A digital item could be a video on Youtube, a tweet on Twitter, or a photo on Instagram. Each digital item possesses specific attributes. For example, a Youtube video could have a certain duration, a certain title structure, or some specific graphical codes in its thumbnail. These attributes are the cultural traits we aim to track and understand.

For the sake of simplicity, we assume that items are characterized by one single attribute x that takes a value in the range $[0\ 1]$. Of course, a more realistic approach would need to consider not only one but a set of q attributes $\{x_1\ ,\ x_2,\ ...\ x_q\}$ corresponding to the different dimensions characterizing the item (e.g., a video's duration, title length, content topic, etc.). However, we choose here to start with the simplest possible implementation to facilitate the study of the collective dynamics that will arise.

The value of the attribute x_j of an item j is determined by the preference α_i of the individual i who created it. As a starting point, we define a simple equation:

$$x_j = \alpha_i + \varepsilon \qquad (1)$$

where i is the individual who created the item j and ε is a random variable. The term ε is common to all new creations, as all individuals have the same probability of introducing variation. This equation simply states that individuals tend to create items that reflect their own preference, e.g., individuals who like short videos on Youtube will tend to create short videos as well, with some variability. To facilitate traceability in the simulations, the value of x_j is constrained in the interval $[0\ 1]$.

Item diffusion

One key feature of online environments is the fact that digital content can spread easily from one person to another: tweets can be retweeted, Youtube videos can be shared with friends and relatives, photos circulate on Instagram, etc. In our framework, this is expressed by a so-called *diffusion function* that determines the subset of individuals in the population who will be exposed to each item.

The diffusion function can take various forms. For instance, an *epidemic* approach would describe the diffusion of an item as a contagion process similar to SIR models (see Mesoudi, 2021, for an overview and examples of similar models): the item creator shares the item with their "neighbors", who would themselves share it with their own neighbors, and so forth. This results in a diffusion cascade where the item spreads progressively from person to person in the population [31–33]. A simple implementation of the epidemic approach requires a single parameter p delineating the probability that an individual exposed to the item would share it in turn with their neighbors. Put simply, after an agent is exposed to a given item, they subsequently share it with their neighbors with a probability p (only once, if at all). The process then repeats with the new agents who received the item, and so on until the propagation stops. Hence, a low value of p would tend to create small, local cascades around the item creator, whereas a large value of p would produce large, global cascades that can reach a large fraction of the individuals in the population (**see figure 1**).

The diffusion process requires a formal description of the social network connecting the individuals, to determine who are the "neighbors" of a given individual. For that, various network structures can be used such as a random network, where each individual is connected to n randomly selected others, or a scale-free network characterized by a power-law distribution of node degrees (see, e.g. [34]). For simplicity, we will first assume that individuals are distributed over an N x N squared grid with periodic boundary conditions (for a total of N^2 individuals in the population). That is, each individual is connected to the 8 others located in the neighboring cells of the grid. The periodic boundary conditions indicate that individuals located at the periphery of the grid are neighbors with those located on the opposite side – a common way to avoid detrimental border effects.

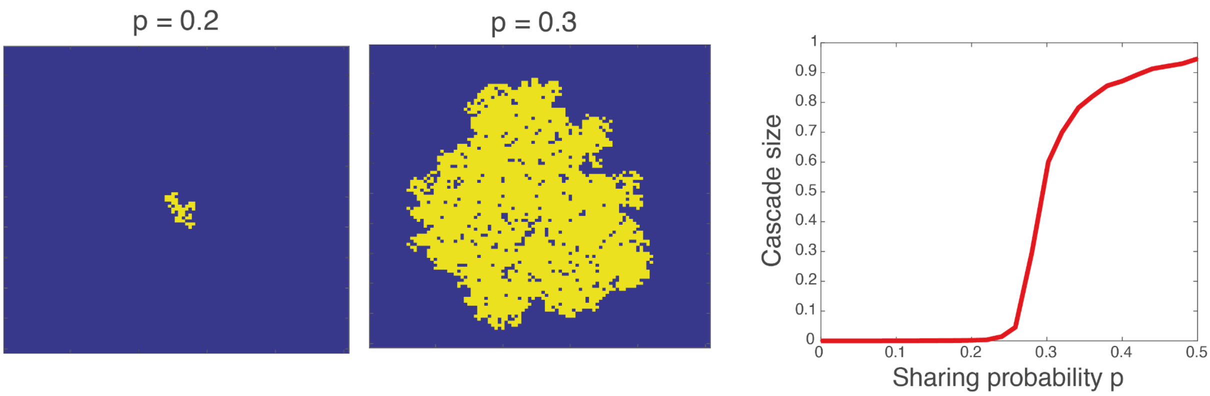

Figure 1: (*Left*) Examples of small and large diffusion cascades produced by the epidemic approach. In this simulation, 10,000 individuals are distributed in a 100 x 100 squared grid. The item creator is located in the middle of the grid and has a probability p of sharing the item with each of the 8 neighbors. Each "infected" neighbor (in yellow) undergoes the same process until the diffusion

stops. (*Middle*) The average cascade size expressed as the fraction of the population that has been exposed to the item. It increases non-linearly with the sharing probability p (*Right*).

In the simplest implementation, the sharing probability p is constant and the same for all individuals and all items. However, it is also interesting to consider that some features of the item may modulate the sharing probability, that is, p could be a function of x_j. The functional relationship $p = f(x)$ could for instance take into account if short videos on Youtube may be shared more often than longer videos, or if sensationalistic tweets on Twitter may be retweeted more often than less exciting ones. Furthermore, the sharing probability could additionally be modulated by the user's preference, formally $p = f(x, \alpha)$. An individual could, in fact, be more likely to share an item if that item has an attribute value that matches their preference α.

In contrast to the epidemic approach, the diffusion function can also take the form of a *broadcast* approach. In this case, a certain subset of the population is immediately exposed to the item, irrespective of the underlying network. This could apply, for instance, if the item is displayed on mass media (e.g., TV or newspapers), or when a recommendation algorithm is designed to expose certain users to certain items. Here again, numerous variations are conceivable. Recommendation systems, for instance, tend to expose individuals to items they may like, thus creating a functional relationship between a user's preference α, the item's attribute x, and the likelihood of exposing that user to that item. The epidemic and the broadcast approaches are not mutually exclusive and typically apply simultaneously. The two mechanisms could even amplify one another, for instance, when mass media report about an item that has triggered a particularly large diffusion cascade.

For the sake of simplicity, however, we will limit ourselves to contagion processes in this manuscript, assuming at first a constant and homogenous sharing probability p in the population, and a regular grid network. Later in the paper, we will show how functional variations can modulate the collective outcome.

Social influence

So far, the framework describes how individuals create digital items and how these items spread in the population. The third and final component introduces a key mechanism from which the collective dynamics will emerge: the *influence function*. For that, we assume that the diffusion of the item affects – to some extent – the preference of the individuals who are exposed to it. For instance, using certain photo filters or enhancing effects on Instagram could increase the likelihood that those who saw that photo would use the same effects in their subsequent posts. This mechanism is crucial because items that spread farther will influence the preference of more people, thus increasing the likelihood that similar items would be created in the future. This feedback loop could then gradually lead to the emergence of a dominant cultural trait.

For the influence function, we use a common approach that is typically used in the social influence literature [35] and for modeling belief updating in simulations of opinion dynamics [29,30]. We assume that an individual i exposed to an item j would update their preference $\alpha_i(t)$ at time t in the following way:

$$\alpha_i(t) = \alpha_i(t-1) + s * (x_j - \alpha_i(t-1)) \quad (2)$$

where x_j is the attribute value of the item j and s is a social influence parameter. With $s = 0$ the individual is indifferent to the item's attribute, whereas with $s = 1$ the individual fully adopts the feature of the item. In other contexts, the social influence parameter is often measured around $s = 0.3$ [35–37].

Here again, numerous variations could be considered. The value of s could depend on the matching between α_i and x_j, or even be negative in the case where the individual dislikes the item's attribute. An item with a clickbait title could spread widely in the population but have little influence on individual preferences because of the psychological disappointment it generates. Conversely, clickbait could as well be heavily copied if individuals noticed its successful diffusion and tried to exploit this feature for their own items, similar to the social learning strategies copy-success and payoff bias [38,11]. The latter assumption would imply a functional relationship between the social influence factor s and the cascade size determined by the diffusion function.

In summary, our modeling framework describes the general processes by which digital items (1) are created by the individuals, (2) spread in the population, and (3) affect individual preferences. The nature of these three components has the potential to create specific collective dynamics, which we now explore in simulations.

Exploratory simulations

We run a first series of simulations to explore the collective dynamics predicted by the model[1]. For the simulations, N=2,500 agents are located on a 50 x 50 regular grid and initialized with a random preference α picked in the interval $[0\ 1]$. For every simulated time step t, one randomly selected agent creates a new item using equation (1). This item then spreads in the population using the epidemic diffusion function with a sharing probability p. Once the diffusion has ended, all the agents who have been exposed to the item update their preference α using equation (2). The simulation continues for a total of T=10,000 time steps. This implementation, therefore, contains three parameters: the random variable ε that is drawn from a normal distribution with mean $\varepsilon_m = 0$ and a fixed standard deviation $\varepsilon_s = 0.05$, plus the sharing probability p and the strength of social influence s which will be varied systematically.

F**igure 2** illustrates the outcome of one simulation run for $p = 0.1$ and $s = 0.75$ at three different time steps. It appears that the initially random distribution of agent preferences tends to cluster into regions where individuals exhibit similar preferences. This collective pattern results from the diffusion of digital items within a certain social distance, causing individuals in that area to converge towards a common preference. At the scale of the population, however, different cultures continue to coexist because the same clustering process occurs simultaneously in different regions of the social space.

[1] The simulation code implemented for Matlab R2018a is available at https://arc-git.mpib-berlin.mpg.de/nickl/cultural_evolution_online

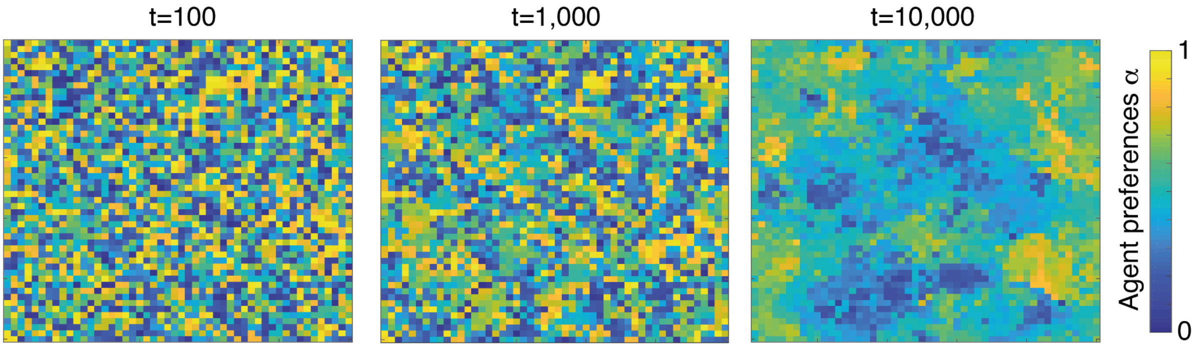

Figure 2: Evolution of one cultural trait in a population of 2,500 individuals after t=100, 1,000, and 10,000 simulation steps. The model parameters are $\varepsilon_s = 0.05$, $p = 0.1$ and $s = 0.75$.

To better understand this pattern, we systematically varied p in the interval $[0\ 0.3]$ with steps of 0.025, and s in the interval $[0\ 1]$ with steps of 0.05. For the analysis, we introduce four coefficients characterizing the emerging cultural structure: (1) the diversity of preferences D, (2) the instability of preferences I, (3) the neighborhood similarity N, and (4) the polarization of preferences P.

The diversity of preferences $D(t)$ at a certain time t is defined as the standard deviation of all the agents' preferences α at that time:

$$D(t) = std(\alpha(t))$$

A large value indicates that multiple preferences coexist in the population (as in **figure 2**) and a small one indicates that all the agents tend to share similar preferences.

The instability of preferences $I(t)$ indicates whether the agents' preferences are constantly changing over the simulation time or whether a stable state has been reached. It is calculated by measuring how much on average the preferences α have changed within a certain time interval δ_t:

$$I(t) = avg_i(\left|\alpha_i(t - \delta_t) - \alpha_i(t + \delta_t)\right|)$$

where avg_i denotes an average over all agents i. In the following, the time interval δ_t is set to $\delta_t = 1,000$ simulation steps.

The neighborhood similarity, is first calculated locally around an agent i by measuring the difference between that agent's preference α_i and the preferences α_j of all the neighbors j:

$$n_i(t) = 1 - avg_j(\left|\alpha_i(t) - \alpha_j(t)\right|)$$

The overall neighborhood similarity coefficient $N(t)$ is then calculated by averaging $n_i(t)$ over all the agents:

$$N(t) = avg_i(n_i(t))$$

Finally, we introduce the polarization coefficient $P(t)$ which indicates how much the population is divided between opposed preferences. We define $P(t)$ as the product of diversity and neighborhood similarity:

$$P(t) = D(t) \cdot N(t)$$

Hence, a large value of P (i.e. for P approaching one) indicates a strong clustering pattern (i.e., neighboring agents have similar preferences) coupled with a high diversity of preferences at the population scale – a signature of group polarization.

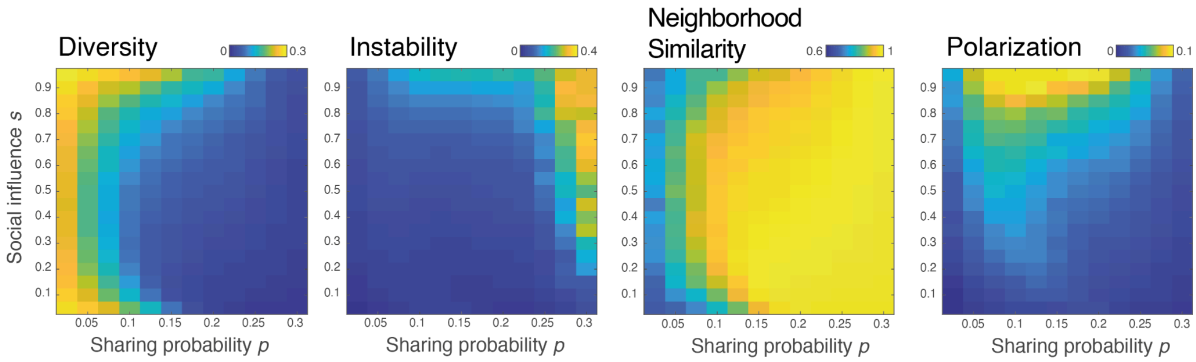

Figure 3: The diversity, instability, neighborhood similarity, and polarization coefficients in a population of 2,500 individuals, after 10,000 simulation steps, when varying the sharing probability p and the social influence s.

As shown in **figure 3**, the four coefficients vary considerably as p and s change. With a very low sharing probability (i.e. $p < 0.1$), items tend to spread only locally to the first few neighbors. As a result, the initially random preferences of the agents remain rather unstructured, which is reflected in a high diversity and low neighborhood similarity values in the population. As the low instability shows in this case, the preferences do not change much in the course of the simulation. When p increases, however, the diffusion gradually goes from local to global and items tend to spread farther and farther. Consequently, the diversity of preferences drops because individuals are increasingly influenced by the preference of others. In addition, clustering patterns emerge. Above a certain threshold (i.e. for $p > 0.25$), the diffusion is so strong that items tend to spread to the entire population: the diversity is close to zero, the clustering pattern slowly vanishes and the preferences become very unstable (because every new item triggers a large-scale wave of changes). Note that values of $p > 0.3$ almost systematically yield a global diffusion (as shown in **figure 1**) and have not been simulated.

A strong polarization (i.e. a combination of high similarity in local neighborhoods and high diversity) is only observed in a particular region of the parameter space: when social influence is strong and for intermediate sharing probabilities. In this case, clusters of agents with high α coexist with clusters of agents with low α. The impact of the social influence factor is less marked. A low value of s tends to increase diversity and reduce clustering because individuals exert little influence on one another: collective patterns hardly emerge.

On the contrary, a high value of s is also associated with high diversity, but this time together with a marked clustering: This is where the polarization is maximized.

The above simulations show how collective patterns can emerge from local interaction rules between simulated agents. In the following, we collect empirical data and use our model to explore candidate mechanisms that can explain the observed trends.

Collecting online data

Observational data can be collected from various online platforms. So far, a lot of academic work has focused on Twitter, informing research on collective dynamics, including opinion dynamics [39], social network structures [40,41], diffusion patterns [31,42–44] and cultural evolution [45]. In this paper, we focus on titles, which are not a platform-specific phenomenon, though titles are important for sharing content on social media, and investigate a less-researched platform, namely YouTube [46,47] for initial descriptions of the YouTube ecosystem see [48,49]. To illustrate the methodology, we track single cultural traits (corresponding to the attribute x in our modeling framework) on a sample of items collected at two time points (in 2010 and 2020) from two different platforms: YouTube and the New York Times. In particular, we explore a potential increase in clickbait style as a case study of cultural evolution online.

Clickbait is an online phenomenon that has received a lot of public attention [26,50–54]. While generally decried as annoying and misleading, it seems to be a common strategy to present content online, embraced increasingly by traditional media outlets [55] and even scientific journals [56]. Therefore, we take clickbait-style as a formal feature of a headline, orthogonal to its content (e.g. its truth content, being "fake news" or not). For this paper, we take three easily identifiable characteristics and track their diffusion in online contexts: title length, use of wh-words and demonstratives.

First, headline length is an indicator of clickbait style: In contrast to the succinct headlines of traditional print media (e.g., “Natural disaster in Country X, 300 dead”), clickbait headlines are written in a conversational style featuring often featuring entire sentences (e.g., “This woman survived a natural disaster. How she did it may surprise you”), and hence containing more words [51]. Second, wh-words (who, what, why etc.) and demonstratives (this, those) are also common characteristics of clickbait headlines (e.g., “This is why ... ”). Referring to something in the body of the text, they act as a forward reference [50], which serves as a device to introduce a curiosity gap [57].
A good example is introduced by Kuiken et al. [53] who examined how editors rewrote news headlines for an online context to increase clicks (an instance of real-life guided variation in this domain; for the potential of guided variation in a fake news context, see [25]. Rewritten headlines were indeed longer (both more characters and more words) and more likely to contain wh-words and demonstratives. Interestingly, the authors did not find that headlines containing more characters were clicked more often, but the use of wh-words and demonstratives lead to significantly more clicks. In contrast, Robertson et al. [22] find that more words in the headline increase clicks. Increasing headline length may by itself be an ineffective strategy, though a behavior does not need to be effective to spread if it has intuitive appeal [58]. For a discussion of the effectiveness of clickbait strategies, see

[52]. In the following, we take headline length, as well as the use of wh-words and demonstratives as theoretically motivated yet straightforward variables to track across platforms and time.

The first platform used for data collection is Youtube. Through an official API, Youtube allows data collection to look back at video attributes over nearly two decades. Precious information about the video items is available, such as the topic each video belongs to (e.g., music, gaming, sports, etc…), and some attributes such as the title, description, upload time, view, like and comment count or its thumbnail image. For this exploratory data collection, however, we only collected English videos posted in 2010 and 2020, sampled based on their content category, and measured the number of words in the video titles, the presence of a wh-word, and the presence of a demonstrative..

The second platform is a newspaper outlet. For this, we collected data from the New York Times database via its official API. While traditional newspapers used to be sold as packages of information, single online news articles could go viral on social media. As such, they compete for the public's attention with other digital items and are thus subject to similar selective pressures. It is therefore possible that online newspaper headlines would undergo a similar evolution as the titles observed on other platforms.

For comparison, we use headlines from the soft-news outlet *Upworthy*, which became notorious for its clickbait titles and frequently appears in the literature on clickbait [26,50,51,59]. We used the data available in the Upworthy Research Archive [60], namely the exploratory portion of the full dataset used e.g. by Robertson et al. [22].

For this illustrative data analysis, we collected the titles of 112,439 and 55,490 news articles from the New York Times published in 2010 and 2020, respectively, and 23,213 and 21,782 videos posted on Youtube in the same period. **Figure 4** shows the mean number of words in the items' titles, the percentage of titles containing a wh-word and the percentage of titles containing a demonstrative on these two digital platforms. A small but significant trend is visible: in both environments, item titles have a tendency to become longer (Mann-Whitney-U-test for difference between 2010 and 2020 samples: for the New York Times corpus U=4701452572, p<0.001, common language effect size = 0.754; for the YouTube corpus U=305553600, p<0.001, common language effect size=0.604). Similarly, by 2020, the percentage of headlines containing a demonstrative has increased on both platforms, with a larger effect for the New York Times: X^2 (1, N = 167929) = 2741.12, p < .01, Cramer's phi = .13 (for New York Times headlines), compared to X^2 (1, N = 44995) = 77, p < .01, Cramer's phi = .04 (YouTube titles). For the percentage of headlines containing a wh-word, we find a similar pattern for the New York Times X^2 (1, N = 167929) = 6658.07, p < .01, Cramer's phi = .2 and no effect for the YouTube titles.

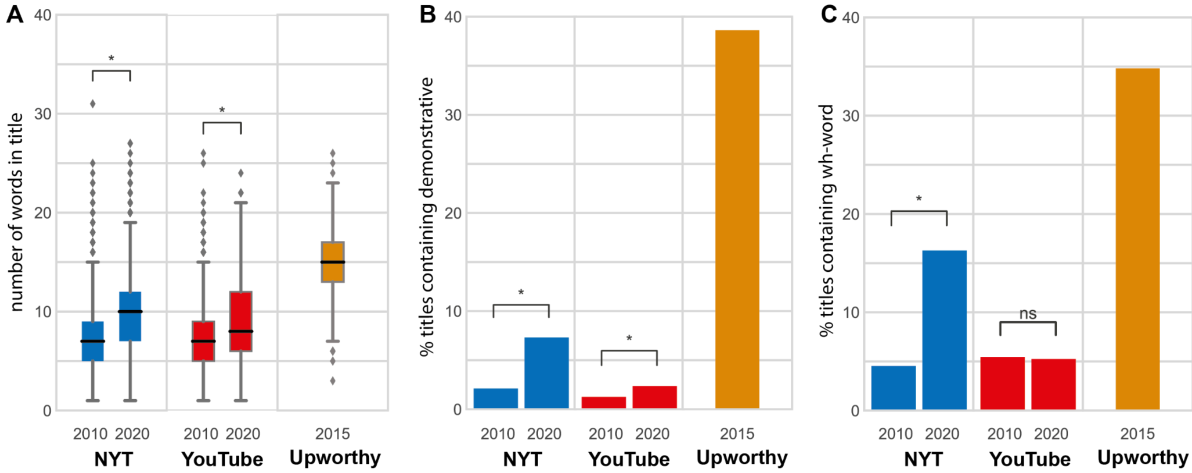

Figure 4: Number of words in the titles at different years across platforms (A), percentage of titles containing a demonstrative at different years across platforms (B), and percentage of titles containing a wh-word compared between different years across platforms (C). The clickbait outlet Upworthy scores high on these three metrics, the other two platforms show a general trend in this direction.

As a reference point, **figure 4** also displays the same item attributes for a sample of 1.626 titles posted on *Upworthy* in 2015, showing that these title attributes are evolving in a similar way across platforms, possibly resulting from the trend introduced by *Upworthy*.
It is worth noting that the dictionary-based approach we use here is inherently limited: It simply records if a certain string occurs, without recognising its function. For example, we do not know if a particular wh-word introduces a question or another construction (e.g., "this is why/how…" etc.), similarly, the word "that" has more uses beyond being a demonstrative. Determining its specific function in context would require more in-depth syntactic processing.
Although very preliminary, this analysis shows that it is in principle possible to collect relevant online data and to interpret it in the scope of cultural evolution. In the following, we use our model to explore possible mechanisms to account for this observed empirical trend.

Additional model simulations

Our data suggest that titles of digital items have undergone a systematic change in the past decade. Using the formalism outlined in our model, this observation corresponds to an increase in the attribute x of created items over simulation rounds. How can the model replicate this trend in simulations? Interestingly, three potential mechanisms could explain this collective pattern, each of which is associated with one of the three building blocks of the model.

First, it could be the result of a bias in the **item creation process**. This mechanism simply suggests that agents may find it easier to create items with a larger attribute value x (cf. Model 2b: Biased Mutation in [61]; and chapter 2 in [62]). Assuming x represents the title length, this bias could be due to the fact that longer titles are easier to produce, or that

shorter ones are increasingly taken by others. We tested this mechanism in simulations by introducing a biased noise term in equation (1), replacing the mean value $\varepsilon_m = 0$ by $\varepsilon_m = k$ where $k = 0.01$ represents the magnitude of the bias. This mechanism produced a global increase in attribute x over simulation rounds. Even though the bias only affects the creation process, the preference of the agents also increased over time. This surprising side effect is induced by the social influence function: Because agents are exposed to items with larger values of x, they eventually develop a preference for this cultural trait – even though it has no inherent impact on the fitness of the item.
Second, this pattern could arise from a bias in the **diffusion function**. Here we consider that the value of the attribute x affects the item's diffusion range (cf. models of biased transmission in [61,62]. This mechanism is consistent with the effective-clickbait hypothesis: the title length serves as a proxy for clickbait which effectively favors item diffusion [63]. This can be modeled by assuming a functional dependency between an item's attribute and the sharing probability $p = f(x)$ instead of a fixed probability. More specifically, we simulated $p = p_0 + k.x$, where $p_0 = 0.1$ is the baseline sharing probability and $k = 0.05$ is the effect strength. We found that this mechanism also yields a cultural shift where agents gradually begin to prefer larger values of x.
Finally, the third mechanism involves a bias in the **social influence function**. Here we assume that items with larger values of x impact more strongly the preference of the agents exposed to them. This could arise from clickbait-style titles being more attention-grabbing or memorable. This can be modeled by introducing a functional dependency between an item's attribute and the social influence parameter $s = f(x)$ used in equation (2). Specifically, we simulated a situation where $s = s_0 + k.x$, where $s_0 = 0.5$ is the baseline social influence factor and $k = 0.5$ is the effect strength. Again, we found that this bias also yields a gradual increase of x over time. Agents exposed to an item with a long title are more likely to update their preference than if the title was shorter – even if the trait is ineffective [58]. Over time, the preference of the agents increases, resulting in the production of items with longer titles.

As shown in **figure 5A**, the three mechanisms outlined above produce the same collective pattern: an increase in the item's attribute over time. Hence, the mere observation of this pattern is insufficient to determine the underlying mechanism with certainty. However, each mechanism generates distinct population dynamics. For instance, a bias in the diffusion function induces a rise in cascade size, which is not observable with the other two (**figure 5B**). Similarly, a bias injected in the social influence function creates a more pronounced polarization of the agent's preferences (**figure 5C**). Hence, additional observations could help discriminate between the candidate mechanisms by measuring changes in cascade size or polarization level over the past years.

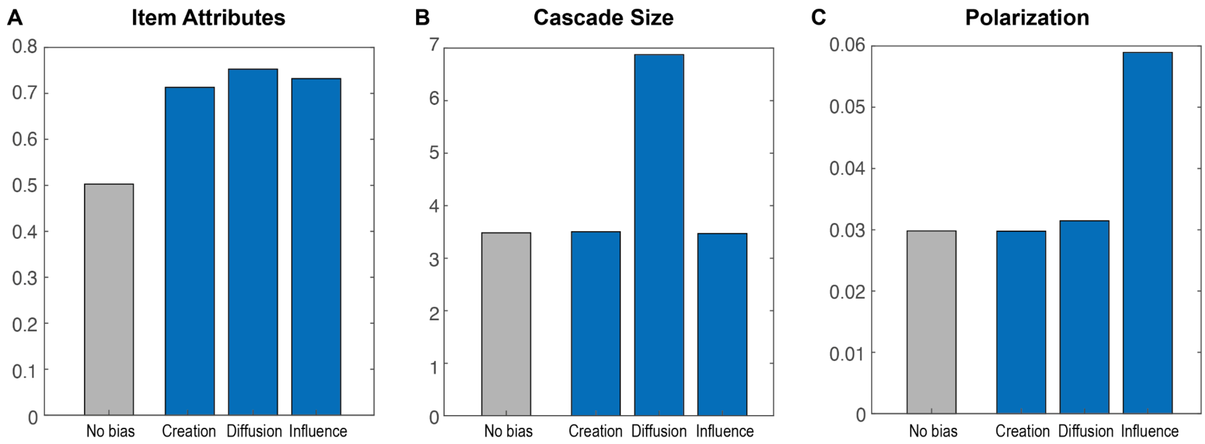

Figure 5: A) A trend towards longer headlines can result from a bias in any component in our model: item creation, item diffusion or social influence. However, injecting biases in different components of the model lead to different population dynamics and outcomes, such as average cascade size (B) or polarization (C). For example, a bias in the diffusion function may lead to deeper cascades, but not greater polarization.

In sum, the model constitutes an effective way to (1) determine the candidate mechanisms that could explain collective patterns, (2) facilitate the comprehension of the underlying dynamics, and (3) provide cues to help discriminate between competing mechanisms.

Discussion

While research has traditionally focused on the mechanisms of cultural evolution in offline environments, the omnipresence of the internet has created a new ecosystem where culture can emerge and evolve. Our research agenda, therefore, aims at applying existing concepts and methods of cultural evolution to the online realm. For this, we have elaborated a general framework describing what we believe are the key driving forces of online cultural evolution.

Still, our model has room for improvement that may need to be addressed in the future. First, our diffusion function does not capture algorithmic personalization. In fact, numerous social media platforms tend to personalize the newsfeed of their users, by presenting them primarily items with attributes matching the user’s preferences (e.g. a preferred topic). Such personalized recommendation systems can strongly bias the diffusion of digital items, interfering with the collective dynamics in place. The impact of this algorithmic bias will need to be carefully evaluated in the future, e.g. by introducing a functional dependency between the diffusion function and the agents' preferences.
Second, our model does not capture competition explicitly: we assumed for simplicity that digital items are produced one after the other and could diffuse in the population, eventually influencing subsequent productions. However, the current online environment is characterized by fierce competition for user attention and algorithmic selection. The model could be extended such that at each timestep, a certain number of items are produced simultaneously and compete with each other for attention and algorithmic traction. Only the most successful items would then be candidates for further sharing cascades.

Thirdly, our model lacks memory: older items are not stored to remain accessible to users. This may capture the dynamics of time-sensitive, novelty-driven, fast-paced platforms such as news outlets, Twitter or TikTok. However, digital environments facilitate the cumulation of information over time – a functionality we rely on when searching the internet. An item uploaded just now may compete for attention with a body of similar items uploaded over the last years. Blogs, Reddit posts, and YouTube tutorials, for instance, are not time-sensitive. This specificity may entail different strategies in the presentation of content, e.g. using attention-grabbing language for short-lived, time-sensitive content versus highly searchable language in the titles of long-lived content.

Further, the above is a single-trait model. We take a diffusionist approach, meaning we look at title length, wh-words and demonstratives as single traits that individuals could pick up independently. In reality, several traits may go hand-in-hand as part of a larger cultural package. If it were easier to define clickbait style as a bundle of necessarily co-occurring features (e.g. a style that requires both wh-words and demonstratives), one could model diffusion as holistic or hierarchical [64]. A holistic approach would mean that transmission is all-or-nothing: either people use all features of clickbait, resulting in clickbait headlines, or none, resulting in non-clickbait headlines. A hierarchical approach would require at least some features to co-occur and to be transmitted as a bundle. At the moment, we do not define clickbait as a set of defining features, but rather as a loose collection of strategies with a common goal, i.e. to maximize clicks. We take increased headline length as a common by-product of the strategies used, and the use of wh-words and demonstratives as straightforward (but not the only) devices to introduce forward reference, and with that, a curiosity gap. For subsequent models, it may be interesting to model multiple traits at once.

Lastly, our model treats cultural evolution as occurring in a closed system. In reality, however, items can exert influence beyond the platform they originated in: social media feeds, for instance, provide aggregating environments where diverse cultural traits – from news items to YouTube videos – compete for online attention. Similarly, online environments are not closed systems, but are influenced by offline culture – news and videos reflect offline events and general societal trends. While this does not invalidate our framework, it is important to remember that cultural influence may originate from a source not formally described in the simulations.

The natural next step of this research agenda will consist in testing the model's assumptions in light of empirical data collected from different online platforms. One promising research direction is the comparison of cultural traits between different platforms. In fact, each platform constitutes a specific environment characterized by different selective pressures: they differ in how strictly content is regulated, whether authors contribute with their real name or under a pseudonym (with implications for reputation costs), and in the production costs of the content (which is e.g. higher for a YouTube video than for a tweet or Reddit post). Social media platforms like Facebook or Twitter exhibit a particular social network structure, whereas others like Reddit do not involve connections between users – with implications for diffusion on the platform.

For a fruitful comparison, a common denominator between these diverse cultural outputs is necessary. Since many digital formats feature a title (e.g. newspaper articles, blogs, YouTube videos, etc.), this attribute could constitute a good basis for a cross-platform comparison, as we demonstrated in **figure 4**. For media with a longer history, like newspapers, it is even possible to compare the titles before and after the medium went online. Another promising attribute to consider is the image characteristics of thumbnails, a

format that similarly applies to news articles, YouTube videos and much other content designed to be shared on social media. This would allow discerning which trends are platform-specific and which pressures apply to online environments more generally.
This research can help us better understand the mechanisms of online culture, with possible insights into human psychology and applied perspectives for the design of social media platforms and the regulation of problematic online content.

Author Contributions

Conceptualization: P.N., M.M.; Methodology: P.N., M.M.; Formal Analysis: P.N., M.M, writing—original draft: P.N.; Writing—review and editing: P.N., M.M.; Supervision: M.M..

Data Availability:

The MatLab code for the simulations, as well as a jupyter notebook documenting the empirical findings are available at
https://arc-git.mpib-berlin.mpg.de/nickl/cultural_evolution_online

The New York Times headlines: The headlines can be collected via the official *The New York Times* API at https://developer.nytimes.com/apis

Upworthy headlines: The exploratory package can be downloaded from OSF at https://osf.io/3vqmp

YouTube titles: YouTube titles can be collected through the official YouTube API at https://developers.google.com/youtube/v3

Acknowledgements

We thank Philipp Lorenz-Spreen for his ongoing advice on the overall project.

References

1. Hills TT. The Dark Side of Information Proliferation. Perspect Psychol Sci. 2019;14: 323–330. doi:10.1177/1745691618803647
2. Lorenz-Spreen P, Mønsted BM, Hövel P, Lehmann S. Accelerating dynamics of collective attention. Nat Commun. 2019;10: 1–9.
3. Mesoudi A. Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences. Chicago, Ill.: University of Chicago Press; 2011.
4. O'Brien MJ, Lyman RL, Mesoudi A, VanPool TL. Cultural traits as units of analysis. Philos Trans R Soc B Biol Sci. 2010;365: 3797–3806. doi:10.1098/rstb.2010.0012
5. Claidière N, Scott-Phillips TC, Sperber D. How Darwinian is cultural evolution? Philos Trans R Soc B Biol Sci. 2014;369: 20130368. doi:10.1098/rstb.2013.0368
6. Acerbi A, Mesoudi A. If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. Biol Philos. 2015;30: 481–503. doi:10.1007/s10539-015-9490-2
7. Mesoudi A, Thornton A. What is cumulative cultural evolution? Proc R Soc B Biol Sci. 2018;285: 20180712. doi:10.1098/rspb.2018.0712
8. Boyd R, Richerson PJ. Culture and the evolutionary process. Paperback ed. Chicago: University of Chicago Press; 1988.
9. Sperber D. Explaining culture: a naturalistic approach. [Repr.], 1. publ. 1996. Oxford: Blackwell; 2002.
10. Acerbi A, Charbonneau M, Miton H, Scott-Phillips T. Culture without copying or selection. Evol Hum Sci. 2021;3: e50. doi:10.1017/ehs.2021.47
11. Kendal RL, Boogert NJ, Rendell L, Laland KN, Webster M, Jones PL. Social learning strategies: Bridge-building between fields. Trends Cogn Sci. 2018;22: 651–665.
12. Chudek M, Heller S, Birch S, Henrich J. Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning. Evol Hum Behav. 2012;33: 46–56. doi:10.1016/j.evolhumbehav.2011.05.005
13. Eriksson K, Coultas JC. Corpses, Maggots, Poodles and Rats: Emotional Selection Operating in Three Phases of Cultural Transmission of Urban Legends. J Cogn Cult. 2014;14: 1–26. doi:10.1163/15685373-12342107
14. Acerbi A. A Cultural Evolution Approach to Digital Media. Front Hum Neurosci. 2016;10. doi:10.3389/fnhum.2016.00636
15. Acerbi A. Cultural evolution in the digital age. First edition. Oxford: Oxford University press; 2020.
16. Mesoudi A, Whiten A, Dunbar R. A bias for social information in human cultural transmission. Br J Psychol. 2006;97: 405–423. doi:10.1348/000712605X85871
17. Derex M, Feron R, Godelle B, Raymond M. Social learning and the replication process: an experimental investigation. Proc R Soc B Biol Sci. 2015;282: 20150719. doi:10.1098/rspb.2015.0719
18. Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ. Emotion shapes the diffusion of moralized content in social networks. Proc Natl Acad Sci. 2017;114: 7313–7318. doi:10.1073/pnas.1618923114
19. Hills TT. The Dark Side of Information Proliferation. Perspect Psychol Sci. 2019;14: 323–330. doi:10.1177/1745691618803647
20. Goldenberg A, Gross JJ. Digital Emotion Contagion. Trends Cogn Sci. 2020;24: 316–328. doi:10.1016/j.tics.2020.01.009
21. Stubbersfield JM. Content biases in three phases of cultural transmission: A review. Cult Evol. 2022;19: 41–60. doi:10.1556/2055.2022.00024
22. Robertson CE, Pröllochs N, Schwarzenegger K, Pärnamets P, Van Bavel JJ, Feuerriegel S. Negativity drives online news consumption. Nat Hum Behav. 2023;7: 812–822. doi:10.1038/s41562-023-01538-4
23. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science.

2018;359: 1146–1151. doi:10.1126/science.aap9559
24. Acerbi A. Cognitive attraction and online misinformation. Palgrave Commun. 2019;5: 15. doi:10.1057/s41599-019-0224-y
25. Stubbersfield J, Tehrani J, Flynn E. Faking the News: Intentional Guided Variation Reflects Cognitive Biases in Transmission Chains Without Recall. Cult Sci J. 2018;10: 54–65. doi:10.5334/csci.109
26. Scott K. You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. J Pragmat. 2021;175: 53–66. doi:10.1016/j.pragma.2020.12.023
27. Nowak A, Szamrej J, Latané B. From private attitude to public opinion: A dynamic theory of social impact. Psychol Rev. 1990;97: 362–376. doi:10.1037/0033-295X.97.3.362
28. Axelrod R. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. J Confl Resolut. 1997;41: 203–226. doi:10.1177/0022002797041002001
29. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. Rev Mod Phys. 2009;81: 591–646. doi:10.1103/RevModPhys.81.591
30. Deffuant G, Neau D, Amblard F, Weisbuch G. Mixing beliefs among interacting agents. Adv Complex Syst. 2000;03: 87–98. doi:10.1142/S0219525900000078
31. Goel S, Watts DJ, Goldstein DG. The structure of online diffusion networks. Proceedings of the 13th ACM Conference on Electronic Commerce. Valencia Spain: ACM; 2012. pp. 623–638. doi:10.1145/2229012.2229058
32. Watts DJ. A simple model of global cascades on random networks. Proc Natl Acad Sci. 2002;99: 5766–5771. doi:10.1073/pnas.082090499
33. Watts DJ, Dodds PS. Influentials, Networks, and Public Opinion Formation. J Consum Res. 2007;34: 441–458. doi:10.1086/518527
34. Newman M, Barabási A-L, Watts DJ. The Structure and Dynamics of Networks: Princeton University Press; 2011. doi:10.1515/9781400841356
35. Soll JB, Larrick RP. Strategies for revising judgment: How (and how well) people use others' opinions. J Exp Psychol Learn Mem Cogn. 2009;35: 780–805. doi:10.1037/a0015145
36. Molleman L, Kurvers RHJM, Van Den Bos W. Unleashing the BEAST: a brief measure of human social information use. Evol Hum Behav. 2019;40: 492–499. doi:10.1016/j.evolhumbehav.2019.06.005
37. Moussaïd M, Kämmer JE, Analytis PP, Neth H. Social Influence and the Collective Dynamics of Opinion Formation. Szolnoki A, editor. PLoS ONE. 2013;8: e78433. doi:10.1371/journal.pone.0078433
38. Laland KN. Social learning strategies. Learn Behav. 2004;32: 4–14. doi:10.3758/BF03196002
39. Gaumont N, Panahi M, Chavalarias D. Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election. Araujo N, editor. PLOS ONE. 2018;13: e0201879. doi:10.1371/journal.pone.0201879
40. Gonçalves B, Perra N, Vespignani A. Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number. Perc M, editor. PLoS ONE. 2011;6: e22656. doi:10.1371/journal.pone.0022656
41. Grandjean M. A social network analysis of Twitter: Mapping the digital humanities community. Mauro A, editor. Cogent Arts Humanit. 2016;3: 1171458. doi:10.1080/23311983.2016.1171458
42. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science. 2019;363: 374–378. doi:10.1126/science.aau2706
43. Juul JL, Ugander J. Comparing information diffusion mechanisms by matching on cascade size. Proc Natl Acad Sci. 2021;118: e2100786118. doi:10.1073/pnas.2100786118
44. Liang H, Fung IC-H, Tse ZTH, Yin J, Chan C-H, Pechta LE, et al. How did Ebola information spread on twitter: broadcasting or viral spreading? BMC Public Health.

2019;19: 438. doi:10.1186/s12889-019-6747-8
45. De Oliveira DVB, Albuquerque UP. Cultural Evolution and Digital Media: Diffusion of Fake News About COVID-19 on Twitter. SN Comput Sci. 2021;2: 430. doi:10.1007/s42979-021-00836-w
46. Rieder B, Coromina Ò, Matamoros-Fernández A. Mapping YouTube: A quantitative exploration of a platformed media system. First Monday. 2020 [cited 25 Apr 2022]. doi:10.5210/fm.v25i8.10667
47. Munger K. The YouTube Apparatus. Elem Polit Commun. 2024 [cited 20 Jun 2024]. doi:10.1017/9781009359795
48. Munger K. The YouTube Apparatus. 1st ed. Cambridge University Press; 2024. doi:10.1017/9781009359795
49. Rieder B, Coromina Ò, Matamoros-Fernández A. Mapping YouTube. First Monday. 2020 [cited 13 Sep 2024]. doi:10.5210/fm.v25i8.10667
50. Blom JN, Hansen KR. Click bait: Forward-reference as lure in online news headlines. J Pragmat. 2015;76: 87–100. doi:10.1016/j.pragma.2014.11.010
51. Chakraborty A, Paranjape B, Kakarla S, Ganguly N. Stop clickbait: Detecting and preventing clickbaits in online news media. 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE; 2016. pp. 9–16. Available: https://ieeexplore.ieee.org/abstract/document/7752207/
52. D. Molina M, Sundar SS, Rony MMU, Hassan N, Le T, Lee D. Does Clickbait Actually Attract More Clicks? Three Clickbait Studies You Must Read. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Yokohama Japan: ACM; 2021. pp. 1–19. doi:10.1145/3411764.3445753
53. Kuiken J, Schuth A, Spitters M, Marx M. Effective Headlines of Newspaper Articles in a Digital Environment. Digit Journal. 2017;5: 1300–1314. doi:10.1080/21670811.2017.1279978
54. Munger K. All the News That's Fit to Click: The Economics of Clickbait Media. Polit Commun. 2020;37: 376–397. doi:10.1080/10584609.2019.1687626
55. Bulik M. Which Headlines Attract Most Readers? The New York Times. 13 Jun 2016. Available: https://www.nytimes.com/2016/06/13/insider/which-headlines-attract-most-readers.html. Accessed 25 Apr 2024.
56. Lockwood G. Academic clickbait: articles with positively-framed titles, interesting phrasing, and no wordplay get more attention online. [cited 13 Sep 2024]. Available: https://www.authorea.com/doi/full/10.15200/winn.146723.36330?commit=e6ce98db83ee601f0e5e90ae70bc3baeafd98c77
57. Loewenstein G. The psychology of curiosity: A review and reinterpretation. Psychol Bull. 1994;116: 75–98. doi:10.1037/0033-2909.116.1.75
58. Miton H, Claidière N, Mercier H. Universal cognitive mechanisms explain the cultural success of bloodletting. Evol Hum Behav. 2015;36: 303–312. doi:10.1016/j.evolhumbehav.2015.01.003
59. Ball J. Read this to find out how Upworthy's awful headlines changed the web. The Guardian. 16 Mar 2014. Available: https://www.theguardian.com/media/2014/mar/16/upworthy-website-generation-y-awful-headlines. Accessed 11 Feb 2023.
60. Matias JN, Munger K, Le Quere MA, Ebersole C. The Upworthy Research Archive, a time series of 32,487 experiments in U.S. media. Sci Data. 2021;8: 195. doi:10.1038/s41597-021-00934-7
61. Mesoudi A. Simulation models of cultural evolution in R. Zenodo; 2021. doi:10.5281/ZENODO.5155821
62. Acerbi A, Mesoudi A, Smolla M. Individual-Based Models of Cultural Evolution: A Step-by-Step Guide Using R. 1st ed. London: Routledge; 2022. doi:10.4324/9781003282068
63. Manjoo F. Can Facebook Fix Its Own Worst Bug? The New York Times. 25 Apr 2017. Available:

https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html. Accessed 29 Apr 2021.
64. Mesoudi A, O'Brien MJ. The Learning and Transmission of Hierarchical Cultural Recipes. Biol Theory. 2008;3: 63–72. doi:10.1162/biot.2008.3.1.63