Cognition 249 (2024) 105809

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Why do languages tolerate heterography? An experimental investigation into the emergence of informative orthography

Jon W. Carr *, Kathleen Rastle

Department of Psychology, Royal Holloway, University of London, United Kingdom

ARTICLE INFO

Keywords:
Communication
Iterated learning
Language evolution
Orthography
Reading
Writing

ABSTRACT

It is widely acknowledged that opaque orthographies place additional demands on learning, often requiring many years to fully acquire. It is less widely recognized, however, that such opacity may offer certain benefits in the context of reading. For example, heterographic homophones such as ⟨knight⟩ and ⟨night⟩ (words that sound the same but which are spelled differently) impose additional costs in learning but reduce ambiguity in reading. Here, we consider the possibility that—left to evolve freely—writing systems will sometimes choose to forego some simplicity for the sake of informativeness when there is functional pressure to do so. We investigate this hypothesis by simulating the evolution of orthography as it is transmitted from one generation to the next, both with and without a communicative pressure for ambiguity avoidance. In addition, we consider two mechanisms by which informative heterography might be selected for: differentiation, in which new spellings are created to differentiate meaning (e.g., ⟨lite⟩ vs. ⟨light⟩), and conservation, in which heterography arises as a byproduct of sound change (e.g., ⟨meat⟩ vs. ⟨meet⟩). Under pressure from learning alone, orthographic systems become transparent, but when combined with communicative pressure, they tend to favor some additional informativeness. Nevertheless, our findings also suggest that, in the long term, simpler, transparent spellings may be preferred in the absence of top-down explicit teaching.

1. Introduction

Writing systems, particularly those employing alphabetic scripts, are commonly regarded as providing a visual representation of speech, with letters or chunks of letters corresponding to distinct sounds. However, it is also well understood that writing systems diverge from their spoken counterparts in important ways (Biber, 1988; Bolinger, 1946; Coulmas, 1991). The insertion of spacing between words, for example, is almost ubiquitous across alphabetic writing systems, even though no such spacing exists between words in speech (Parkes, 1992; Saenger, 1997). It seems likely that graphic innovations such as these exist because they confer some benefit that is not required in the spoken modality (Rastle, 2019). In the case of spacing, for example, the separation of words into discrete chunks presumably aids in the targeting and extraction of visuo-linguistic information—constraints that do not exist in the auditory modality. In principle, the same may be true of spelling: Words may be spelled in ways that diverge from the spoken language because such divergence confers some benefit in reading (Ulicheva, Harvey, Aronoff, & Rastle, 2020).

One potential case of such functional divergence is heterographic homophony—words that sound alike but which are written differently (e.g., ⟨meat⟩ and ⟨meet⟩ for /miːt/). Heterographic spellings such as these may serve a valuable function in reading. For example, an English speaker faced with a spoken sentence beginning /ðɛr…/ will have high uncertainty about what word—or even what sentence structure—is likely to come next: a noun, as in /ðɛr kat/, a form of the verb *to be*, as in /ðɛr ɪz/, or the progressive form of a verb, as in /ðɛr gəʊɪŋ/. In writing, by contrast, this uncertainty is greatly reduced; the spellings ⟨their⟩, ⟨there⟩, and ⟨they're⟩ differentiate these cases, giving the reader a headstart on processing the upcoming syntactic structure and semantic content. Heterographic homophony is also common below the word level, since many orthographies forego the phonological principle in favor of the morphological principle in the spelling of affixes (Sandra, Ravid, & Plag, 2024). The English suffixes *-er* (denoting the comparative form of an adjective; e.g., *nicer*) and *-or* (denoting the performer of an action; e.g., *actor*) are homophonous in speech (/ər/), but their spellings differentiate these meanings in writing. Of course, English orthography is suboptimal here in that *-er* may also indicate agentive status (e.g.,

* Corresponding author.
E-mail address: jon.carr@rhul.ac.uk (J.W. Carr).

https://doi.org/10.1016/j.cognition.2024.105809
Received 3 November 2023; Received in revised form 25 April 2024; Accepted 29 April 2024

builder); nevertheless, statistical patterns such as these hold across a variety of English affixes (Berg & Aronoff, 2017) and it has been shown that readers are sensitive to and make use of such cues in reading (Ulicheva et al., 2020). Heterography might be especially important given the differing constraints of the written modality, including the lack of other cues to meaning, such as stress, context, and body language, and the inability for reader and writer to engage in immediate feedback and repair. Furthermore, written language has richer vocabulary and more complex syntax than spoken language (Biber, 1988; Korochkina, Marelli, Brysbaert, & Rastle, 2024; Nation, Dawson, & Hsiao, 2022), placing different pressures on ambiguity resolution.

Heterography is particularly notable in English, but it is also a feature of many other languages and writing systems. French is similar to English in having a large number of heterographic homophones: *cent* (*hundred*), *sang* (*blood*), *sans* (*without*), and *sens* (*feel*), for example, are all pronounced /sɑ̃/ (although their pronunciations will sometimes be distinguished through liaison). In Danish, the words *hver* (*every*), *vejr* (*weather*), *vær* (*be*), and *værd* (*worth*) are all pronounced /vɛˀɐ̯/. In Vietnamese, the graphemes ⟨d⟩, ⟨gi⟩, and ⟨r⟩ are homophonous, resulting in sets like *dao* (*knife*), *giao* (*delivery*), and *rao* (*advertise*), all pronounced /zāw/. In some cases, features of an orthography designed for other purposes can inadvertently result in homophone disambiguation: Noun capitalization in German contrasts *Wagen* (*car*) and *wagen* (*to dare*), both pronounced /vaːgn̩/; eclipsis marking in Irish contrasts *bpáistí* (*children*) and *báistí* (*rain*), both pronounced /bɑʃtʲi/; and the morphological principle of Russian orthography contrasts приступить (*to start*) and преступить (*to transgress*), both pronounced /pristupʲitʲ/. Even in the most transparent of orthographies it is possible to find some instances of heterography: In Italian, a residual initial ⟨h⟩ inherited from Latin contrasts *hanno* (*have*) and *anno* (*year*), both pronounced /anno/, while the grave accent is sometimes used to distinguish common homophonous words, such as *la* (*the*) and *là* (*there*).

Perhaps the most elaborate example of how a writing system can deal with homophony head-on is the Chinese orthography. The Chinese spoken languages are rich in homophones, making heterographic spellings—and therefore a logographic writing system—particularly useful (Frost, 2012). In Mandarin Chinese, the words 糖 (*sugar*), 塘 (*embankment*), 溏 (*pond*), and 搪 (*to block*) are homophonous in speech but heterographic in writing—the phonetic radical on the right (⟨唐⟩, /táŋ/) represents the spoken syllable, while the semantic radicals on the left differentiate the meanings (Coulmas, 1991, p. 101). In addition, 唐 itself is a surname/dynasty (*Tang*), and another unrelated word 堂 (*hall*) is also pronounced /táŋ/, yielding at least six ways to write the same sound depending on the meaning. This property allows the written form of Chinese to convey more information about meaning—to be more *informative*—than its spoken counterpart.

Despite the benefits that heterography may provide in reading, it comes with two main costs. Firstly, by definition, heterography implies that a single sound can be spelled multiple ways. In English, the heterographic spellings ⟨meat⟩ and ⟨meet⟩ imply that /iː/ can be spelled ⟨ea⟩ or ⟨ee⟩. Readers are therefore required to learn alternate spellings for a single sound, resulting in longer learning periods and more difficult decoding (Reis, Araújo, Morais, & Faísca, 2020; Seymour, Aro, & Erskine, 2003; Spencer & Hanley, 2003; Taylor, Plunkett, & Nation, 2011; Zhao, Li, Elliott, & Rueckl, 2018). Secondly, the arbitrary mapping between heterographic forms and meaning must also be learned. From the point of view of a modern English speaker, there is no intrinsic reason why *meat* is spelled ⟨ea⟩ and *meet* is spelled ⟨ee⟩. Nevertheless, these arbitrary spelling distinctions must be learned if they are to be useful, and they presumably place an additional burden on reading and—perhaps even more-so—on writing (Frith, 1979; Shankweiler & Lundquist, 1992).

In this paper, we consider the possibility that—left to evolve freely—writing systems will sometimes choose to forego some simplicity for the sake of informativeness. A simple spelling system would be one that is easy to learn, use, and process; for example, by being transparent with respect to phonology. An informative spelling system, on the other hand, would be one that precisely conveys meaning. This idea of a tradeoff between simplicity and informativeness in the writing system has long been noted (e.g., Coulmas, 1991), and such a tradeoff has also been discussed within the study of language more broadly (e.g., Gabelentz, 1891; Martinet, 1952; Rosch, 1978; Zipf, 1949). Recent typological (e.g., Kemp, Xu, & Regier, 2018) and experimental (e.g., Kirby, Tamariz, Cornish, & Smith, 2015) studies have also subjected these ideas to empirical investigation in various domains. Of particular note here is the finding that complex, systematic structure emerges under concurrent pressures to be both simple and informative, a point we return to shortly.

First, however, it is useful to consider the mechanisms by which selection could occur if it is indeed the case that heterography emerges for functional reasons. Berg and Aronoff (2021, pp. 325–326) outline two models of how a word might enter a state of heterography. The first model, *the differentiation model* (Fig. 1A), explains heterography through the creation of new orthographic forms. For example, the spelling ⟨lite⟩ for the word *light* is frequently used in food products to mean light-in-calorific-weight; in British English, the spelling ⟨cheque⟩ (perhaps influenced by French *chèque*) differentiates the bank draft from other meanings of the word *check*; and the word *byte* was a deliberate respelling of *bite* to avoid accidental mutation into the closely related term *bit* (Buchholz, 1977). Many monosyllabic words that are homophonous with common function words also tend to adopt alternate spellings, often by appending ⟨e⟩ or by doubling the final consonant: *be–bee*, *but–butt*, *by–bye–buy*, *for–fore–four*, *in–inn*, *or–oar–ore*, *so–sew*, *to–too–two*, *we–wee*. Differentiation is also common in surnames–*Clarke*, *Greene*, *Wilde*; *Carr*, *Hogg*, *Mann*—and trade names—*Blu Tack*, *Froot Loops*, *Wite-Out* (Carney, 1994, *sec*. 6). One important way in which differentiation can occur—especially in a language like English that has historically contained a lot of spelling variation (Nevalainen, 2012; Stenroos & Smith, 2016)—is by the conditioning of variant spellings on meaning; pairs like *discreet–discrete*, *flour–flower*, and *plain–plane*, which were once variant spellings of the same word, have taken on distinct meanings over time (Carney, 1994, sec. 5.4). Berg and Aronoff (2017, p. 58) have referred to this as the "functionalization of leftovers": The spelling variants that survive are those that "can find distributional or functional niches."

The second model, *the conservation model* (Fig. 1B), explains heterographic homophones as the historical residue of sound change: Two spoken forms merge and become homophonous, but the original spellings are conserved in the orthography. For example, the *meat–meet* merger that occurred during the Great Vowel Shift ultimately resulted in Middle English /ɛː/ (spelled ⟨ea⟩) and /eː/ (spelled ⟨ee⟩) being pronounced /iː/ in Early Modern English (Lass, 2000), but the spellings were never changed accordingly, thus giving rise to a set of heterographic homophones that persist in present-day English (Wells, 1982, pp. 140–141): *heal–heel*, *leak–leek*, *meat–meet*, *read–reed*, *sea–see*, *team–teem*, *weak–week*. The same is true of the *pain–pane* merger (Wells, 1982, pp. 141–142): *maid–made*, *main–mane*, *pain–pane*, *raise–raze*, *sail–sale*, *vain–vane*. Sound changes involving consonants have also resulted in (or contributed to) pairs of words entering a state of heterography, such as the reduction of /kn/ into /n/ (e.g., *knight–night*, *know–no*, *knot–not*), the loss of /ç/ (e.g., *eight–ate*, *right–rite*, *sight–site*), and the merger of /ʍ/ into /w/ (e.g., *whale–wail*, *which–witch*, *whine–wine*).[1] A more recent (and perhaps in-progress) example can be

[1] Although we can never be entirely certain how words were pronounced before the advent of sound recording technology, historical linguists have compiled persuasive evidence by a variety of methods. Comparison to modern German, for example, offers an insight into how these words might have been pronounced in the past (e.g., *knot* is cognate with *Knoten* where the /kn/ cluster continues to be fully rendered and *eight* is cognate with *acht*, where the palatal fricative still exists).

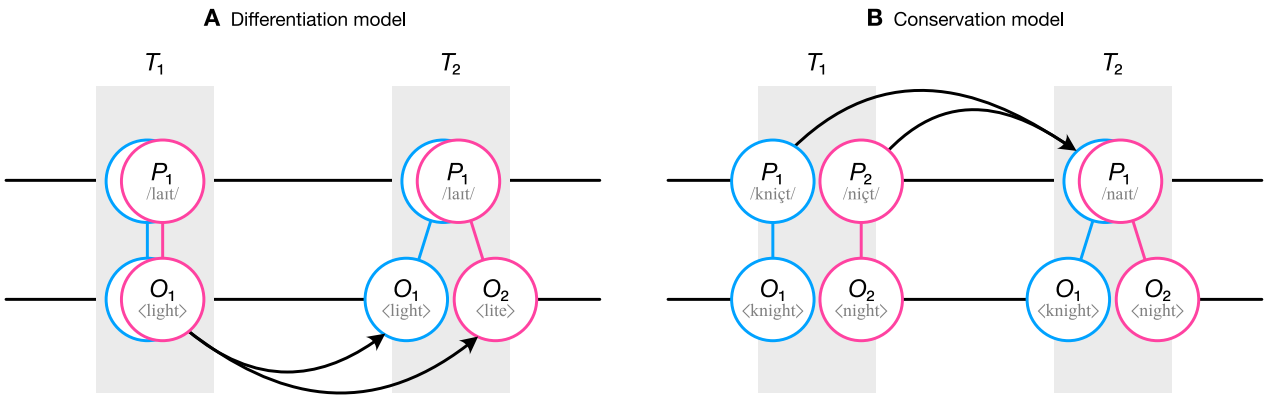

Fig. 1. Two models of heterography. **A** In the differentiation model, two meanings are, at time T_1, expressed by a single phonetic form P_1 and a single orthographic form O_1; however, by time T_2, two orthographic forms have emerged to differentiate the meanings in writing. **B** In the conservation model, the two distinct phonetic forms that existed at time T_1 have become homophonous by time T_2, but the two corresponding orthographic forms have been conserved, resulting in the same state of heterography as in the differentiation model. Adapted from Berg and Aronoff (2021, pp. 325–326) with permission.

found in dialects that have undergone the *father–bother* merger, including most varieties of American English (Labov, Ash, & Boberg, 2005, p. 169), which has resulted in, for example, *balm–bomb* (/bɑm/), *lager–logger* (/lɑgər/), and *mach–mock* (/mɑk/); these word pairs continue to be spelled with ⟨a⟩ vs. ⟨o⟩ despite being homophonous in such dialects.

In some cases, it is debatable whether a given case of heterography was delivered by the differentiation or conservation mechanism. For example, while the etymological (and folk-etymological) respellings introduced during the Renaissance might appear to be cases of conservation (the most notorious example being the replacement of ⟨dout⟩ with ⟨doubt⟩ to indicate the word's Latin derivation from *dubitare*; Crystal, 2005, p. 268), it has also been argued that such respellings were motivated in part by a desire to differentiate homophones such as *scene–seen*, *scent–sent*, and *whole–hole* (Scragg, 1974, pp. 58–59). Nevertheless, regardless of the particular mechanism behind specific cases in English or any other language, our primary contention here is that both of these mechanisms provide adaptive, functional explanations for heterography. Differentiated spellings that prove communicatively useful will be more likely to survive; likewise, conserved spellings that prove communicatively useful will be more likely to survive.

Our aims in this paper are twofold. First, we test the idea that heterography emerges in response to a functional pressure to disambiguate meaning in writing. Second, we seek to understand how the emergence of heterography plays out under the two mechanisms of differentiation and conservation. Is one of these a better candidate explanation than the other? Approaching these evolutionary questions using data from natural languages is challenging. In particular, the available diachronic data (for any language) will necessarily be limited and impoverished—languages do not fossilize well, especially in their spoken forms. In addition, any answer derived from such datasets will have to rely on correlational, as opposed to causal, evidence—we cannot rerun history many times under different conditions.

We therefore turn to a different approach. Here we experimentally simulate the processes of differentiation and conservation using the experimental iterated learning paradigm (Kirby, Cornish, & Smith, 2008). In this paradigm, an artificially constructed language (or so-called "alien language") is passed along a *transmission chain* of human participants, simulating what happens during the cultural transmission and evolution of language. Participant i in a transmission chain learns the system based on the linguistic output of participant $i-1$ and, subsequently, produces new linguistic output for participant $i+1$ to learn from, although the participants themselves are not aware of this generational structure. It has been demonstrated in a wide variety of studies that, after several generations of cultural transmission, artificial languages can gradually adapt to the biases of the human learners and the environments in which they are used, yielding emergent linguistic phenomena, such as compositionality (Beckner, Pierrehumbert, & Hay, 2017; Kirby et al., 2008, 2015), combinatoriality (Verhoef, Kirby, & Boer, 2015), semantic category structure (Canini, Griffiths, Vanpaemel, & Kalish, 2014; Carr, Smith, Cornish and Kirby, 2017; Silvey, Kirby, & Smith, 2019), regularization (Smith & Wonnacott, 2010), and argument marking (Motamedi, Smith, Schouwstra, Culbertson, & Kirby, 2021), among many other things. For reviews, see Bailes and Cuskley (2023), Kirby et al. (2014), Kirby (2017), Smith (2022), and Tamariz (2017).

Kirby et al. (2008) described the first experimental application of the iterated learning framework (which had previously been confined to computational modeling), showing that compositional structure—a systematic relationship between recombinant linguistic units and meaning—could spontaneously emerge under a *bottleneck on transmission*. This "bottleneck" defines a limit on the amount of information that can flow from one generation to the next (Brighton, 2002). Under a tight bottleneck, where little data passes from one generation to the next, the learner must perform more generalization from less input, such that the cognitive biases that the learner brings to the table become more important in shaping the structure of the emergent language. Generalization from limited input is a major driver of systematic structure in the language as a whole, since human learners tend to generalize in ways that increase the simplicity (i.e., systematicity) of the system, albeit unconsciously (Culbertson & Kirby, 2016). However, left unchecked, this bias for simplicity would ultimately result in the emergence of maximally simple, degenerate languages. Kirby et al. (2015) therefore extended the framework by including a communicative task; instead of each generation consisting of a single participant, each generation now consisted of a pair of participants engaged in a shared task requiring communicative precision. Crucially, this communicative component prevented the artificial languages from degenerating; instead, the languages find a tradeoff between simplicity on the one hand and informativeness on the other, just as in natural language (Kemp et al., 2018; Kemp & Regier, 2012; Mollica et al., 2021; Regier, Kemp, & Kay, 2015; Zaslavsky, Kemp, Regier, & Tishby, 2018).

Our paper reports the results of two experiments—focusing on the differentiation and conservation models respectively—with the goal of demonstrating that functional heterography arises preferentially under a communicative need for disambiguation. All data and code is available from https://osf.io/7auw6/. To increase the transparency of our work, we created a preregistration at https://aspredicted.org/p8aw9.pdf. Note, however, that due to the more exploratory nature of this project, our preregistration did not specify strong confirmatory hypotheses or precise statistical models, focusing instead on the research question, experimental conditions, general predictions, primary measurement constructs, sample size, and exclusion criteria.

2. Experiment 1

Our first experiment tests the ability of the differentiation model to explain the emergence of informative orthography. Can variant spellings become conditioned on meaning, such that the written form of the language diverges from the spoken form in a way that is expressive, despite the extra cost in learning? We had two main hypotheses:

1. Under pressure from learning alone, we expect to see the emergence of an increasingly transparent orthography.
2. Under additional pressure for disambiguation, we expect to see greater use of differentiated, non-transparent spellings.

2.1. Methods

Our methods follow the experimental iterated learning literature, as described above, with one main difference: The artificial language has both spoken and written forms that may diverge or converge over time. Broadly, participants are first asked to learn a simple alien language (consisting of words for colored shapes) and are then asked to reproduce what they learned in a test phase. The written form of the language may change over time, since the orthographic output of participant i becomes the input to participant $i + 1$ in a transmission chain design, but the spoken form of the language remains fixed and under experimenter control. To explore the hypotheses outlined above, we conducted the experiment under two different conditions: Transmission-only, in which the test phases emphasizes simple reproduction, and Transmission + Communication, in which the test phase encourages disambiguation.

2.1.1. Participants

We recruited 287 participants via the Prolific platform. Participants were paid £2.00 for participation plus additional bonuses of up to £1.08 as detailed below (median bonus: £0.74). The median completion time was 15 m with a median hourly rate of £8.05 (£10.94 including bonus). We limited recruitment to (self-declared) native English speakers, since it was important that participants would perceive the spoken forms in a relatively consistent way (particularly in the case of Experiment 2). 14 participants were excluded because they (or their communication partners) used English color words (8) or failed the auditory attention checks (6). A further three participants were lost to communication-game pairing failures. The final dataset comprises 270 participants: 90 in the Transmission-only condition (10 chains of 9 participants) and 180 in the Transmission + Communication condition (10 chains of 9 pairs of participants).

2.1.2. Stimuli

Participants were taught words for nine alien objects—three shapes (pentagon, star, torus) in three colors (pink, yellow, blue), as depicted in Fig. 2. The alien words had a spoken and written form composed of a stem and suffix. The stems, which always express the shape dimension, were /buvɪ/ ⟨buvi⟩ (the pentagon), /zɛtɪ/ ⟨zeti⟩ (the star), and /wɒpɪ/ ⟨wopi⟩ (the torus). These stems were fixed and unchanging throughout both experiments reported in this paper and were designed to be easy to learn by being graphically and phonetically iconic of the shapes they represent (e.g., the round torus shape is represented by "round" sounds/letters). Throughout Experiment 1, the spoken form of the suffix was always pronounced /kəʊ/, but its spelling was free to change over time. Thus, the spoken form of the language consists of just three unique words—/buvɪkəʊ/, /zɛtɪkəʊ/, and /wɒpɪkəʊ/—that mark only a shape distinction; however, the spelling of the suffix could potentially take on different forms to mark color.

Each of the 10 transmission chains was seeded with a randomly-generated suffix spelling system, which was created by randomly mapping the following nine spellings onto the nine objects: ⟨co⟩, ⟨coe⟩, ⟨coh⟩, ⟨ko⟩, ⟨koe⟩, ⟨koh⟩, ⟨qo⟩, ⟨qoe⟩, ⟨qoh⟩. In other words, the /k/ sound may be spelled ⟨c⟩, ⟨k⟩, or ⟨q⟩ and the /əʊ/ sound may be spelled ⟨o⟩, ⟨oe⟩, or ⟨oh⟩, although the initial seed system contained no particular regularity. This procedure models an initial state of high spelling variation (every object has a unique suffix spelling), but these spellings may, over time, become transparent (the /k/ and/or /əʊ/ sounds take on consistent spellings) or expressive (spellings of /k/ and/or /əʊ/ become systematically associated with meaning). The spoken forms were synthesized using the Apple text-to-speech synthesizer (Tessa voice).

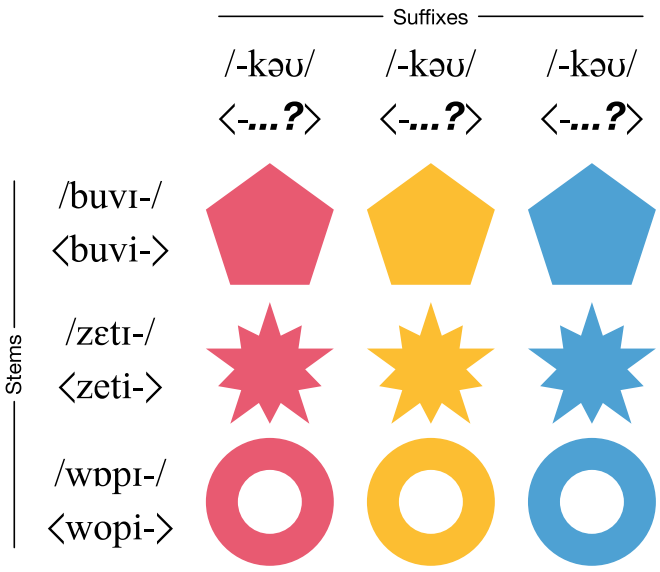

Fig. 2. The nine object stimuli with their stems and suffixes. The spoken and written forms of the stems were fixed and unchanging throughout the experiment, as were the spoken forms of the suffixes, which were always homophonous; however, the written forms of the suffixes were free to evolve over time, potentially taking on differentiated forms to indicate color (e.g., ⟨co⟩, ⟨ko⟩, and ⟨qo⟩ to represent pink, yellow, and blue).

2.1.3. Transmission procedure

Participants were arranged into transmission chains such that the spellings produced by one participant would subsequently be taught to the next participant in the chain (see Fig. 3A). The first participant in a chain was taught the initial, randomly generated seed system, and this system was then free to evolve as it was transmitted to subsequent generations. Importantly, this process was subject to a bottleneck on transmission: Not all nine spellings were transmitted from one generation to the next; rather, the participant at generation i would observe only six of the nine spellings produced at generation $i - 1$ (at least one of each shape and at least one of each color).[2] Nevertheless, participants were asked to produce a spelling for all nine objects, meaning that generalization was required for three unseen items. Transmission continued for nine generations in each of ten independent chains. In the Transmission + Communication condition (see Fig. 3B), each generation consisted of a pair of participants, but the productions of only one of the two (the primary participant; determined by whichever participant started the experiment first) were iterated to the next generation. The productions of the secondary participant were not iterated any further and were thus a cultural deadend; the role of the secondary participant was to act as a genuine communicative partner for the primary participant, inducing pressure for the language to become informative.

[2] The 2/3 bottleneck parameter was chosen based on common practices in the field and some piloting. In general, increasing this parameter will lead to a slower evolutionary process, so the value was chosen to allow for sufficient change to occur within nine generations.

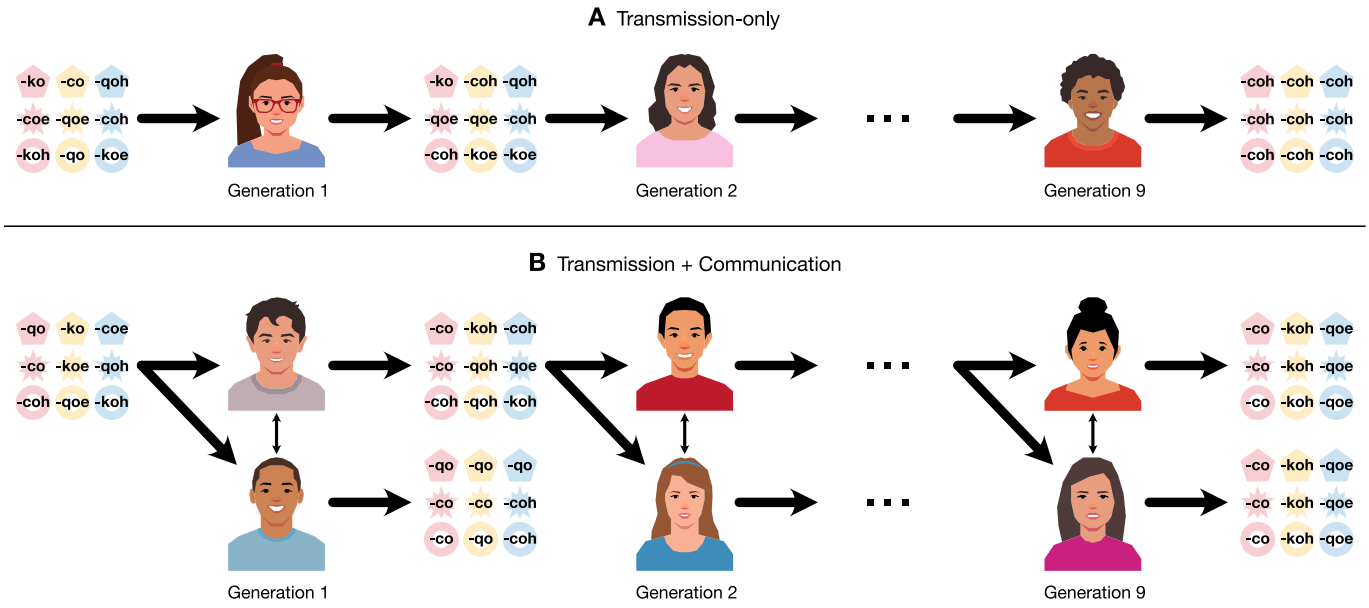

Fig. 3. **A** Transmission-only procedure. Each generation consists of a single participant, who first receives training on six of the nine suffixes and then produces suffixes for all nine items. These productions are then used as the training material for the next generation in the chain. The initial system of suffixes is randomly generated with high spelling variation; but by the ninth generation, the system is expected to become transparent. **B** Transmission + Communication procedure. Each generation now consists of two participants who engage in a communicative task. Both participants receive training on the same system from the previous generation. Under a communicative pressure, the suffix spellings are expected to become expressive of color, despite the spoken forms being homophonous.

2.1.4. *Training procedure*

All participants were trained on the spoken and written forms through a combination of passive exposure trials and "mini-test" trials, lasting around 8 min. Participants were also told explicitly before starting the training session that the stems looked and sounded like the objects' shapes, allowing participants to focus on learning the suffixes during the training phase. In passive exposure trials, the alien objects were presented alongside the written and spoken forms in quick succession for 2 s each. In mini-test trials, which were interleaved among the passive exposure trials, the participant was asked to type the appropriate written form for an object and was given feedback on any errors (deleted characters shown in red strikethrough text and additions shown in bold green text). The participant received a 2p bonus for spelling the word correctly but had to submit their response within 20 s. Each of the six object–word pairs in the training set (i.e., the seen items that passed through the bottleneck on transmission) was passively exposed 18 times and mini-tested six times, resulting in a total of 108 passive exposure trials and 36 mini-test trials (the maximum bonus in training was therefore 72p). To check that participants were listening to the spoken forms, they were asked auditorily to click on the alien object at three random points during training; participants who did not follow this instruction were excluded. The instructions provided to participants are provided in Appendix A in the supplementary material.

2.1.5. *Test procedure in transmission-only*

After training, participants assigned to the Transmission-only condition completed a test phase, alternating between production and comprehension trials. In production trials, the participant was shown an object and heard its associated pronunciation. The participant's task was to type the appropriate spelling. The input box was limited to eight lowercase Latin characters, and participants had to spell the stem correctly to continue to the next trial. Since participants heard the word pronounced aloud, typing the stem correctly should have been trivial, but in cases where the stem was initially spelled incorrectly, a popup message explicitly reminded the participant of the correct spelling of the stem.[3] This restriction was imposed to prevent the stems from diverging from their spoken forms over time; however, no such restriction was imposed on the spelling of the suffix. Since the overall word length was restricted to eight characters and since all stems were four letters long, participants could use, at a minimum, a zero suffix and, at a maximum, a four-letter suffix. In comprehension trials, the participant was shown a word and had to click on the matching object from an array of all nine objects arranged in random order (in cases where multiple objects were described by the same wordform, any of the objects was a valid choice). In both types of test trial, the participant was awarded a bonus of 2p for each correct answer, but no explicit feedback was provided on the correctness of the signal or object selection. Each of the nine object–word pairs (i.e., including unseen items) was tested once in production and once in comprehension, resulting in 18 trials (the maximum bonus in test was therefore 36p).

2.1.6. *Test procedure in transmission + communication*

Participants assigned to the communicative condition completed a live, over-the-internet communication game with another participant. Both participants received training on the same orthographic system inherited from the previous generation.[4] The communication game, which shares similarities with Kirby et al. (2015), closely mirrored the overall structure of the test administered to participants in the non-communicative condition described above, with the production and comprehension trials becoming the two sides of a single communicative interaction. On a given trial, one participant (the director) would complete a production trial under the same input restrictions described

[3] Although artificial, these stem correction messages were rarely encountered. 81% of participants never encountered this message (they always spelled the stem correctly), 14% encountered it on one trial (out of nine), and 5% encountered it on two or three trials.

[4] The seen items were selected independently for each participant. We felt this was more ecological than giving both participants identical input.

above (i.e., produce a form for a target meaning), and the word they used was relayed to the other participant (the matcher), who would then complete a comprehension trial in response to that word (i.e., pick an object from the matcher array). The two participants then switched roles, resulting in the same overall trial structure as the Transmission-only condition (i.e., alternation between production and comprehension trials).

The framing and goal of the communication game was, however, quite different from the non-communicative test. In communication, the shared goal of the director and matcher was to have a successful communicative interaction, not necessarily to reproduce what they had learned in training. Both participants received the 2p bonus each time an interaction was successful; that is, the reward structure is not based on using the "correct" forms taught in training but based on accurately conveying meaning. The director is thus incentivized to produce a wordform that is unambiguous, and the matcher is incentivized to carefully interpret what the director has attempted to convey. The second important difference from the non-communicative test was that participants received rich feedback on the interaction: The director saw which object the matcher clicked on, and the matcher saw which object was the correct target. As such, we view feedback as an intrinsic part of communicative interaction; thus, in the non-communicative test described above, no feedback was provided, as is the case in similar studies (Carr et al., 2017; Motamedi, Schouwstra, Smith, Culbertson, & Kirby, 2019; Saldana, Kirby, Truswell, & Smith, 2019; Silvey, Kirby, & Smith, 2019).

Overall, the communication game is identical to the non-communicative test in terms of the task to be performed (nine productions and nine comprehensions), but the goal is quite different. In the non-communicative test, the goal and reward structure are based on accurately reproducing the orthography learned during training, whereas in the communication game, the goal and reward structure are based on successfully communicating a target object.

2.2. Results

The results from all ten chains (labeled A–J) in the Transmission-only condition are shown in Fig. 4. Each 3×3 matrix represents the suffix spelling system in use at a particular generation with shape represented along the rows and color represented along the columns, following the same 3×3 layout used in Fig. 2. The color-coding of these matrices indicates similarity in suffix form: Similar colors are used to represent similar suffixes, making it easier to see how the suffixes pattern with meaning.[5] For example, the system at Generation 9 in Chain D uses three spellings (⟨koe⟩, ⟨ko⟩, and ⟨co⟩) to express the shape dimension, yielding a horizontal stripes pattern in the matrix representation. We describe such a system as "redundant" because shape was consistently and reliably expressed by the stem, so the suffix spelling system that emerged in this case conveys no additional information—the suffix simply repeats whichever shape was marked by the stem. Redundant suffix systems are characteristic of the Transmission-only condition, with similar outcomes occurring in Chains C, E, F, and H. We also saw the emergence of fully transparent suffix spelling systems in Chains A, B, I, and J. Chain I, for example, ultimately uses a single spelling, ⟨coe⟩, to represent the /kəʊ/ sound; that is, the written suffix forms make no distinction between shapes or colors, just like the spoken language. Chain G did not settle on a clear pattern, using ⟨co⟩, ⟨coe⟩, and ⟨coh⟩ somewhat interchangeably in the final generation, although there were some signs of a color-expressive system emerging in, for example, Generation 6, where yellow is consistently spelled ⟨coh⟩, blue is consistently spelled ⟨coe⟩, and pink uses both ⟨co⟩ and ⟨coe⟩. The only other signs of color-expressive

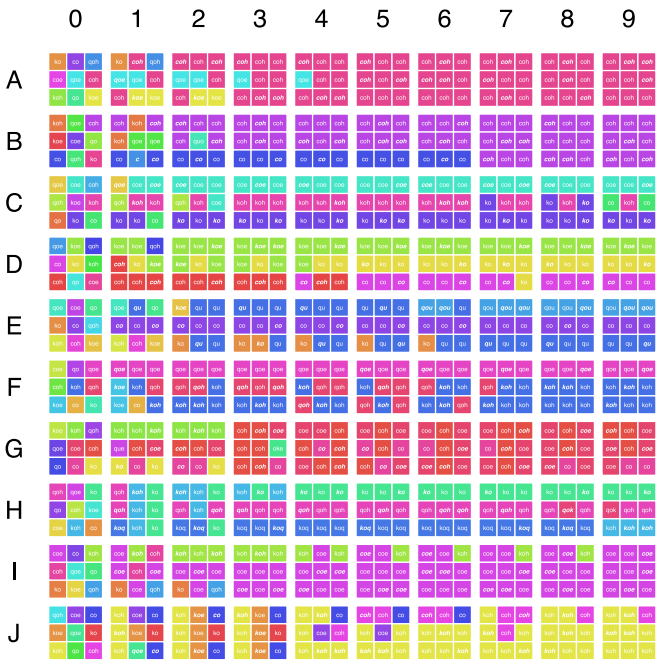

Fig. 4. Results from the Transmission-only condition in Experiment 1 (differentiation). Each matrix shows the suffix spelling system in use at a particular generation (shape on the rows, color on the columns, as in Fig. 2). Chains are labeled A–J and generations are labeled 0–9 (0 is the randomly generated seed system). Each chain uses an independent color palette, with each color representing a particular suffix spelling; similar colors indicate similar spellings. Spellings in bold-italic are the generalizations on unseen items. By the ninth generation, most systems are degenerate (e.g., Chains A, B, and I), redundant (e.g., Chains D, E, and H), or a mixture of the two (e.g., Chain F).

systems are H1, which was rapidly converted into a redundant system in subsequent generations, and J2, which ultimately degenerated toward transparency.

The results for the Transmission + Communication condition (Chains K–T) are shown in Fig. 5. Like Transmission-only, degeneration to a single spelling by the ninth generation is a relatively common outcome (e.g., Chains N, R, and S and to a lesser extent Chains L, M, and O). In contrast, redundant, shape-expressive systems (as indicated by horizontal stripes) are relatively rare (e.g., K8, M6 and Q1–6). Instead, the presence of the communicative task appears to favor color-expressive systems, although these are far from common and often unstable. In particular, there are two main kinds of color-expressive system that emerge. The first are the generalization-based expressive systems: K5, M1, P5, and S2. In these cases, the participant tended to generalize their input in a way that is consistent with expressing color over shape. For example, in K5, pink was consistently spelled ⟨ko⟩ and blue was consistently spelled ⟨co⟩. In the case of M1, color is expressed by the final vowel letters (⟨oe⟩ for pink, ⟨o⟩ for yellow, and ⟨oh⟩ for blue) with the spelling of the /k/ sound conditioned on the stem (⟨buvic-⟩, ⟨zetik-⟩, and ⟨wopiq-⟩). We note, however, that in all these cases the expressive system was not reciprocated by the participant's partner—resulting in low communicative accuracy—and not sustained or elaborated on in subsequent generations.

The second type of color-expressive system is one that simply appropriates the expressive power of English to differentiate color: K9, M7, Q7, and to a lesser extent L7 and O4. In M7, for example, the participant added ⟨r⟩, ⟨g⟩, and ⟨b⟩ (presumably red, gold, blue) to the ends of the words, although the participant's partner only reciprocated the ⟨g⟩ spelling and seemingly failed to understand what was meant by ⟨r⟩ and ⟨b⟩. In the case of Q7, the pair of participants added ⟨r⟩, ⟨o⟩, and ⟨b⟩ (presumably red, orange, blue) to communicate with high accuracy, but, although this system was retained into Generation 8, it started to

[5] The color-coding was generated independently for each chain by selecting *n* evenly-spaced hues, where *n* is the number of unique forms that emerged across the chain, and mapping these hues onto the suffixes in alphabetical order.

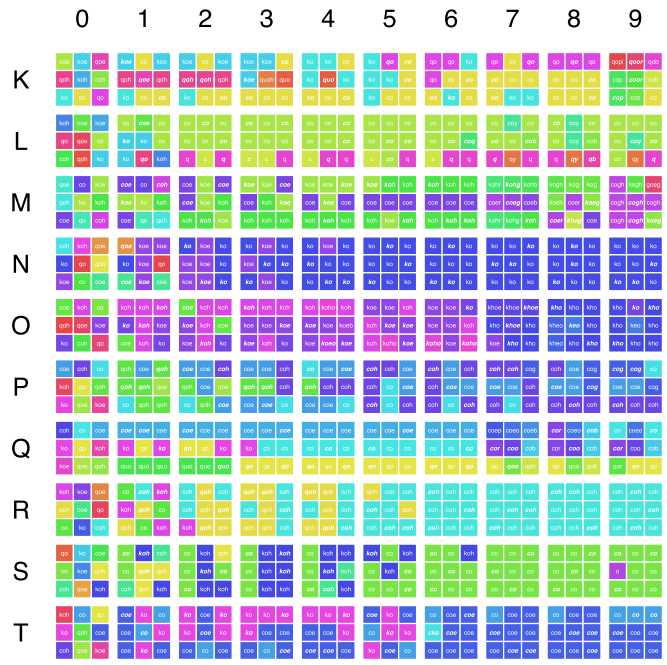

Fig. 5. Results from the Transmission + Communication condition in Experiment 1 (differentiation). Each matrix shows the suffix spelling system in use at a particular generation (shape on the rows, color on the columns, as in Fig. 2). Chains are labeled K–T and generations are labeled 0–9 (0 is the randomly generated seed system). Each chain uses an independent color palette, with each color representing a particular suffix spelling; similar colors indicate similar spellings. Spellings in bold-italic are the generalizations on unseen items. There are some isolated examples of differentiation through implicit generalization (K5, M1, P5, S2) and explicit innovation (K9, L7, M7, O4, Q7).

disintegrate by Generation 9. Overall, in the five cases where English color letters were used, the systems did not really catch on, perhaps because they retained spelling redundancy from the previous generation, resulting in unnecessarily complex suffix spellings that were too difficult to learn for subsequent generations. Q7, for example, uses ⟨-coe-⟩, ⟨-co-⟩, ⟨-qo-⟩ for shape plus ⟨-r⟩, ⟨-o⟩, ⟨-b⟩ for color. In addition to this handful of cases, there were a further four pairs of participants who used full English words as the suffix (typically ⟨-pink⟩ or ⟨-red⟩, ⟨-oran⟩ or ⟨-yell⟩, and ⟨-blue⟩), but these pairs were excluded and replaced before iteration to the next generation.[6] We mention these cases, however, because they are still examples of a conscious effort to differentiate—something that never happened in the Transmission-only condition.

To analyze more formally what types of system tend to emerge under different conditions, we first designated four suffix systems that are of particular interest: holistic, expressive, redundant, and degenerate. These four typological categories may be positioned along the simplicity–informativeness continuum in terms of the number of suffix

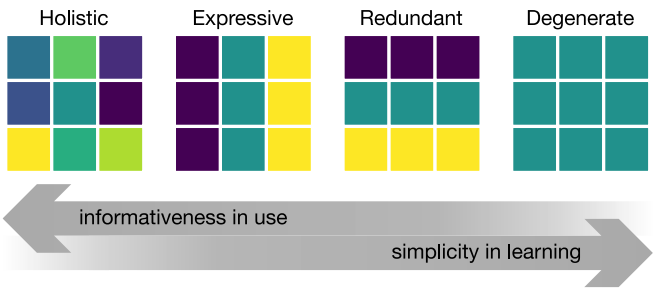

Fig. 6. Four primary systems of interest (or typological categories) arranged along the cost/complexity continuum. A holistic system uses a unique suffix spelling for each shape–color combination. An expressive system only expresses color. A redundant system only expresses shape (which is already conveyed by the stem). A degenerate system expresses nothing.

forms they make use of and how these forms are conditioned on meaning, as illustrated in Fig. 6. The holistic[7] system has nine unique forms each of which expresses a particular color–shape combination. This is complex to learn, but the suffix alone can pick out exactly one meaning. The expressive and redundant systems have three unique forms that express either color (expressive) or shape (redundant). These systems are easier to learn, but only the expressive system is fully informative (when acting in combination with the stem). The degenerate system uses just one suffix form. This makes it trivial to learn, but entirely uninformative. To classify a participant's output into one of these four typological categories, we computed which of the reference systems was most similar in terms of its information content. To do this, we formalized the systems as set partitions (i.e., a partitioning of the universe of meanings into disjoint subsets) with variation of information (Meilă, 2007) defined as the distance metric between any two such partitions.[8] A given participant's output system is then classified into one of the four typological categories based on whichever reference system is closest.

The typological distributions are plotted in Fig. 7A, revealing the proportion of the 10 chains that fall into each typological category at each generation. In Transmission-only, the holistic systems used to initialize the chains are rapidly replaced by redundant systems by Generation 2. The dominance of the redundant category is then gradually eroded as the chains transition to degeneracy. In Transmission + Communication, there is initially a fairly even mix of all four kinds of system, but by Generation 9, degenerate systems tend to be most common. There is also a notable increase in holistic systems emerging in later generations; these are the cases of compositional suffixes that arose through the addition of English color letters on top of redundant suffix spellings (i.e., K9, M1, M7, and Q7). Although there was not much evidence of expressive systems emerging in either condition, it is interesting to note that the communicative condition did seem to disfavor the inexpressive redundant systems.

[6] In our preregistration, we stated that we would reject participants who used English color words. We allowed single letters because these were typically less transparent, especially in cases where the choice of letter (e.g., ⟨g⟩ or ⟨r⟩) did not match how another person might have perceived the color (yellow vs. gold vs. orange or pink vs. red). Indeed, as we see in the results, subsequent generations often did not pick up on the color letters (L9, M8, and Q9).

[7] Our use of the term *holistic* is slightly unusual here, owing to the fact that we are focused on the suffix level and not the word level. By holistic, we only mean that each shape–color combination has a unique suffix form. We do not distinguish between truly holistic suffixes (nine unique suffixes with no structure in how they relate to each other) and compositional suffixes (nine unique suffixes that can be generated from compositional rules, as was the case in, for example, Q8).

[8] Variation of information is a proper metric on set partitions, measuring the amount of information (in bits) that is lost and gained in the transformation of one partition into another. Under this metric, the holistic and degenerate systems are considered very dissimilar because they carry very different levels of information content. The expressive and redundant systems are also considered quite dissimilar because, although they carry the same amount of information, the information they carry is orthogonal (shape in the case of redundant, and color in the case of expressive).

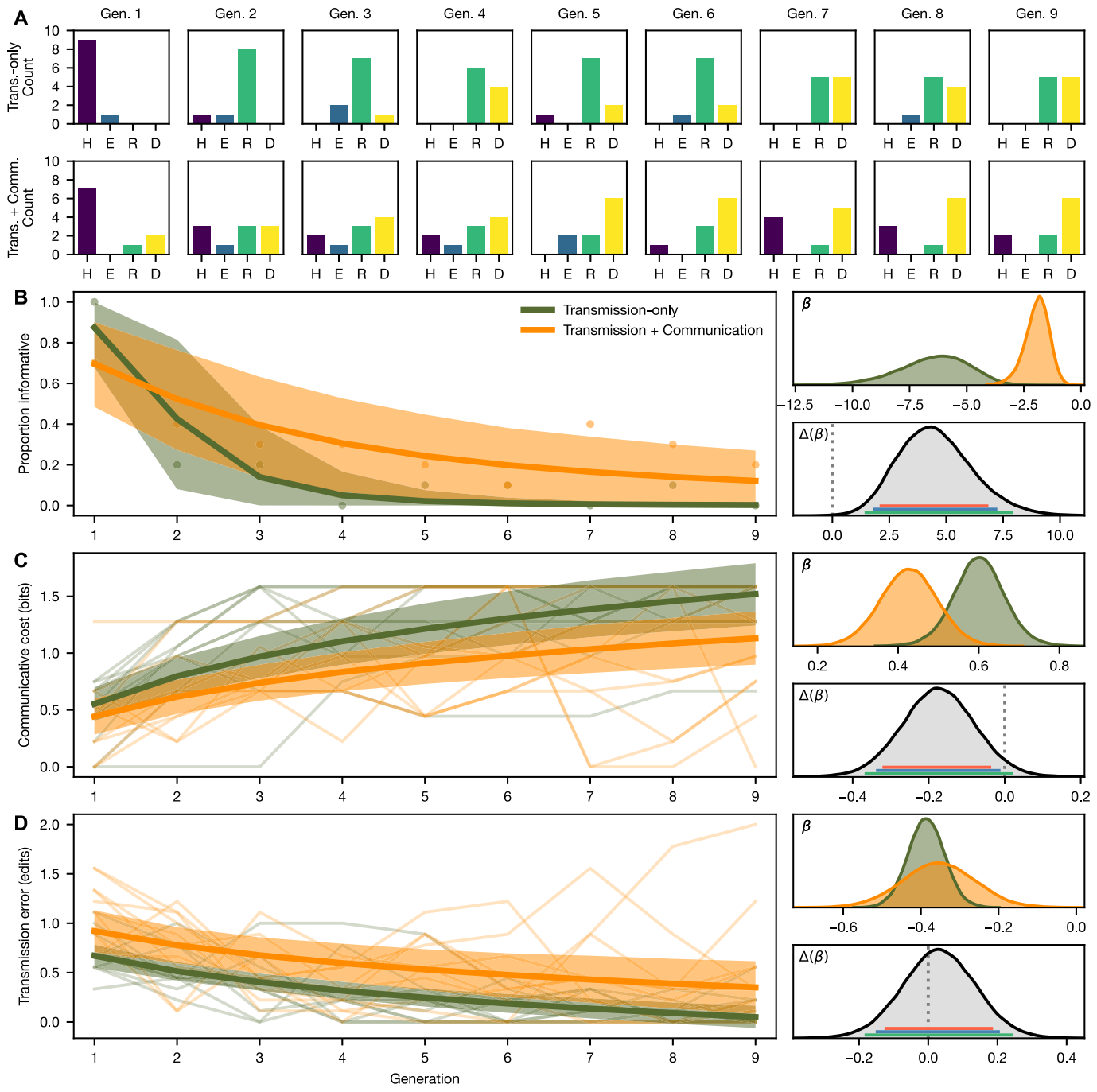

Fig. 7. Results of Experiment 1. **A** Typological distribution by generation and condition over the four typological categories: holistic (H; purple), expressive (E; blue), redundant (R, green), and degenerate (D; yellow). **B** Proportion of systems classified as informative (holistic or expressive) by generation. The dots show the observed proportions and the curves show logistic regression models fit to the data. **C** Communicative cost by generation along with regression models fit to the data. **D** Transmission error by generation along with regression models fit to the data. The panels on the right show the posterior estimates of the slope (β) parameters by condition as well as the posterior differences between conditions. The green, blue, and red bars indicate respectively the 95%, 90%, and 85% HDIs (credible intervals).

To analyze how informativeness changes over time and how this compares between the two conditions, we reduced the typological classifications into two broader categories: informative systems (i.e., holistic or expressive) and uninformative systems (i.e., redundant or degenerate). We then fit a Bayesian mixed-effects logistic regression model that predicts whether or not a system is informative as a function

of generation with by-chain random slopes and intercepts,[9] which is the standard model structure used to analyze iterated learning experiments (Winter & Wieling, 2016). Our Bayesian approach produces posterior estimates of two key parameters: α, which represents the intercept of the regression model (i.e., the model estimate of the dependent variable at Generation 0), and β, which represents its slope (i.e., the model estimate of how much the dependent variable changes per generation). To determine if there is a statistical difference between conditions, we compute the difference in slopes, $\Delta(\beta) = \beta_{\text{comm}} - \beta_{\text{trans}}$, and check if this posterior difference satisfactorily rejects zero. In other words, we test to see whether the dependent variable is changing over time more rapidly in one condition compared to the other. Here we follow the convention that the 95% highest density interval (HDI; the narrowest interval that contains 95% of the posterior probability mass) should not include zero. We emphasize, however, that the posterior is a complete description of the evidence (given the data and model assumptions) and does not strictly need to be reduced to a binary yes/no decision. The results are shown in Fig. 7B. In both conditions, there is a decrease in informativeness over time (the slopes are negative), but in the Transmission + Communication condition, the slope is shallower, suggesting that informativeness decreases more slowly in the presence of communicative pressure. The difference in slopes, $\Delta(\beta) = 4.53$ (95% HDI: 1.48, 7.88), clearly rejects zero, pointing to a meaningful difference between conditions.

One issue with the above approach is that collapsing the systems into binary categories (informative vs. uninformative) results in a loss of information about *how informative* the systems are. In addition, there was evidence to suggest that the model was a suboptimal description of the data, since there was also a difference in the α estimates (the intercepts; Table B1 in the supplementary material), which should theoretically be the same (i.e., there should be no difference between conditions at Generation 0). We address these limitations with a second measure of informativeness, communicative cost (Kemp et al., 2018; Kemp & Regier, 2012; Regier et al., 2015), which has previously been used in similar experimental studies (Carr, Smith, Culbertson, & Kirby, 2020; Carstensen, Xu, Smith, & Regier, 2015; Smith, Frank, Rolando, Kirby, & Loy, 2020) and was proposed in our preregistration as the primary measure of informativeness. Communicative cost is an information-theoretic measure that expresses how much information will be lost, on average, when a speaker/writer attempts to convey a meaning to a listener/reader using some shared signaling system. If the system contains no ambiguities (all meanings are expressed by unique signals), communicative cost will be zero bits—that is, zero information will be lost during each attempt to communicate using that system. Communicative cost will take some larger value if the system contains ambiguity. It is given by $\sum_{m \in U} Pr(m)(-\log \Pr(m|s_m))$, where U is the universe of meanings that may be expressed, $Pr(m)$ is the probability that a particular meaning would need to be expressed, and $Pr(m|s_m)$ is the probability that a reader would infer meaning m given that a writer produced signal s for meaning m. In our case, U is the set of nine alien objects, $Pr(m)$ is set to $1/|U|$ (all objects need to be talked about with equal probability), and $Pr(m|s_m)$ is given by $1/|M_s|$, where M_s is the set of meanings labeled s according to the system. The results are shown in Fig. 7C. We used the same mixed-effects linear regression model described above (except that the likelihood is now Gaussian). In line with the previous analysis, cost increases with generation in both conditions but increases more slowly under communicative pressure. However, the support for a difference between conditions was weaker under this more nuanced measure, $\Delta(\beta) = -0.17$ (95% HDI: −0.36, 0.02); although we were not able to reject zero at the 95% level, we were able to reject it at the 92% level (92% HDI: −0.35, −0.01).

Aside from the informativeness of the orthographic systems, we also predicted in our preregistration that the systems would become easier to learn over time in both conditions, albeit for slightly different reasons. In Transmission-only, the orthographic system is expected to become increasingly learnable as it degenerates into a single, transparent suffix form. In Transmission + Communication, the system is expected to become more learnable as the unsystematic, holistic systems transform into other easier to learn systems (notably expressive systems, although—as noted already—expressive systems rarely emerged). Following prior work, we operationalized learnability as transmission error—the amount of error that the participant at Generation i made in reproducing the orthographic system that existed at Generation $i - 1$; transmission error is defined as the mean Levenshtein edit distance between the corresponding orthographic forms in consecutive generations (see e.g., Kirby et al., 2008). These results are plotted in Fig. 7D. In both conditions, the estimates of the β parameters are negative and clearly reject zero, suggesting that the systems do indeed become increasingly learnable over time as hypothesized. Our analysis also suggested that there was little difference between the conditions in terms of how rapidly transmission error decreases over time (i.e., $\Delta(\beta)$ is highly compatible with zero), although we did note that transmission error tended to be a little higher in the communicative condition. This is to be expected because the goal in the Transmission-only condition is to reproduce the forms taught in training, whereas the goal in the communicative condition is to devise a system that permits accurate communication, which necessitates greater change to the system taught in training.

2.3. Summary

Our first experiment tested whether informative, heterographic orthography could emerge through spelling differentiation and whether it would emerge preferentially under communicative pressure. Although there was evidence to suggest that the orthographic systems remain informative for longer under communicative pressure, both conditions ultimately converged on degenerate, uninformative systems and there was little evidence of systematic differentiation in spelling. Other than resorting to English, such differentiation could have been achieved in a number of ways, most obviously through the conditioning of spelling variation on meaning (e.g., ⟨ko⟩ for pink, ⟨co⟩ for yellow, and ⟨qo⟩ for blue), but also through less obvious strategies such as the use of length (e.g., ⟨ko⟩ for pink, ⟨kko⟩ for yellow, and ⟨kkko⟩ for blue) or the use of arbitrary silent letters (e.g., ⟨kox⟩ for pink, ⟨kof⟩ for yellow, and ⟨kom⟩ for blue). Such differentiation was not forthcoming, however, even under communicative pressure. Instead, if spelling variation was conditioned on anything, it was conditioned on shape, resulting in redundant suffix spellings. This result is in stark contrast to most prior experimental iterated learning studies, in which informative, compositional systems do typically emerge, especially under communicative pressure (e.g., Kirby et al., 2015).

So, what was different about the present experiment compared to the large body of prior experimental iterated learning studies? The primary difference was the presence of a spoken language that is decidedly *not* informative about one of the dimensions. Indeed, the point of our experiment is to see whether orthography can resist phonology under sufficient pressure for informativeness. The presence of homophonous suffix forms acts as a cue to participants that the language itself does not mark color and that, therefore, the orthography should also not mark color. In support of this explanation, we conducted an additional experiment during review, which showed that when homophony is removed, the systems tend to resist degeneration in line with prior work. We discuss this experiment in more detail in the Discussion section and in Appendix C of the supplementary material.

Faced with orthographic variation that could be conditioned on

[9] Model: informative ~ log(generation) + (1 + log(generation) | chain). Bernoulli likelihood with logit link function, weakly informative default priors, six chains of 12,000 samples. All statistical models were fit using Bambi 0.13 (Capretto et al., 2022). Parameter estimates and diagnostics are provided in Appendix B of the supplementary material.

either dimension, learners appear to rule out the possibility that it might be conditioned on color, since such a hypothesis would be in conflict with the spoken language. As a result, any variation in spelling comes to be associated with particular stems, resulting in the emergence of redundant suffix spellings that serve no real purpose. Similar outcomes have been noted before in the context of artificial language learning experiments (Smith & Wonnacott, 2010), and a rough analog can be found in English in the spelling of /-ʃən/ (⟨cian⟩, ⟨cion⟩, ⟨sion⟩, ⟨ssion⟩, or ⟨tion⟩), which is conditioned on the stem (e.g., *magician suspicion*, *expulsion*, *transmission*, and *station*) following a complex set of rules (Carney, 1994, pp. 420–421). Interestingly, however, these redundant systems were relatively uncommon under communicative pressure, suggesting that communicating participants recognized the futility of using the suffix to mark shape.

Overall, although the orthographic systems tended to remain slightly more informative under communicative pressure, the emergent orthographies ultimately preferred to transparently encode sound rather than meaning. This finding seems to align with our general experience of the world: If someone decided to start using the spelling ⟨banque⟩ to differentiate the financial institution from river banks, would anyone take that spelling seriously or even understand the intention? Without top-down diktat, it is hard to get spelling differentiation off the ground in the written modality.

3. Experiment 2

We now turn our attention to the conservation model of heterographic homophones: Given that an informative system already exists (both in speech and in writing), does that informative system persist in writing even after the spoken language has degenerated into homophony? And, importantly, does this happen preferentially in the presence of communicative pressure? Our hypotheses were as follows:

1. Under pressure from learning alone, we expect to find that orthography will track the spoken form of the language, becoming increasingly degenerate as homophony increases.
2. Under additional pressure for disambiguation, we expect the orthography to conserve archaic (but informative) spelling distinctions even after these distinctions cease to exist in the spoken form of the language.

3.1. Methods

The methods were identical to Experiment 1 with two exceptions: The artificial language used to seed each chain started out fully compositional (in both its spoken and written forms), and two sound mergers were artificially induced during cultural transmission, resulting in the spoken forms of the suffixes becoming increasingly homophonous and uninformative over time. This was designed to model the historical processes of sound change and conservation described in the Introduction.

3.1.1. Participants

The experiment was completed by 297 native-English participants recruited through Prolific. The payment and bonusing scheme was identical to Experiment 1 (median bonus: £0.78). The median completion time was 15 m with a median hourly rate of £7.99 (£10.68 including bonus). 17 participants were excluded because they (or their partners) failed the auditory attention checks (15) or used English color words (2). A further 10 participants were lost to communication-game pairing failures. Like Experiment 1, the final dataset comprises 270 participants: 90 in the Transmission-only condition (10 chains of 9 participants) and 180 in the Transmission + Communication condition (10 chains of 9 pairs of participants).

3.1.2. Stimuli

The alien objects and word stems were identical to Experiment 1. Unlike Experiment 1, however, the transmission chains were seeded with a fully compositional language that used three distinct suffixes to systematically express each of the colors. A separate set of suffixes was created for each chain by concatenating a randomly drawn consonant from {/f/, /s/, /ʃ/} and a randomly drawn vowel from {/ə/, /ɛɪ/, /əʊ/}, both without replacement. For example, one chain might use the suffixes /fəʊ/, /ʃə/, /sɛɪ/ to represent the colors pink, yellow, and blue, while another chain might use /sə/, /fɛɪ/, /ʃəʊ/ for those colors. The initial orthographic system was transparent and based on the following phoneme–grapheme mapping: {/f/→⟨f⟩, /s/→⟨s⟩, /ʃ/→⟨x⟩, /ə/→⟨a⟩, /ɛɪ/→⟨ei⟩, /əʊ/→⟨oe⟩}. The suffixes were designed to be distinctive (and therefore easy to memorize and associate with colors), but also similar enough to (mostly) allow for somewhat plausible sound mergers and result in somewhat plausible spellings following sound merger (e.g., it is plausible that /f/ might supplant /s/ or /ʃ/ in speech or that /ʃ/ might be spelled ⟨s⟩ or ⟨x⟩ in writing). We attempted to achieve this balance by combining consonants that are very similar with vowels that are very dissimilar, while also avoiding reuse of any sounds present in the stems. The spoken forms were synthesized using the Apple text-to-speech synthesizer (Moira voice).

3.1.3. Sound change

Each transmission chain was run for nine generations, which were divided into three epochs. During Epoch I (Generations 1 to 3), the three spoken suffixes were distinct (as described above), allowing the spoken language to express all three colors without ambiguity. During Epoch II (Generations 4 to 6), the spoken language had two distinct suffix forms, reducing its informativeness. During Epoch III (Generations 7 to 9), all three spoken suffixes were homophonous, just as in Experiment 1, making the spoken language entirely uninformative about color. This was achieved through two sets of sound changes, the first occurring in the transition from Epoch I to II and the second occurring in the transition from Epoch II to III. An example is illustrated in Fig. 8. In the first sound change, two of the spoken suffix forms were chosen at random and the consonant from one (chosen at random) was paired with the vowel from the other, resulting in a new suffix form that replaced the original two. In the second sound change, the remaining two suffix forms were merged in the same way, resulting in full homophony. Crucially, the spellings did not automatically change following a sound change event; rather, the orthographic system was free to adapt (or not) in response to the sound changes. Note also that individual participants did not directly experience the sound changes; a Generation 4 participant, for example, would always hear Epoch II sounds, while observing spellings produced by a Generation 3 participant (presumably representing the Epoch I sounds). In reality, of course, sound change is more gradual, with individual speakers experiencing both outgoing and

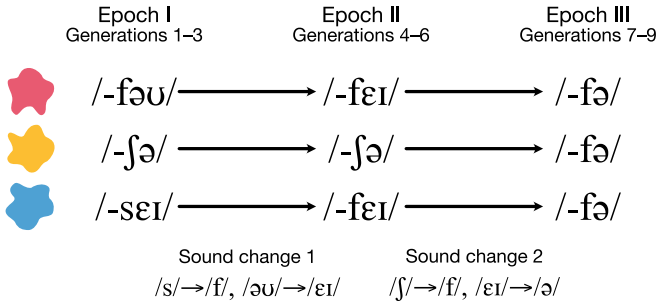

Fig. 8. Examples of the spoken suffixes in Experiment 2. During Epoch I, color is represented by three distinct spoken suffixes. During Epoch II, two of the suffixes are homophonous, reducing the informativeness of the spoken language. During Epoch III, the spoken form of the language makes no color distinction.

incoming spoken forms within their lifetimes.

3.2. Results

The results for all ten chains (labeled A–J) in the Transmission-only condition are shown in Fig. 9. It is immediately clear that color-expressive suffixes (as indicated by vertical stripes) are maintained fairly reliably through the first epoch; perfectly in the case of Chains A, B, D, F, and I, and with some errors in the other five chains. Some of these errors are very minor, such as the use of ⟨sha⟩ instead of ⟨xa⟩ for one item in H3, while others are more catastrophic, such as the early loss of the ⟨sei⟩/⟨xa⟩ distinction in Chain J. Generation 4 represents the first real test of the orthographic systems in the face of sound change, and, in most cases, the Generation 3 systems are preserved quite faithfully (Chains A, B, D, E, H, and I), but by the end of the epoch (Generation 6), many have degenerated into redundant systems that encode shape (Chains B, G, and J) or transparent systems that mirror the Epoch II pattern of homophony (Chains C, E, F, and I). These processes continue into Epoch III and by the ninth generation, all systems have become degenerate, redundant, or some mixture of the two. The one exception is Chain D, whose original spellings were conserved perfectly through to the final generation with only one generalization error in Generation 8, which was quickly reverted in Generation 9.

Like Experiment 1, redundant systems are characteristic of the Transmission-only condition, especially in Epoch III. As the spoken suffixes become more homophonous, the variant spellings are increasingly conditioned on shape rather than color, perhaps because the spoken language signals to learners that the language does not mark color, so they rationalize the system as three words with idiosyncratic spellings. Interestingly, however, all chains exhibited conservation of spelling *form*, even if the way in which spelling was conditioned on meaning was lost. For example, Chain A ultimately represents the sound /fə/ with the spellings ⟨foe⟩ and ⟨xa⟩, spellings that are internally inconsistent and contrary to standard uses of the Latin alphabet, but which trace their origins back to the original seed orthography. Overall, then, the Transmission-only condition in Experiment 2 is characterized by the conservation of spelling *form* without conserving how form patterns with meaning.

The results for the Transmission + Communication condition (Chains K–T) are shown in Fig. 10. Like Transmission-only, the seed systems are mostly maintained faithfully through Epoch I; perfectly in the case of Chains K, L, N, O, and T, and with some errors in the other five chains, although some of these errors are non-destructive changes, such as ⟨x⟩ being replaced with ⟨sh⟩ in R1. Several systems were then maintained through Epoch II, notably Chains O, P, R, and T, and, in one case, through to the end of Epoch III (Chain R, albeit with the original <xei> spelling replaced with ⟨shei⟩). The Chain O system was preserved up to Generation 7, Chain P was maintained up to Generation 7 and almost to Generation 9 with two modifications (⟨sha⟩ instead of ⟨xa⟩ and ⟨fa⟩ instead of ⟨fei⟩), and Chain T was conserved faithfully up to Generation 8. The final form of Chain Q was also fully expressive, albeit through a combination of both conservation and differentiation: The ⟨fei⟩ form was conserved from the seed orthography, the ⟨oxie⟩ spelling appears to derive from a misremembering of ⟨xoe⟩ (partial conservation), and the ⟨fe⟩ spelling, which began as a typo introduced in Q6, seems to have been generalized across the blue items in Q7, perhaps to differentiate them from the yellow items (indeed, the participant's partner made the same generalization). A similar case of differentiation might also have occurred in L9, where the ⟨sol⟩ spelling (originally a typo on ⟨so⟩) was generalized across the yellow items, resulting in a semi-expressive system (although the participant's partner generalized the ⟨sol⟩ spelling across shape).

Unlike Experiment 1, no participant pairs attempted to use English color letters and there was only one case of a pair using English color

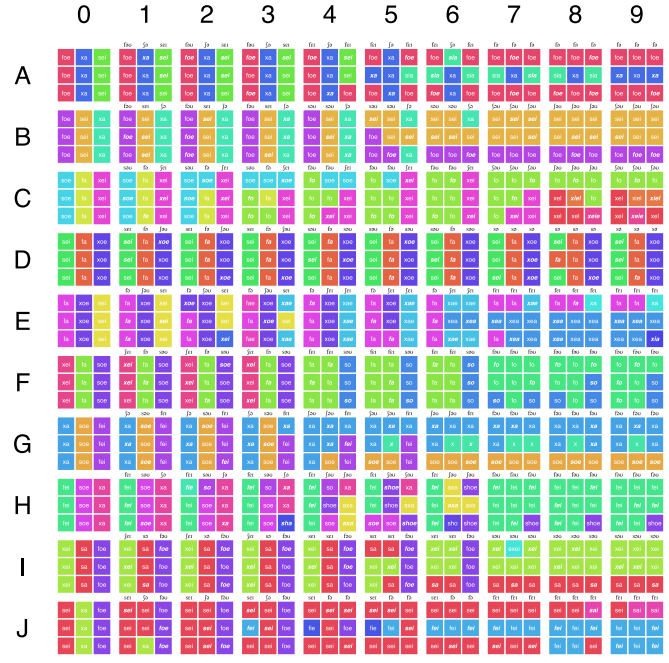

Fig. 9. Results from the Transmission-only condition in Experiment 2 (conservation). Each matrix shows the suffix spelling system in use at a particular generation (shape on the rows, color on the columns, as in Fig. 2). Chains are labeled A–J and generations are labeled 0–9 (0 is the randomly generated seed system). Each chain uses an independent color palette, with each color representing a particular suffix spelling; similar colors indicate similar spellings. Spellings in bold-italic are the generalizations on unseen items. The final systems are characterized by the conservation of form without the conservation of expressivity.

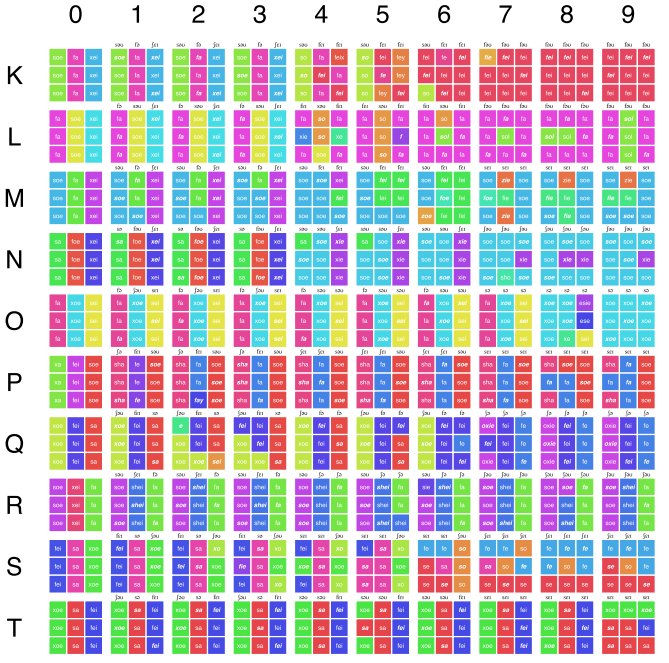

Fig. 10. Results from the Transmission + Communication condition in Experiment 2 (conservation). Each matrix shows the suffix spelling system in use at a particular generation (shape on the rows, color on the columns, as in Fig. 2). Chains are labeled K–T and generations are labeled 0–9 (0 is the randomly generated seed system). Each chain uses an independent color palette, with each color representing a particular suffix spelling; similar colors indicate similar spellings. Spellings in bold-italic are the generalizations on unseen items. Five chains (O, P, Q, R, T) remain fully expressive into the final epoch, in most cases conserving the original forms.

words (Generation 9 of Chain K), although this pair was excluded and replaced (this generation was the only case in which the training input was fully degenerate, which would have resulted in a strong pressure to find a communicative solution in the form of English). Presumably, the general conservation of expressive spelling in Experiment 2 negated the need to innovate novel systems.

Our quantitative analysis of Experiment 2 is identical to that of Experiment 1 with a slight change to the statistical model. Rather than predict the outcome variables as a function of generation, we now predict the outcome variables as a function of the epoch number (1, 2, or 3) and the generation number within the epoch (1, 2, or 3).[10] This yields two slope parameters: β, which represents the effect of epoch, and γ, which represents the additional effect of generational turnover. This model is more appropriate to the Experiment 2 setup, where the experimentally induced homophony results in discontinuities from one epoch to the next, and it allows us to separate out the effect of the homophony pressure from the more general effect of generational turnover.

Fig. 11A plots the typological distributions by generation and condition. Initially, all systems are expressive, but the dominance of this category is gradually eroded over time, particularly during the second and third epochs once the spoken forms had become homophonous. Notably, however, the loss of expressive systems appears to be slower in Transmission + Communication, and redundant systems were also less popular under communicative pressure. As in Experiment 1, we further collapsed the typological categories into two broader categories (informative vs. uninformative) to analyze the trends over time. The results, shown in Fig. 11B, show that the probability of a system being informative drops over time in both conditions (primarily as a function of epoch), but does so more slowly in the communicative condition. Although there was some weak evidence of a difference in epoch slopes ($\Delta(\beta) = 2.65$; 95% HDI: −0.58, 5.95), we could not conclusively reject zero under this first measure of informativeness.

The results in terms of the preregistered measure of informativeness, communicative cost, are presented in Fig. 11C. Here we find a nonzero effect of both epoch (β) and generation (γ) on cost, as well as a difference between conditions in terms of epoch: $\Delta(\beta) = -0.2$ (95% HDI: −0.4, −0.003). Like the previous measure, the γ slopes were in close alignment, so $\Delta(\gamma)$ is highly compatible with zero difference. This suggests that the effect of generational turnover is very similar between conditions and that the difference between conditions is mostly driven by the increases in homophony induced in each epoch. The overall result is that, in Transmission + Communication, the increase in cost is linear across the nine generations, whereas in the Transmission-only condition, the increase in cost follows something more akin to a step function, with sudden increases in cost in response to each additional bout of homophony. In other words, in the Transmission-only condition, the orthographic systems respond rapidly to the changing spoken forms, while in the Transmission + Communication condition, the orthographic systems are more resistant to the homophony.

For completeness, Fig. 11D also plots transmission error; however, we did not hypothesize any particular differences in learnability in Experiment 2, either over time or by condition. The expressive orthographic systems used to initialize the chains start out very easy to learn, and learnability remains fairly consistent throughout the experiment in both conditions, albeit with some constant level of change over time as the systems gradually come into alignment with the spoken forms.

3.3. Summary

Experiment 1 asked whether an informative, heterographic orthography may be *created* de novo under pressure from homophony. Experiment 2, by contrast, asked whether an informative, heterographic orthography can simply be *maintained*, even under the same levels of homophony encountered in Experiment 1. In the Transmission-only condition, only one chain (Chain D) remained expressive into the fully homophonous Epoch III, while in Transmission + Communication, five chains (O, P, Q, R, and T) remained expressive, albeit not necessarily all the way to Generation 9. The fact that expressive spellings persisted longer and across more chains under communicative pressure suggests that an informative orthography—despite running contrary to the spoken language—may be maintained when it serves a useful purpose. That being said, the fact that informativeness could, in principle, be maintained without communicative pressure (most notably in Chain D) suggests that a strong communicative pressure is not a strictly necessary condition for conservation: Learning alone can, to a limited extent, maintain informative heterography.

Many of the chains did, however, eschew informativeness entirely in favor of greater transparency, and the inevitable long-term consequence for all chains appears to be degeneracy. This is to be expected under Transmission-only, where the systems are adapting under learnability pressure, but is somewhat surprising in Transmission + Communication. Our findings ultimately suggest that, in the long term, alphabetic orthographic systems might favor the faithful encoding of speech over the useful encoding of meaning, although there may exist brief windows of time during which informative heterography can resist the spoken language. Interestingly, although participants were resistant to encoding into writing something that is not encoded in speech, they were—at the same time—content to conserve spelling forms that were internally inconsistent and unusual. Tradition has a powerful hold over writing systems.

4. Discussion

The written and spoken forms of a language are never perfectly identical; they diverge in many ways as a result of the differing constraints relevant to each. Spacing between words for example does not exist in speech but constitutes a useful innovation in writing that permits rapid reading (Rayner, Fischer, & Pollatsek, 1998; Sainio, Hyönä, Bingushi, & Bertram, 2007; Zang, Liang, Bai, Yan, & Liversedge, 2013). Similarly, the consistent spelling of affixes, such as the English past-tense marker ⟨-ed⟩, which diverges from its spoken realization (/d/, /t/, or /ɪd/ depending on the preceding sound), permits faster access to meaning (Ulicheva et al., 2020). Might it be the case that—left to evolve freely—the written form of a language will become better adapted to the needs of writers and readers to the detriment of its alignment with the spoken form of the language? Do writing systems adapt to the affordances and constraints of the written modality (Rastle, 2019)?

We addressed these questions by focusing on the particular case of heterographic homophones—morphemes that sound the same but that are spelled differently. Heterographic homophones permit the written language to be more informative than the spoken language; the spellings ⟨knight⟩ and ⟨night⟩, for example, convey a distinction in meaning that cannot be conveyed in speech without supplying additional information. We investigate whether heterography might arise for functional reasons by experimentally simulating the cultural evolution of orthography under two distinct mechanisms, differentiation and conservation, as described by Berg and Aronoff (2021).

4.1. Experiment 1: Differentiation

In our first experiment, we focused on the differentiation mechanism: Might variant spellings be used to differentiate meanings that are otherwise identical in speech? If so, we would expect levels of spelling differentiation to be greater when there is greater pressure for disambiguation, which we induced through the addition of a communication game. Importantly, the initial randomly generated orthographic systems that we used to seed the transmission chains contained high

[10] Model: dependent ~ epoch + (1 + epoch | chain) + generation_in_epoch + (1 + generation_in_epoch | chain).

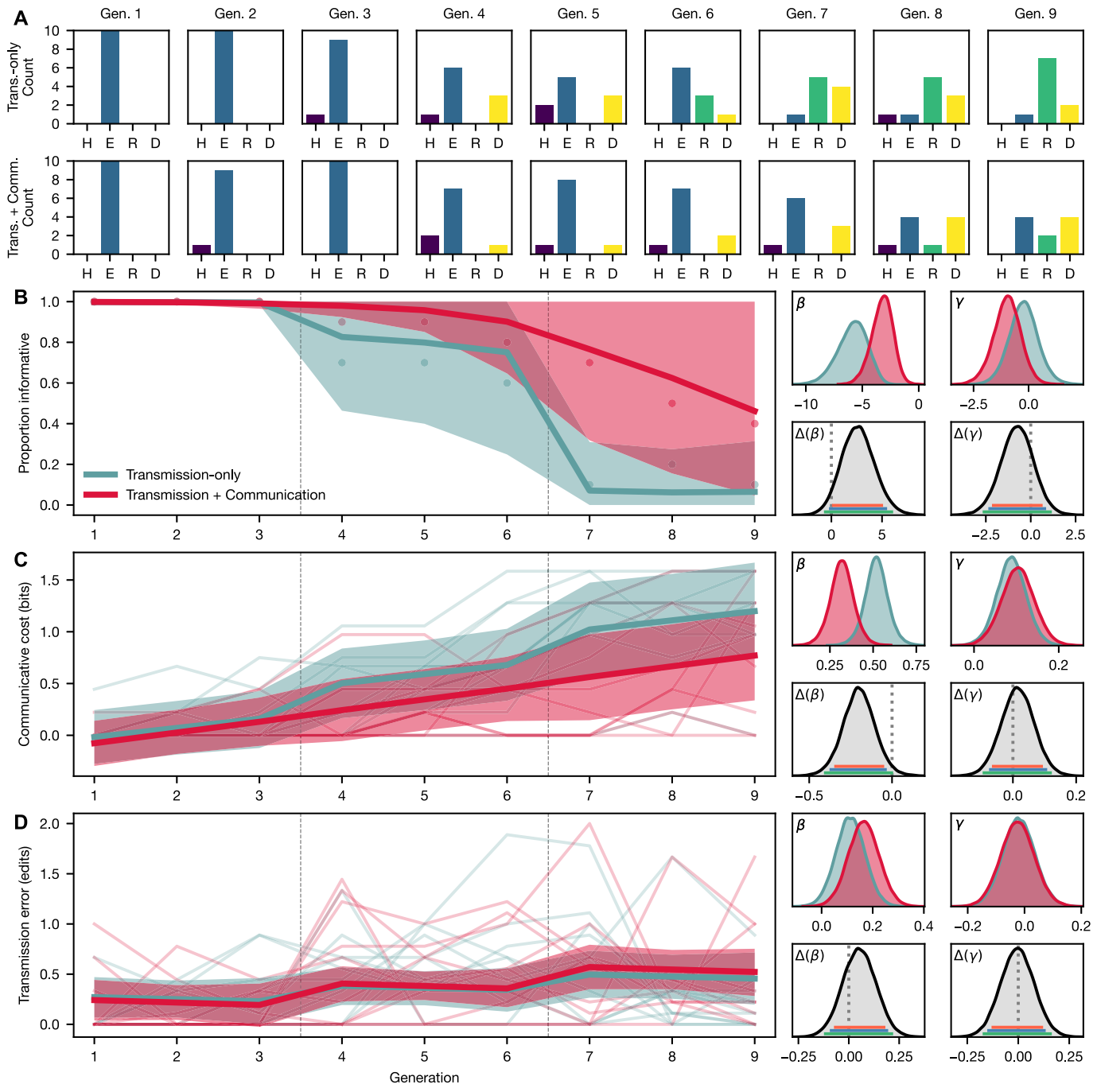

Fig. 11. Results of Experiment 2. **A** Typological distribution by generation and condition over the four typological categories: holistic (H; purple), expressive (E; blue), redundant (R, green), and degenerate (D; yellow). **B** Proportion of systems classified as informative (holistic or expressive) by generation. The dots show the observed proportions and the curves show logistic regression models fit to the data. **C** Communicative cost by generation along with regression models fit to the data. **D** Transmission error by generation along with regression models fit to the data. The panels on the right show the posterior estimates of the slope (β and γ) parameters by condition as well as the posterior differences between conditions. β represents the effect of epoch and γ represents the effect of generation number within epoch. The green, blue, and red bars indicate respectively the 95%, 90%, and 85% HDIs (credible intervals).

variation—that is, multiple ways of spelling the same sound. We included this variation because, for the differentiation mechanism to be viable, the writing system has to be receptive to spelling variation. A writing system that does not permit a one-to-many phoneme-to-grapheme mapping would not be capable of differentiating homophonous words. Only when spelling variation is permitted and available can variants be conditioned on meaning.

Although the emergent orthographies tended to be slightly more informative under communicative pressure, systematic differentiation was rare, unstable, and fleeting, be it through implicit generalization of the supplied variants or explicit innovation of new variants. This is *not* because participants were unable to learn variant spellings; in many cases variant spellings were retained but ineffectually conditioned on shape. Nor was it because participants were unable to learn a system of color marking that is not expressed in the spoken language; we know from Experiment 2 that participants can learn and reproduce such systems. Instead, it seems that, in Experiment 1, differentiation could not get off the ground. We see two reasons for this. First, participants seemed

disinclined to directly encode meaning in how they chose to spell, preferring instead to "write by ear" (Frith, 1979). When asked to type in a word for a pink pentagon called /buvɪkəʊ/, participants were inclined to type a sequence of graphemes that reflected the sound they heard, without encoding meaning. This behavior might be connected to the concept of "functional fixedness" (German & Defeyter, 2000), which states that learners find it difficult to adduce a new function (e.g., writing meaning) when they are accustomed to another function (e.g., writing sound), which highlights a potentially important role for generational turnover in the development of writing systems, since new learners will be more receptive to new functions. Second, even when participants *did* appreciate the need to differentiate the written forms to be successful, they often appeared reluctant or unable to do so, perhaps because they viewed the spelling as immutable or because the problem of aligning with a partner—even in a synchronous setting—was too difficult to overcome without the ability to coordinate over an extended period of time. It is notable, for example, that in the communication games, many attempts to differentiate using English color letters were not reciprocated.

This conclusion is in partial agreement with work by Treiman, Seidenberg, and Kessler (2015). In this study, participants were asked to provide spellings for novel English words (e.g., /hæf/ meaning *alehouse*) that were homophonous with preexisting English words (in this case, *half*). Participants tended to provide the same spelling as the preexisting word (i.e., ⟨half⟩) rather than other possible alternatives that would have had the benefit of differentiating meaning (e.g., ⟨haf⟩, ⟨haff⟩, ⟨haph⟩). The authors argue that participants prefer the "lesser effort that is required to use a familiar whole-word orthographic form compared to that needed for assembling a novel spelling" (p. 544), which aligns with our findings. Treiman et al. (2015) also found, however, that, when given two alternatives to choose from (e.g., ⟨half⟩ vs. ⟨haff⟩), participants generally did prefer the novel spelling. This runs contrary to our first experiment, since our participants are similarly provided with multiple possible spellings of the sound /kəʊ/ (⟨coe⟩, ⟨koh⟩, ⟨qo⟩, etc.), but they nevertheless tended not to condition these on the color dimension, even under communicative pressure with financial incentive. Thus, although the preference for simplicity might be relatively weak at the individual level, it might nevertheless be amplified by the iterated learning process at the population level.

During the review process, a concern was raised that our participants might not have fully understood the communicative nature of the task, thus explaining why we did not observe the emergence of systematic differentiation in spelling. Our position, as outlined above, was that the lack of differentiation was due to the very strong homophony pressure. To test whether the lack of informativeness might be attributed to the homophony and to check that the design of our experiment and implementation of the communicative pressure was sufficient to promote more informative systems, we ran an additional experiment. This experiment was identical to the communicative condition of Experiment 1, except that we removed the spoken forms (thereby replicating previous iterated learning experiments, which are generally only orthographic in nature; e.g., Beckner et al., 2017; Kirby et al., 2008, 2015) and altered the orthographic forms so that no homophony was implied in the spellings. This experiment is described in Appendix C of the supplementary material, but in short, we observed much greater levels of innovation and informativeness in this new experiment, with a clear statistical difference between it and Experiment 1. Broadly, the effect of removing the homophony pressure was that the systems remained more informative compared to Experiment 1 (in terms of both the proportion of systems classified as informative and communicative cost). There was also a commensurate increase in communicative success as a result of the emergence of these more informative systems (a point we return to shortly). Importantly, this suggests that the participants in Experiment 1 did indeed understand the communicative imperative, but nevertheless preferred their spellings to encode sound rather than meaning. Or, rather, the cultural evolutionary process ultimately tended to favor simplicity over informativeness in this particular domain.

4.2. Experiment 2: Conservation

In our second experiment, we tested a different mechanism by which orthographies may end up possessing additional informativeness beyond that of the spoken form of the language: conservation. Under this mechanism, expressive forms do not emerge but are simply fossils representing an earlier form of the spoken language that *was* expressive of a particular meaning distinction. Over time, the orthography may experience a ratcheting effect, in which heterographic forms accumulate (due to successive sound changes) but rarely recede (due to the informativeness they provide). Over longer periods of time, this mechanism might even shift an orthography from a phonographic principle to a logographic one. This parallels what we know about many of the heterographic homophones in English, which arose as byproducts of either the preservation of etymology or phonological changes that were never assimilated into written forms (Berg & Aronoff, 2021), and which are sometimes argued to give English a semi-logographic character (Chomsky & Halle, 1968; Coulmas, 1991; DeFransis, 1989; Zachrisson, 1931). To be clear, this is not to say that such heterography is nonadaptive or an accident of history; rather, such heterographs may have been preserved precisely *because* of the informativeness they inadvertently provide in reading. Thus, just as the spoken language avoids *sound* mergers that increase ambiguity (e.g., Wedel, Kaplan, & Jackson, 2013), so the written language might likewise avoid *spelling* mergers that increase ambiguity. If correct, this would predict that expressive orthography should be preserved preferentially under communicative pressure.

Our findings did indeed show that informative heterography may be conserved more frequently and for longer periods under communicative pressure for disambiguation. There are two important caveats, however. First, we found that cultural transmission alone—that is, blind learning and reproduction—will result in at least some conservation, not only in form but also in the conditioning of form on meaning. Cases such as Chain D correspond to the "accident of history" explanation: Expressive orthography is preserved not because it serves any useful purpose (recall that in Transmission-only there is no functional need for the language to be informative), but because participants are simply reproducing what they learned, and what they learned has not (yet) placed a significant enough burden on learning for simplification to kick in. The additional level of conservation that occurs in Transmission + Communication corresponds to the repurposing explanation; that is, expressive orthography that originally served one purpose (representing speech) is maintained for a new purpose (representing meaning directly). The second caveat is that, in the long term, it appears that transparency might ultimately win the day, even under communicative pressure. Chain O, for example, went from expressive to degenerate in two generations under full homophony pressure, and based on the trajectories of the communicative cost curves (Fig. 11C), it seems likely that all chains will ultimately undergo the same transformation eventually. Informative heterography that arises through conservation is but a temporary oasis on the march toward transparency.

4.3. Differentiation or conservation?

It is important to note, at this point, that the two experiments cannot be compared directly, although we made every effort to keep the two as close as possible. Fundamentally, participants—or more generally, the evolutionary systems—are being asked to do something quite different across the two experiments: create in Experiment 1 and maintain in Experiment 2. The demands of these two tasks are different, and one task or the other may be better suited to our experimental paradigm. However, our experiments do serve to highlight the comparative difficulties involved in differentiation vs. conservation. For *differentiation* to operate, participants must overcome several challenging hurdles: They must

grasp the mechanics of the game and its incentive structure (the apprehension problem), they must be able to put themselves in the shoes of their partners (the theory of mind problem), they must be capable of devising a linguistic solution (the innovation problem), they must be able to align with an interlocutor separated in time and space (the alignment problem), and they must be prepared to rebel against their input, overcoming social stigma in the process (the social problem). Furthermore, once a system has been created, it needs to be reliably transmitted across multiple generations (the learnability problem). The *conservation* of an expressive orthography is, by comparison, plain sailing—it is only the learnability problem that applies.

In general, it might be said that the maintenance of an optimal system is easier than the construction of a new one (see also Smith, 2002). This is made particularly salient by Fig. 12, which compares the experiments in terms of communicative success (the proportion of trials in which the comprehending participant selected the correct target item in response to their partner). In Experiment 1, communicative success remains around chance level (a one in three probability of selecting the right color) because the orthographic systems tend to become uninformative, mirroring the spoken form of the language. In Experiment 2, by comparison, communicative success remains high in several of the chains (i.e., those chains that preserved the expressive system). This suggests that, while it may be difficult for participants to establish an informative writing system, it is comparatively easy to preserve an informative system that offers a clear advantage. It is also interesting to observe what happens to communicative success in Experiment 3 where the homophony pressure is removed. Here, communicative success *does* increase over time, as the participants—unencumbered by having to represent sound—find communicative strategies to differentiate meaning. Taken together, these results suggest that it is difficult for the differentiation mechanism to operate in the face of homophony, but comparatively easy for the conservation mechanism to operate under these same circumstances.

We emphasize, however, that we did not make a-priori predictions about which of the two mechanisms might represent a better theory of the emergence of informative heterography and our experiments were not designed to test the two theories against each other. Instead, we draw this conclusion on the basis that it was difficult for informative heterography to get off the ground in Experiment 1 due to the homophony present in the spoken form (as clarified by Experiment 3), while some degree of informative heterography did persist in Experiment 2 in the face of increasing homophony. A useful way that future work could explore this hypothesis would be to fit models of differentiation and conservation to historical data on spelling change to see which model offers a better fit to the data.

4.4. Limitations

These conclusions must be interpreted within the limitations of these experiments, which are, after all, highly simplified simulacra of real-world processes. Besides the general scaling-down of orthography, phonology, morphology, and semantics to an experimentally tractable test case, one notable issue we faced was how to induce pressure for informativeness in the written modality. We follow a large body of recent studies by using a real-time communication game (e.g., Carr et al., 2017; Kanwal, Smith, Culbertson, & Kirby, 2017; Kirby et al., 2015; Raviv, Meyer, & Lev-Ari, 2018; Saldana et al., 2019; Silvey, Kirby, & Smith, 2019; Winters, Kirby, & Smith, 2015), but such games are not very representative of the dynamics involved in asynchronous written communication (although see Winters & Morin, 2019, for some approaches). That being said, much written communication in the present day is indeed synchronous (e.g., text messaging), potentially allowing the dynamics typically associated with synchronous communication, such as feedback, to play a role in the development of written forms of the language (Lupyan & Dale, 2016).

Another limitation of this work is the extent to which lexical disambiguation really matters in real-world reading scenarios, since the syntactic and semantic context usually makes the meaning clear. If *knight* and *night* were spelled the same way, it is hard to imagine a context in which they might be confused. That being said, cultural transmission has been argued to have a strong amplifying effect on small cognitive biases (Thompson, Kirby, & Smith, 2016), so perhaps even a minor benefit in reading could have a large effect on orthography. An important issue in pursuit of this hypothesis will be to better understand the mechanism by which the biases of readers might place selective pressure on a writing system that is primarily shaped by the needs and preferences of writers, especially given that writing systems are often fixed cultural fossils that do not readily adapt to external pressures. It is also important to note that the functional explanation for heterography advanced here (i.e., ambiguity avoidance) is likely to be one of many. For example, Stenroos and Smith (2016) take the view that English spelling has generally remained opaque with respect to phonology because its primary function was to be a record keeper across time and space. From this perspective, written forms of language resist change because they need to be accessible across decades or centuries and across different jurisdictions or dialect areas.

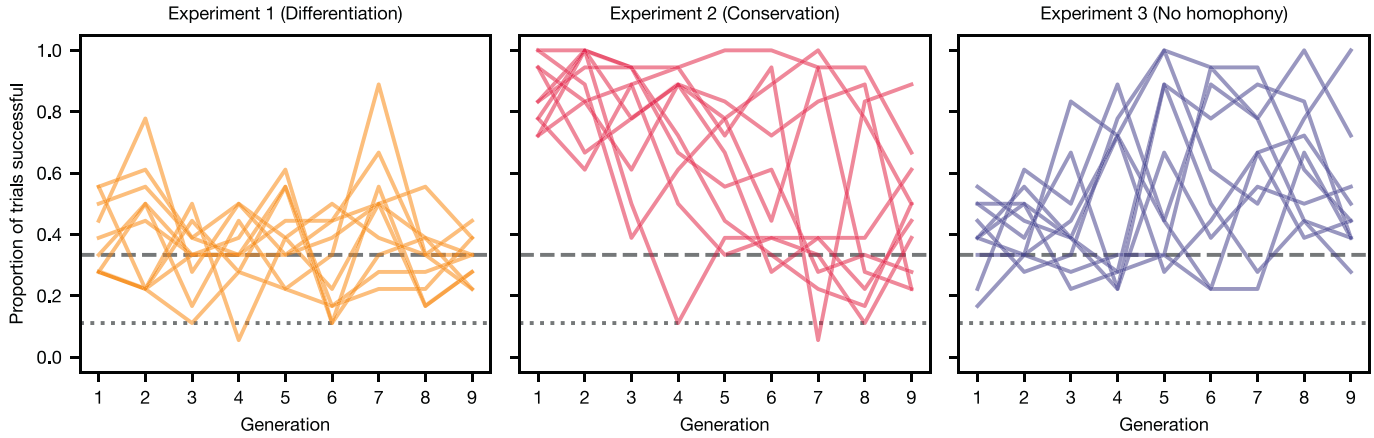

Fig. 12. Communicative success by generation in the communicative conditions of all three experiments. The dotted line shows chance level if the comprehending participant selects from the array of nine items at random (i.e., 1/9), and the dashed line shows chance level if the comprehending participant knows the correct shape but selects color at random (i.e., 1/3). Experiment 3 is a replication of Experiment 1 without any auditory component (see Appendix C in the supplementary material).

The results of both experiments were relatively weak statistically, with the differences in terms of informativeness between the Transmission-only and Transmission + Communication conditions only just (or not quite) meeting the 95% criterion. This is likely to be related to a combination of two factors: a relatively small effect size combined with a relatively small sample size, with only ten chains—ten independent sampling units—in each condition. Our decision to run ten chains per condition was primarily based on norms in the field (e.g., Kempe, Gauvrit, & Forsyth, 2015; Kirby et al., 2008, 2015; Raviv et al., 2018; Roberts & Fedzechkina, 2018; Smith & Wonnacott, 2010; Tamariz & Kirby, 2015), since we had little idea of what effect size we could expect to find when designing the studies. Nevertheless, the small effect sizes we observed do suggest some risk of type one error and future work with this paradigm would benefit from increasing the number of chains in light of these relatively small effect sizes.

Lastly, one important thing to note is that the participant population we draw from (native English speakers) is already accustomed to heterography and opacity; informative orthography might be even less forthcoming in other populations used to more transparent writing systems. This brings us to a much deeper issue with iterated learning experiments in general: We cannot avoid the fact that our participants come into the lab with prior linguistic baggage, whether that baggage is for the encoding of sound or the encoding of meaning. Ideally, our experiments would be performed with participants who have no writing experience at all, but since that would be very difficult to achieve, perhaps the second-best option is a participant population that is relatively open-minded to both types of writing systems. In this sense, our use of English speakers is actually quite appropriate, since the English writing system is neither fully phonographic nor fully logographic.

5. Conclusion

It has long been known that heterography makes reading and learning to read difficult (Pexman, Lupker, & Jared, 2001; Seymour et al., 2003). As a result, heterography has often been derided as a source of unnecessary complexity, and the orthographic reforms implemented in many languages have tended to focus on its elimination. However, recent research suggests that heterography may in some circumstances be functional because it permits rapid access to meaning (Rastle, 2019; Ulicheva et al., 2020). The novel research presented in this article suggests that the cultural evolution of writing systems may prefer to trade some simplicity for greater informativeness when the communicative need for disambiguation is strong enough. These results imply that writing systems may, under some circumstances, evolve to fill a "reading niche." However, our research also shows that creating heterography, and even maintaining it, is challenging given the demands it poses on learning. These findings raise the prospect of a third major issue relevant to the cultural evolution of writing systems: education. Instead of yielding to the pressure of learnability, orthographies like English and Chinese have developed and maintained a high degree of informativeness because those societies have invested in education systems that spend many years teaching children to read (e.g., Wu, Li, & Anderson, 1999). Thus, informative writing systems that contribute to rapid, skilled reading may not only impose learning costs, but may also require ongoing economic investment.

CRediT authorship contribution statement

Jon W. Carr: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Kathleen Rastle:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data availability

The data and code are available from https://osf.io/7auw6/.

Acknowledgement

This work was funded by a Leverhulme Trust Research Project Grant (grant number: RPG-2020-034) awarded to KR.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105809.

References

Bailes, R., & Cuskley, C. (2023). The cultural evolution of language. In J. J. Tehrani, J. Kendal, & R. Kendal (Eds.), *The Oxford handbook of cultural evolution (pp. C59P1–C59S14)*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198869252.013.59.

Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution, 2*(2), 160–176. https://doi.org/10.1093/jole/lzx001

Berg, K., & Aronoff, M. (2017). Self-organization in the spelling of English suffixes: The emergence of culture out of anarchy. *Language, 93*(1), 37–64. https://doi.org/10.1353/lan.2017.0000

Berg, K., & Aronoff, M. (2021). Is the English writing system phonographic or lexical/morphological? A new look at the spelling of stems. *Morphology, 31*(3), 315–328. https://doi.org/10.1007/s11525-021-09379-5

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Bolinger, D. L. (1946). Visual morphemes. *Language, 22*(4), 333–340. https://doi.org/10.2307/409923

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

Buchholz, W. (1977). The word "byte" comes of age. *Byte, 2*(2), 144.

Canini, K. R., Griffiths, T. L., Vanpaemel, W., & Kalish, M. L. (2014). Revealing human inductive biases for category learning by simulating cultural transmission. *Psychonomic Bulletin & Review, 21*(3), 785–793. https://doi.org/10.3758/s13423-013-0556-3

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2022). Bambi: A simple interface for fitting Bayesian linear models in Python. *Journal of Statistical Software, 103*(15), 1–29. https://doi.org/10.18637/jss.v103.i15

Carney, E. (1994). *A survey of English spelling*. Routledge.

Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science, 41*(4), 892–923. https://doi.org/10.1111/cogs.12371

Carr, J. W., Smith, K., Culbertson, J., & Kirby, S. (2020). Simplicity and informativeness in semantic category systems. *Cognition, 202*, Article 104289. https://doi.org/10.1016/j.cognition.2020.104289

Carstensen, A., Xu, J., Smith, C. T., & Regier, T. (2015). Language evolution in the lab tends toward informative communication. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual conference of the Cognitive Science Society (pp. 303–308)*. Cognitive Science Society.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Coulmas, F. (1991). *The writing systems of the world*. Blackwell.

Crystal, D. (2005). *The stories of English*. Penguin.

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

DeFransis, J. (1989). *Visible speech: The diverse oneness of writing systems*. University of Hawaii Press.

Frith, U. (1979). Reading by eye and writing by ear. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 379–390). Plenum Publishing. https://doi.org/10.1007/978-1-4684-0994-9_23.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841

Gabelentz, G. (1891). *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse. T.O.* Weigel Nachfolger.

German, T. P., & Defeyter, M. A. (2000). Immunity to functional fixedness in young children. *Psychonomic Bulletin and Review, 7*(4), 707–712. https://doi.org/10.3758/BF03213010

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf`s law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition, 165*, 45–52. https://doi.org/10.1016/j.cognition.2017.05.001

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science, 336*(6084), 1049–1054. https://doi.org/10.1126/science.1218811

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

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition, 136*, 247–254. https://doi.org/10.1016/j.cognition.2014.11.038

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review, 24*, 118–137. https://doi.org/10.3758/s13423-016-1166-7
Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the USA, 105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105
Kirby, S., Griffiths, T. L., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology, 28*, 108–114. https://doi.org/10.1016/j.conb.2014.07.014
Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87–102. https://doi.org/10.1016/j.cognition.2015.03.016
Korochkina, M., Marelli, M., Brysbaert, M., & Rastle, K. (2024). The Children and Young People's Books Lexicon (CYP-LEX): A large-scale lexical database of books read by children and young people in the United Kingdom. *Quarterly Journal of Experimental Psychology.*. https://doi.org/10.1177/17470218241229694
Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of North American English*. De Gruyter Mouton.
Lass, R. (2000). Phonology and morphology. In R. Lass (Ed.), *Vol. 3. The Cambridge history of the English language* (pp. 56–186). Cambridge University Press. https://doi.org/10.1017/CHOL9780521264761.004.
Lupyan, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences, 20*(9), 649–660. https://doi.org/10.1016/j.tics.2016.07.005
Martinet, A. (1952). Function, structure, and sound change. *Word, 8*(1), 1–32. https://doi.org/10.1080/00437956.1952.11659416
Meilă, M. (2007). Comparing clusteringsan information based distance. *Journal of Multivariate Analysis, 98*(5), 873–895. https://doi.org/10.1016/j.jmva.2006.11.013
Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., & Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences, 118*(49), Article e2025993118. https://doi.org/10.1073/pnas.2025993118
Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition, 192*, Article 103964. https://doi.org/10.1016/j.cognition.2019.05.001
Motamedi, Y., Smith, K., Schouwstra, M., Culbertson, J., & Kirby, S. (2021). The emergence of systematic argument distinctions in artificial sign languages. *Journal of Language Evolution, 6*(2), 77–98. https://doi.org/10.1093/jole/lzab002
Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book language and its implications for children's language, literacy, and development. *Current Directions in Psychological Science, 31*(4), 375–380. https://doi.org/10.1177/09637214221103264
Nevalainen, T. (2012). Variable focusing in English spelling between 1400 and 1600. In S. Baddeley, & A. Voeste (Eds.), *Orthographies in early modern Europe* (pp. 127–165). De Gruyter Mouton. https://doi.org/10.1515/9783110288179.127.
Parkes, M. B. (1992). *Pause and effect: An introduction to punctuation in the West*. Ashgate Publishing.
Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 139–156. https://doi.org/10.1037/0278-7393.27.1.139
Rastle, K. (2019). EPS mid-career prize lecture 2017: Writing systems, reading, and language. *Quarterly Journal of Experimental Psychology, 72*(4), 677–692. https://doi.org/10.1177/1747021819829696
Raviv, L., Meyer, A., & Lev-Ari, S. (2018). Compositional structure can emerge without generational transmission. *Cognition, 182*, 151–164. https://doi.org/10.1016/j.cognition.2018.09.010
Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research, 38*(8), 1129–1144. https://doi.org/10.1016/S0042-6989(97)00274-5
Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney, & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 237–263). John Wiley & Sons. https://doi.org/10.1002/9781118346136.ch11.
Reis, A., Araújo, S., Morais, I. S., & Faísca, L. (2020). Reading and reading-related skills in adults with dyslexia from different orthographic systems: A review and meta-analysis. *Annals of Dyslexia, 70*(3), 339–368. https://doi.org/10.1007/s11881-020-00205-x
Roberts, G., & Fedzechkina, M. (2018). Social biases modulate the loss of redundant forms in the cultural evolution of language. *Cognition, 171*, 194–201. https://doi.org/10.1016/j.cognition.2017.11.005
Rosch, E. H. (1978). Principles of categorization. In E. H. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum.
Saenger, P. (1997). *Space between words: The origins of silent reading*. Stanford University Press.
Sainio, M., Hyönä, J., Bingushi, K., & Bertram, R. (2007). The role of interword spacing in reading Japanese: An eye movement study. *Vision Research, 47*(20), 2575–2584. https://doi.org/10.1016/j.visres.2007.05.017
Saldana, C., Kirby, S., Truswell, R., & Smith, K. (2019). Compositional hierarchical structure evolves through cultural transmission: An experimental study. *Journal of Language Evolution, 4*(2), 83–107. https://doi.org/10.1093/jole/lzz002
Sandra, D., Ravid, D., & Plag, I. (2024). The orthographic representation of a word's morphological structure: Beneficial and detrimental effect for spellers. *Morphology, 34*, 103–123. https://doi.org/10.1007/s11525-024-09424-z
Scragg, D. G. (1974). *A history of English spelling*. Manchester University Press.
Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174. https://doi.org/10.1348/000712603321661859
Shankweiler, D., & Lundquist, E. (1992). On the relations between learning to spell and learning to read. In *, Vol. 94. Advances in psychology* (pp. 179–192). Elsevier. https://doi.org/10.1016/S0166-4115(08)62795-8.
Silvey, C., Kirby, S., & Smith, K. (2019). Communication increases category structure and alignment only when combined with cultural transmission. *Journal of Memory and Language, 109*, Article 104051. https://doi.org/10.1016/j.jml.2019.104051
Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science, 14*(1), 65–84. https://doi.org/10.1080/09540090210164306
Smith, K. (2022). How language learning and language use create linguistic structure. *Current Directions in Psychological Science, 31*(2), 177–186. https://doi.org/10.1177/09637214211068127
Smith, K., Frank, S., Rolando, S., Kirby, S., & Loy, J. E. (2020). Simple kinship systems are more learnable. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the Cognitive Science Society (pp. 801–807)*. Cognitive Science Society.
Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition, 116*, 444–449. https://doi.org/10.1016/j.cognition.2010.06.004
Spencer, L. H., & Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology, 94*(1), 1–28. https://doi.org/10.1348/000712603762842075
Stenroos, M., & Smith, J. J. (2016). Changing functions: English spelling before 1600. In V. Cook, & D. Ryan (Eds.), *The Routledge handbook of the English writing system* (pp. 125–141). Routledge. https://doi.org/10.4324/9781315670003.ch08.
Tamariz, M. (2017). Experimental studies on the cultural evolution of language. *Annual Review of Linguistics, 3*, 389–407. https://doi.org/10.1146/annurev-linguistics-011516-033807
Tamariz, M., & Kirby, S. (2015). Culture: Copying, compression, and conventionality. *Cognitive Science, 39*, 171–183. https://doi.org/10.1111/cogs.12144
Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 60–76. https://doi.org/10.1037/a0020126
Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences of the USA, 113*, 4530–4535. https://doi.org/10.1073/pnas.1523631113
Treiman, R., Seidenberg, M. S., & Kessler, B. (2015). Influences on spelling: Evidence from homophones. *Language, Cognition and Neuroscience, 30*(5), 544–554. https://doi.org/10.1080/23273798.2014.952315
Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition, 195*, Article 103810. https://doi.org/10.1016/j.cognition.2018.09.013
Verhoef, T., Kirby, S., & Boer, B. (2015). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science, 40*, 1969–1994. https://doi.org/10.1111/cogs.12326
Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition, 128*(2), 179–186. https://doi.org/10.1016/j.cognition.2013.03.002
Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge University Press.
Winters, J., & Morin, O. (2019). From context to code: Information transfer constrains the emergence of graphic codes. *Cognitive Science, 43*(3), Article 12722. https://doi.org/10.1111/cogs.12722
Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution, 1*(1), 7–18. https://doi.org/10.1093/jole/lzv003
Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition, 7*, 415–449. https://doi.org/10.1017/langcog.2014.35
Wu, X., Li, W., & Anderson, R. C. (1999). Reading instruction in China. *Journal of Curriculum Studies, 31*(5), 571–586. https://doi.org/10.1080/002202799183016
Zachrisson, R. E. (1931). Four hundred years of English spelling reform. *Studia Neophilologica, 4*(1), 1–69. https://doi.org/10.1080/00393273108586757
Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S. P. (2013). Interword spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 720–734. https://doi.org/10.1037/a0030097
Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences of the USA, 115*, 7937–7942. https://doi.org/10.1073/pnas.1800521115
Zhao, J., Li, T., Elliott, M. A., & Rueckl, J. G. (2018). Statistical and cooperative learning in reading: An artificial orthography learning study. *Scientific Studies of Reading, 22*(3), 191–208. https://doi.org/10.1080/10888438.2017.1414219
Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison–Wesley.