Communication increases category structure and alignment only when combined with cultural transmission

Catriona Silvey[a,1,*], Simon Kirby[a], Kenny Smith[a]

[a] *School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, United Kingdom*

Abstract

The semantic categories labeled by words in natural languages are used for communication with others, and learned by observing the productions of others who learned them in the same way. Do these processes of communication and cultural transmission affect the structure of category systems and their alignment across speakers? We examine novel category systems that emerge from communication, cultural transmission, and both processes combined. Communication alone leads to category systems that vary widely in their communicative effectiveness, and are no more structured or aligned than those created by individuals. When combined with cultural transmission, communication speeds up convergence on a learnable number of structured, aligned categories that are consistently communicatively effective. However, cultural transmission without communication eventually has similar results. Communication appears to be neither necessary nor sufficient for creating semantic category systems that are robustly effective for communication. Furthermore, category systems that emerge from cultural transmission are more aligned across speakers than the systems created by individuals, suggesting that cultural transmission allows individuals to coordinate their semantic systems more effectively than they can through shared perceptual biases alone.

[*] Corresponding author

Email address: `c.silvey@ucl.ac.uk` (Catriona Silvey)

[1] Present address: Division of Psychology and Language Sciences, University College London, London, United Kingdom

Keywords: categories, convexity, alignment, communication, cultural transmission

Introduction

The meanings of words divide the world into semantic categories. These categories vary across languages, but they have a number of common properties. Firstly, they are *specific* to a certain degree: they are not so general that every referent belongs in the same category, nor so fine-grained that every referent belongs in a different category.[2] Secondly, they divide up similarity space in a systematic way. Categories tend to have a "family resemblance" structure, where referents in the same category share graded similarity on many dimensions, and category members that are more similar to other members and less similar to non-members are more prototypical, or representative of the category (Rosch & Mervis, 1975). Seeking a general principle to account for this, Gärdenfors (2000, 2014) builds on the work of Shepard (1987) and Anderson (1991) to show that these properties follow from semantic categories being *convex*[3] regions in conceptual space, a property that maximizes within-category similarity and between-category difference. Thirdly, semantic categories are *aligned* within a speech community: speakers of the same language broadly agree which items belong in a category labeled by a given word.

While categorization behavior is also found in non-human animals (Bergman et al., 2003; Watanabe et al., 1995), the labeled semantic category systems found in human languages are unique in two ways. Firstly, humans use these labeled categories for communication. We define communication as the use of language to reduce uncertainty about a referent. For example, if someone says 'I'm looking for a <u>dog</u>', you may not know their exact goal,

[2] Exceptions are very general superordinate terms such as 'thing' at one extreme, and proper names at the other. However, the majority of semantic categories fall between these levels of specificity and generality.

[3] A convex region is defined as follows: for every pair of points X and Y in a category, every item located between X and Y is also in the category. Gärdenfors shows that, in one direction, the assumption of concepts being convex regions predicts prototype effects, and in the other direction, that a categorization of a conceptual space generated on the basis of prototypes (if each referent in the space is assigned to the category of its closest prototype) will consist of convex regions. See Gärdenfors (2000) for full argument.

but the use of a category label narrows it down to a set of similar possibilities. Secondly, labeled category systems are culturally transmitted. We define cultural transmission as the multi-generational process of individuals learning a language by observing the productions of other individuals who previously learned it in the same way. While other animals acquire categories via individual learning or an innate endowment, human children inherit a culturally perpetuated system of semantic categories by learning from the communicative productions of others. Furthermore, communication and cultural transmission are not simply an alternative delivery system for categories that individuals would otherwise learn on their own. Evidence suggests that these processes lead to the emergence of categories that individuals do not spontaneously invent. An example is terms that label categorical divisions of space, such as the English words 'left' and 'right', or the left/right relations represented by markers in many sign languages. Deaf children raised without sign language input do not spontaneously invent these terms, and they perform poorly in spatial mapping tasks relative to hearing children who have acquired the spatial categories labeled in their native language (Gentner et al., 2013). Over the course of communication and cultural transmission, however, these categories can arise. In the 1970s, when the first cohort of deaf children entered schools for special education in Nicaragua, a new sign language (Nicaraguan Sign Language, or NSL) arose as the children communicated with each other. A second cohort later entered the schools and acquired the language from the first cohort, changing it in the process. Unlike first-cohort signers, second-cohort signers use a consistent system for denoting left/right relations; correspondingly, second-cohort signers outperform first-cohort signers in spatially guided search tasks (Pyers et al., 2010). Experimental work backs up the role of culturally transmitted conventions in category acquisition: children's performance on a search task is selectively predicted by their ability to produce phrases involving the words 'left' and 'right' (Hermer-Vazquez et al., 2001). Thus, communication and cultural transmission not only constitute the mechanism for semantic category acquisition, but have also been shown to lead to the emergence of categories that individuals do not acquire alone.

Given the roles of communication and cultural transmission in the emergence of semantic category systems, are the properties of these systems – specificity, convexity, and alignment – shaped by these processes?

Experimental models of communication and cultural transmission

In recent years, a growing body of research (building on theoretical work by Deacon, 1997 and Christiansen & Chater, 2008 among others) has approached questions of this kind by using artificial language experiments to simulate communication and cultural transmission. In the *dyadic communication* paradigm (reviewed in Galantucci & Garrod, 2011), participants use artificial languages to help their partner identify a target referent from an array of distractors. In the *iterated learning* paradigm (reviewed in Kirby et al., 2014), participants learn artificial languages by observing the productions of a previous learner who acquired the language in the same way. These two methods illuminate how the processes of communication and cultural transmission, respectively, affect the structure of language.

These methods have shown that aspects of linguistic structure can be explained as the result of interacting pressures imposed by communication and cultural transmission. Kirby et al. (2008, 2015) trained participants on artificial languages where labels referred to images that varied on dimensions of meaning, such as shape and texture. The initial languages were unstructured: each image had its own random label, such that labels were not systematically related to meaning dimensions. The languages that emerged from the experiments became structured in different ways, depending on whether they were used for communication, culturally transmitted over generations, or both. Languages that were culturally transmitted without being used for communication became *degenerate*: the number of labels in the language dropped until most images were referred to by the same label. Languages that were used for communication without being culturally transmitted remained *holistic*: each image was referred to by a unique unstructured label. Languages that were both used for communication and culturally transmitted became *compositionally structured*: parts of the labels came to refer to features of the images (e.g., particular shapes and textures), and these parts could be combined to communicate any image.

By modeling each process in isolation and then together, Kirby et al. were able to observe the pressure that each exerts on the structure of language, as well as the synergistic effect of both combined. They concluded that commmunication exerts a pressure for an *expressive* system that maintains distinctions between meanings, while cultural transmission exerts a pressure for a *simple* system that can be concisely cognitively represented and hence easily learned. In isolation, each of these processes results in a language that is optimized for one of these pressures but not the other. Languages that emerge from communication alone are expressive, but not simple; languages that emerge from transmission alone are simple, but less useful for communication. However, both processes together result in a language that is optimized for both pressures. Compositional languages are both simple (since they consist of only a few label parts) and expressive (since these parts can be systematically recombined to communicate any referent). This work suggests the combined processes of communication and cultural transmission as a mechanism for the origin of the compositional structure we see in natural languages.

Communication, cultural transmission, and semantic categories

Can the same processes of communication and cultural transmission explain the specificity, convexity, and alignment of semantic categories? This section will summarize two previous lines of work: the first focusing on alignment and convexity, and the second on convexity and specificity.

Alignment and convexity

A long tradition of research has shown that communication, through the process of conventionalization (Lewis, 1969), leads to semantic representations that are shared or aligned between interlocutors (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987; Garrod & Pickering, 2009; Markman & Makin, 1998; Schwartz, 1995; Steels & Belpaeme, 2005). Some researchers have further suggested that the process of communicative alignment leads to categories that are structured according to similarity. In her *shareability* account, Freyd (1983) argued that communication results in a set of representations grouped by similar values on particular dimensions. According to this account, similarity-based structure (of which convexity is a special case) is a by-product of the need for alignment: speakers make themselves understood by using similarity between items to motivate the establishment of communicative conventions. Results from experiments and models support this account. For example, Markman & Makin (1998) found that communication increased category consistency between individuals and promoted focus on the commonalities of the items being categorized; Jäger & van Rooij (2007) found that signaling games between agents resulted in convex color categories; Voiklis & Corter (2012) found that communication led to better category learning, in part by heightening participants' attention to family resemblance structure. While these results suggest a causal direction from alignment to convexity, Gärdenfors (2000, 2014) suggests the opposite. In Gärdenfors's geometric conceptual spaces, convex concepts are cognitively simple: they can be represented as a set of prototypes, and are hence easy for individuals to learn from few examples.[4] The convex structure of these concepts then makes it easier for speakers and hearers to align their understanding of particular words (Warglien & Gärdenfors, 2011). Thus, previous research suggests a link between convexity and alignment, but accounts disagree on the causal direction: the shareability account predicts that communication should lead to alignment and hence convexity, whereas the conceptual spaces account predicts that simplicity pressures should favor convex categories, which then make alignment easier.

Convexity and specificity

In a more recent line of work, two research groups have illuminated different aspects of how communication and cultural transmission affect the convexity and specificity of semantic categories. Regier and colleagues (Carstensen et al., 2015; Kemp & Regier, 2012; Regier et al., 2007) argue that semantic category systems are an efficient compromise between *simplicity* and *informativeness*. Simple systems have a concise cognitive representation: one way of achieving this is to have fewer categories (i.e. to be less specific). Informative systems maximize the hearer's accuracy in reconstructing a speaker's intended message. The most informative system is the most specific, with the largest number of categories. At a given level of specificity, informativeness is higher when category members are more similar to each other and more different from members of other categories. While Regier and colleagues do not define the structural constraints on informative categories in terms of convexity, convex categories optimally satisfy these constraints. Using simulations, Regier et al. (2015) showed that semantic category systems in the domains of color and kinship are near-optimally informative for their level of simplicity. Following up on this work, Carstensen et al. (2015) asked whether cultural transmission could be a mechanism for the origins of this trade-off between simplicity and informativeness. They trained participants on 4 initially random semantic categories of spatial relations. They tested the participants on these category systems and passed the systems they produced on to the next generation of participants as training input. Over 10

[4] For experimental evidence that human category learners have a bias for convex categories, see Chemla et al. (2019); Landau & Shipley (2001); Pothos & Chater (1997).

generations of iterated learning, the systems gradually became more informative. The authors interpreted this as evidence that cultural transmission could lead to semantic category systems that are near-optimally informative for their level of simplicity.

Carr et al. (2017) also used iterated learning to investigate the effect of cultural transmission on category systems. However, instead of training their participants on 4 random categories, they trained them on input that lacked categories entirely: a set of 48 randomly generated triangles, each referred to by a unique label. These label-triangle pairs were passed down over 10 generations of iterated learning in two conditions: 1) cultural transmission alone, where individual participants learned and produced labels for a set of triangles; 2) cultural transmission + communication, where pairs of participants first learned labels for a set of triangles and then played a dyadic communication game. In the communication game, participants took turns playing the role of sender and receiver. The sender was shown a target triangle and asked to type a label to communicate that triangle to the receiver. The receiver then saw the label and an array of six triangles from which they had to select the target. The authors found that cultural transmission alone led to less specific systems (with fewer categories), while cultural transmission + communication led to more specific systems (with more categories). The authors interpreted this in the light of Kirby et al. (2008, 2015) as further evidence that cultural transmission exerts a pressure for simple systems, while communication exerts a pressure for expressive systems. In both conditions, the categories that emerged were structured by similarity, with groups of triangles referred to by the same label forming generally contiguous regions in similarity space.[5] Following up on this work, Carr et al. (2018) used an agent-based model to examine how learning biases for simplicity and informativeness influence category system structure. They found that a bias for simplicity leads to category systems that are initially more compact (or convex), and later less specific. A bias for informativeness leads to more specific category systems, and a strong bias for informativeness further leads category systems to become more convex. The authors used a two-dimensional stimulus space in which it was possible to differentiate the

[5] Voronoi tessellations from the final generation of the experiments showed that category systems did not become fully convex. The authors did not analyze whether they tended towards greater convexity over generations.

simplest from the most informative category systems, holding specificity constant: simple systems classified the stimuli by just one of the two dimensions, while informative systems clustered stimuli by similarity on both dimensions in maximally convex categories. Comparing the predictions of their model to human learning experiments, they found that human learners have a bias for simplicity, not informativeness. They argued that the results of Carstensen et al. (2015) can be explained as the consequence of a bias not for informativeness, but for simplicity: the categories became more convex due to the first-stage effect of a simplicity bias, and only became more informative as a side-effect of this process. The results summarized in this section suggest that cultural transmission (with or without communication) initially leads to more convex categories, and that cultural transmission combined with communication leads to more specific category systems than cultural transmission alone.

Open questions

The previous work summarized above suggests that communication and cultural transmission can explain some of the properties of semantic categories. However, a crucial gap in this research is an investigation of the effect of communication alone. This is an issue for two reasons. Firstly, the experimental work that found positive effects of communication on category structure and alignment involved participants using their shared native language, making the categories already lexicalized in that language available as common ground. Therefore, this work showed the effect of communication combined with cultural transmission, rather than communication alone. Does communication alone result in aligned categories that are structured by similarity, as hypothesized by Freyd (1983)? Or is the benefit of convex categories for communication a side-effect of their simplicity, as hypothesized by Warglien & Gärdenfors (2011)? Secondly, the apparent conflict between the results of Carstensen et al. (2015) and Carr et al. (2017) requires further investigation. Carr et al. (2017) found that when category systems are culturally transmitted without any communicative task, they become less expressive, or less useful for communication. However, Carstensen et al. (2015) found that category systems become more informative, or more useful for communication, when culturally transmitted without any communicative task. A possible explanation suggested by the agent-based model of Carr et al. (2018) is that Carstensen et al.'s participants are optimizing not for communication, but for learning. However, in Carr et al. (2018)'s two-dimensional space, the simplest category systems are one-dimensional, whereas the most informative category systems are convex. This distinction between simple and informative structures may not hold in a conceptual space with complex multi-dimensional structure (such as the spatial relations of Carstensen et al.). The work of Gärdenfors (2000, 2014) suggests that in this context, the structures Regier and colleagues define as informative (at a given level of specificity) are also simple (i.e., they have a concise cognitive representation). To investigate this, it is important to compare the structure and communicative effectiveness of category systems that result from cultural transmission (with and without communication) to category systems that result from communication alone, where expressivity pressures should be the main influence on category system structure.

Without examining both pressures in isolation and then together (as in Kirby et al. 2008, 2015), we cannot build up a full picture of their separate and combined effects. Furthermore, communication alone, or the process of building up conventions without prior learning of a culturally transmitted system, is relevant to real situations of language emergence, such as that of Nicaraguan Sign Language (described in the Introduction). In Kirby et al. (2008, 2015), communication alone led to a lack of structure; when combined with cultural transmission, it changed the nature of the structure that emerged. Similarly, in the domain of semantic categories, the effects of the two pressures combined could be markedly different from the effect of each in isolation. To investigate this, we need to observe participants creating category systems via communication, cultural transmission, and both combined, in a conceptual space where we can easily quantify our three variables of interest: specificity, convexity, and alignment.

In summary, this paper aims to answer the following questions:

1) Does communication alone create specific, convex, aligned category systems, as compared to baseline category systems created by individuals?

2) Does cultural transmission alone lead category systems to become more or less effective for communication?

3) Do the separate effects of communication and cultural transmission on category systems change when these pressures are combined?

Experiment 1

Methods

The design of Experiments 1 and 2 is summarized in Figure 1.

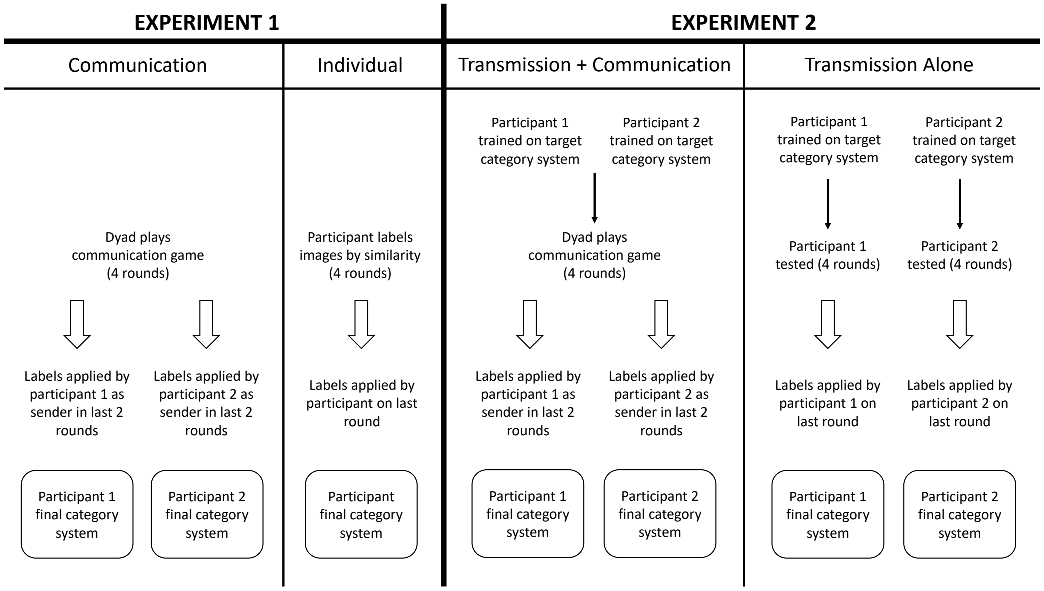

EXPERIMENT 1
EXPERIMENT 2
Communication
Individual
Transmission + Communication
Transmission Alone
Participant 1 trained on target category system
Participant 2 trained on target category system
Participant 1 trained on target category system
Participant 2 trained on target category system
Dyad plays communication game (4 rounds)
Participant labels images by similarity (4 rounds)
Dyad plays communication game (4 rounds)
Participant 1 tested (4 rounds)
Participant 2 tested (4 rounds)
Labels applied by participant 1 as sender in last 2 rounds
Labels applied by participant 2 as sender in last 2 rounds
Labels applied by participant on last round
Labels applied by participant 1 as sender in last 2 rounds
Labels applied by participant 2 as sender in last 2 rounds
Labels applied by participant 1 on last round
Labels applied by participant 2 on last round
Participant 1 final category system
Participant 2 final category system
Participant final category system
Participant 1 final category system
Participant 2 final category system
Participant 1 final category system
Participant 2 final category system

Figure 1: The design of Experiments 1 and 2. In Experiment 1, participants create a category system from scratch, either individually or in pairs through communicative interaction. In Experiment 2, pairs of participants are trained on an existing category system and then either use it in communicative interaction or simply attempt to recall it; category systems are passed from participant to participant in a chain of transmission.

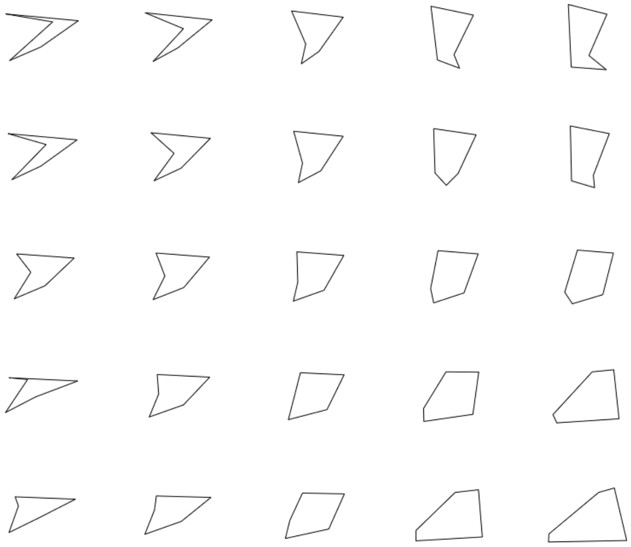


Figure 2: The set of images used as stimuli in the experiments. The four corner images were generated by randomly placing and connecting five vertices. The intermediate images were morphs between these corner images. For each intermediate image, the position of each vertex was set as the average of the positions of the corresponding vertex in each of the four corner images, weighted according to the image's Euclidean distance from each corner.

Experiment 1 compared category systems produced in two conditions: a Communication condition, where pairs of participants created categories over the course of a dyadic communication game, and an Individual condition, where isolated individuals created categories alone. The aim was to compare the specificity, convexity and alignment of the category systems produced in the two conditions. Based on the work summarized above, we predict that communication should lead to systems with higher specificity, higher convexity, and greater levels of alignment within pairs than systems created by individuals.

Stimuli

The stimuli to be categorized were a set of simple images (Figure 2), designed to form a quasi-continuous Euclidean similarity space without clear category boundaries. The corner images that defined the space were randomly generated in order to avoid using familiar shapes which could be easily labeled in participants' native languages.

In all conditions, participants divided the space into categories by applying labels to the images. Images given the same label were considered to be in the same category. Labels were CVCV nonsense words, generated by combining consonants and vowels selected at random from the whole alphabet (under the constraint that they did not form English words). A number of different wordlists were used in order to ensure the results were not dependent on one specific set of labels.

Variables

Category systems in the stimulus space can vary in specificity, convexity, and alignment. Figure 3 shows example systems with corresponding values of these variables.

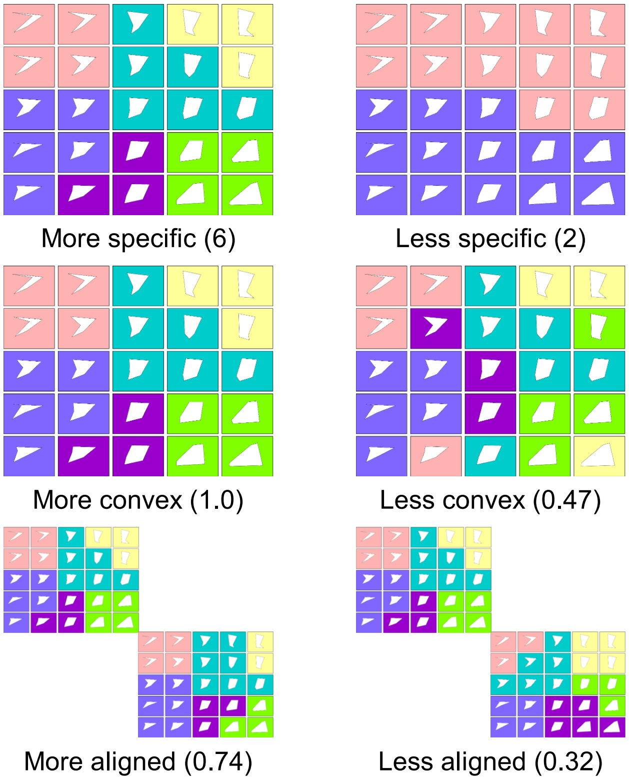
Specificity is the number of categories in a system. Possible values range from 1 (all images are in the same category) to 25 (each image is in its own category).

Convexity is strictly speaking a binary measure (a category is either a convex region of similarity space or it is not). However, since we cannot be sure that participants are categorizing the images using the same similarity gradient as we used to generate them, we do not expect participants to produce absolutely convex category systems. Instead, we are interested in whether category systems in different conditions tend more or less towards convexity. In order to quantify the extent to which a category forms a tightly clustered region in similarity space, we use a measure adapted from Theiler & Gisler (1997). For each image, we calculate the proportion of its neighbors in the Euclidean similarity space that are in the same category. We then average this over all images in the category, then over all categories in the system. Using simulations, we confirmed that the category systems that maximize this index are fully convex.

Convexity can vary between systems of the same specificity, as shown in Figure 3. However, the range of variation of the raw convexity index is constrained by the specificity of the system and the number of images per category. We want to focus on the variation in convexity that is not simply a consequence of these other factors. To do this, we correct our convexity index using the following formula (Hubert & Arabie, 1985):

$$Corrected = \frac{Veridical - Expected}{Maximum - Expected} \tag{1}$$

where $Veridical$ is the true value of the index for the category system,



More specific (6)

Less specific (2)

More convex (1.0)

Less convex (0.47)

More aligned (0.74)

Less aligned (0.32)

Figure 3: The three dependent variables, specificity, convexity, and alignment, illustrated using possible category systems in the experimental stimulus space. Images with the same background color are in the same category. Top row: the system on the left is more specific, with 6 categories, while the system on the right is less specific, with 2 categories. Second row: the system on the left is more convex, with categories that form tightly clustered regions of similarity space, while the system on the right is less convex, with categories that are more discontinuous and dispersed. Third row: the systems on the left are relatively well aligned (the groups of images assigned to the same category are similar across the two systems), while the systems on the right are less aligned (the groups of images assigned to the same category differ substantially across the two systems). Numbers in parentheses are the quantified versions of each variable, calculated as described in the text.

Expected is the value of the index if images were assigned to categories at random (given the number of categories and the number of images per category in the veridical system), and *Maximum* is the highest possible value of the index (given the number of categories and number of images per category in the veridical system). The expected value for a given category system was determined by shuffling the assignment of images to categories 100,000 times, measuring the convexity index each time, and taking the average. The maximum value for a given category system was determined by a stochastic hill-climbing search implemented using a genetic algorithm (see Supplementary Material).

Alignment between two category systems is measured using the Adjusted Rand Index. The original index by Rand (1971) simply calculates the proportion of all unique pairs of images for which the two systems agree on their categorization (i.e., both put them in the same category or both put them in different categories). Similar to convexity, variation in raw alignment is constrained by the number of categories in a system and the number of images per category. Again, we want to measure how aligned two systems are independent of these other factors. The Adjusted Rand Index achieves this using Equation 1. Here, the expected index is calculated by the formula given in Hubert & Arabie (1985), and the maximum index is always 1 (in the case where two category systems are identical). We measured both *alignment* of category systems within pairs of participants, and *convergence*, or the average alignment across all systems produced in each condition. Convergence was calculated via a Monte Carlo procedure, where randomly paired samples were repeatedly drawn from the pool of participants in each condition. The mean alignment of each sample of pairs was collected, and the standard deviation of these means was used to calculate confidence intervals. Importantly, alignment was measured irrespective of the labels used: two participants could use an entirely different set of labels and their systems could still be perfectly aligned, if the images were grouped in the same way.

Communicative success, measured only in the Communication condition, was each pair's score in the experimental communication game, described below.

Participants

Participants were 43 students at the University of Edinburgh (34 female, median age 24). 22 participants (randomly assigned into 11 pairs) took part in the Communication condition. This condition took 1 hour; participants were paid £7, and each member of the pair with the highest communication score was additionally awarded a £10 Amazon voucher. One pair failed to complete the experiment within an hour and were excluded from the analysis. 21 participants took part in the Individual condition. One participant was subsequently excluded due to experimenter error. This condition took 30 minutes. The Individual condition was shorter than the Communication condition because participants did not have to wait for a partner's response on each trial. Participants were paid £3.50. The Linguistics and English Language Ethics Committee of the University of Edinburgh approved the study. All participants provided written informed consent.

Procedure

Full instructions to participants in each condition of each experiment are provided in the Supplementary Material.

Communication Condition. Participants in the Communication condition completed the experiment in pairs, seated in separate cubicles and communicating via computer terminals. In each communication trial, one participant was designated as the sender and one as the receiver. The sender was shown an onscreen array of all 25 images, one of which was highlighted with a red box to indicate it was the target. The positions of images in the array were randomized independently for every trial, ensuring that participants never saw the set of images laid out as in Figure 2. The sender was also presented with one initial label. The sender could reveal a new label at any stage by clicking a "new word" button, up to a maximum of 25 labels. Label sets were identical and revealed in the same order for each participant in a pair. Any labels each sender had revealed on a previous trial remained visible on that sender's screen for all subsequent trials, without any information about which image(s) the sender had previously applied each label to. The sender was instructed to choose a word that would help the receiver pick out the target from the array of images.

Once the sender had picked a label, the receiver was presented with the label and a randomized onscreen array of all 25 images. The positions of images in the array were randomized independently for every trial. The receiver was instructed to select the image they thought the sender had attempted to communicate.

Once the receiver selected an image, both participants were shown a feedback screen. This contained the label the sender had used, the target image, the image the receiver had selected, the score for the trial, and the running total score for the whole experiment. The feedback screen was displayed for 4 seconds. This rich feedback was intended to model a communicative context where dyads can adjust their strategies depending on failure or success.

The maximum communicative success score was awarded if the receiver selected the sender's target image. Success scores then decreased the further away the receiver's selected image was from the target in similarity space. The most communicatively successful category system would be one where every image was in its own category; systems with fewer categories would have lower communicative success on average. The score for each trial was on an ordinal scale, based on the inverse Euclidean distance between the target and the image the receiver selected, from a minimum of 1 (for picking an image at the opposite corner of the space from the target) up to a maximum of 15 (for correctly picking the target).[6] This was based on the assumption that in communication, the cost for being slightly wrong (e.g., thinking 'poodle' means 'Dalmatian') is less than the cost for being very wrong (e.g., thinking 'poodle' means 'ice cream').

After each trial, the sender and the receiver swapped roles. The experiment consisted of 100 trials divided into 4 rounds. Each round featured the 25 images as targets in a pseudo-randomized order. The randomized lists were balanced such that each participant was the sender for every target image once in the first half of the experiment, and once in the second half. Halfway through the experiment, participants had the chance to take an optional break of up to 2 minutes.

We analyzed the final category system produced by each participant in the pair. Each participant's final category system was defined by the label that participant used the last time they were the sender for each image (i.e., during the last 2 rounds). Images given the same label were considered to be in the same category.

Individual Condition. The Individual condition was designed to match the Communication condition's sequential labeling procedure and amount of

[6] We assumed the inverse Euclidean distance between the generated images would correspond to a reasonable degree with their perceptual similarity. After running the experiments, we checked this assumption via an online experiment that collected pairwise similarity judgements of the images on a Likert scale. The Spearman rank correlation of these similarity judgements with the inverse Euclidean distances was $\rho = .73, p < .001$, suggesting that the ordinal Euclidean similarity scale used for the feedback function was overall a good fit to the perceptual similarity of the images.

exposure to the stimuli, but involved a single participant categorizing images on the basis of similarity.

In each trial, the participant was presented with an onscreen array of all 25 images, one of which was selected with a red box to indicate it was the target. As in the Communication condition, the positions of images in the array were randomized independently for every trial. The participant was instructed to label similar images with the same word and different images with different words. As in the Communication condition, the participant was presented with a single initial label, and could reveal a new label at any stage, up to 25 labels; any labels they had revealed on a previous trial remained visible on their screen for all subsequent trials, without any information about which image(s) the participant had previously applied each label to. Once the participant had picked a label for the image, they were presented with the next trial. There were 100 trials in total, divided into 4 rounds, as in the Communication condition. Each round featured the 25 images as targets in a randomized order.

We analyzed the final category system produced by each individual participant. As for the Communication condition, this was defined by the labels the participant used the last time they labeled each image. While participants in the Individual condition did not interact, they were assigned into pseudo-pairs who shared the same wordlist, so that within-pair alignment could be compared with the Communication condition.

Results

All data and scripts for producing the analyses and graphs below are available at `https://github.com/silveycat/categories`.

Specificity

Participants in the Communication condition had on average 10.0 categories in their final systems, 95% CI [7.2, 12.7].[7] Participants in the Individual condition had a mean of 6.0 categories [5.1, 6.9]. This was 4.0 [1.1,

[7] For the Communication condition in Experiment 1, and for the Transmission + Communication condition in Experiment 2, confidence intervals for specificity, convexity and learnability were calculated based on standard errors and degrees of freedom for the average number for each pair (rather than counting each member of the pair separately), since these data points were not independent.

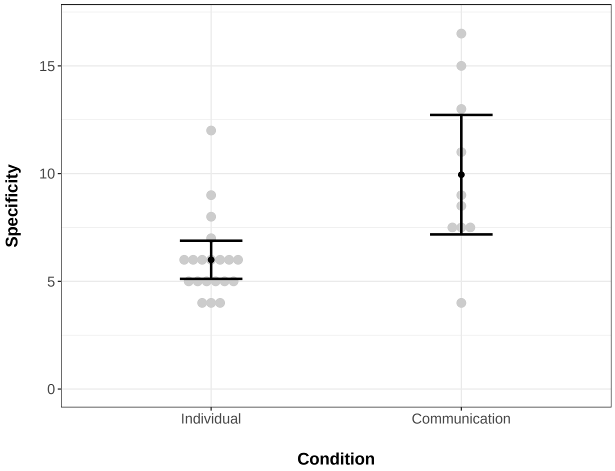


Figure 4: Average specificity (number of categories) of final category systems from the Individual and Communication conditions of Experiment 1. Error bars are 95% confidence intervals. Data points from each participant are shown in grey.

6.8][8] fewer than communicators (Figure 4). This difference was statistically significant: $t(11.2) = 3.05$, $p = .01$. As the larger confidence interval for communicators suggests, communicating pairs varied substantially in how many categories they used: SD for communicators = 3.88 [2.67, 7.08], and for individuals = 1.89 [1.44, 2.76].[9]

Convexity

Category systems produced in the Communication condition had an average convexity of 0.57 [0.42, 0.73]. This was less on average than in the Individual condition, where category systems had an average convexity of 0.65 [0.59, 0.71]. However, this difference of 0.08 [-0.09, 0.24] was not statistically significant: $t(12.36) = -0.99$, $p = .34$. Again, communicating pairs varied substantially in how convex their categories were: SD for communicators = 0.22 [0.15, 0.40], and for individuals = 0.13 [0.10, 0.19].

[8] Between-subjects CIs were calculated by the Welch-Satterthwaite method (Cumming, 2012) where variance between conditions was not homogenous.

[9] These and all following 95% CIs on SDs were calculated by the formula given in Sheskin (2011).

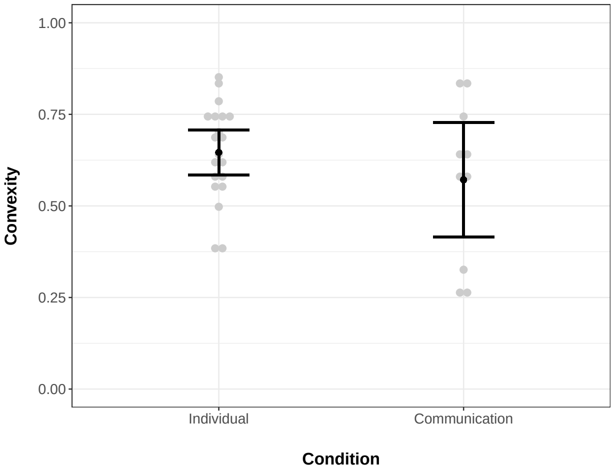
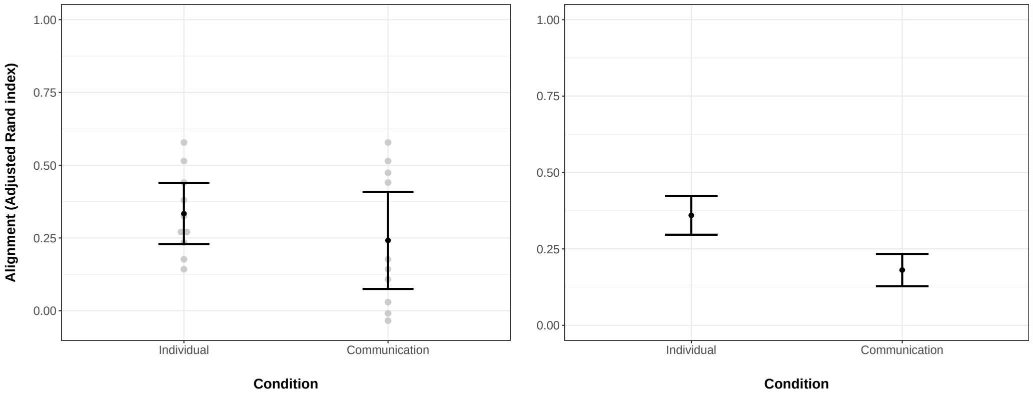


Figure 5: Average convexity of final category systems from the Individual and Communication conditions of Experiment 1. Error bars are 95% confidence intervals. Data points from each participant are shown in grey.

Alignment and convergence

Figure 6a shows alignment of category systems within pairs. Category systems produced in the Communication condition had an average Adjusted Rand index of 0.24 [0.07, 0.41], suggesting they were only around a quarter as aligned as they could be. Category systems produced by pseudo-pairs in the Individual condition (who used the same wordlist but did not interact) were slightly more aligned on average, 0.33 [0.23, 0.44]. This difference of 0.09 [-0.09, 0.28] was not statistically significant, $t(15.1) = -1.06, p = .31$. Again, communicators were more variable than individuals: SD for communicators = 0.23 [0.16, 0.43], and for individuals = 0.15 [0.10, 0.27].

We also measured convergence, or alignment across all category systems within a condition. Since convergence was calculated via a Monte Carlo procedure using randomly paired samples from each condition, standard statistical tests could not be applied. However, we calculated 95% confidence intervals from the distribution of means obtained. Convergence of category systems in the Individual condition was 0.36 [0.30, 0.42], whereas in the Communication condition it was 0.18 [0.13, 0.23]. Systems produced by communicating participants were more diverse overall than systems produced by individuals (Figure 6b).



(a) Within pairs

(b) Across all participants

Figure 6: (a) Alignment of final category systems within pairs; (b) convergence across all participants in each condition in Experiment 1. Error bars are 95% confidence intervals (for (a), empirical, and for (b), calculated via Monte Carlo simulation). For alignment, data points from each pair of participants are shown in grey.

Communicative success

In the first round of communication, pairs achieved an average score of 257 [245, 268] points, out of a maximum 375 (chance was 245). In the final round, pairs scored an average of 295 points [272, 317], an improvement of 38 [14, 62] points. First-round scores were close to chance, suggesting that pairs started by predominantly guessing. By the final round, pairs were reliably performing above chance on average. However, as the large confidence interval shows, pairs varied in their success.

We did not find that higher specificity in pairs' final systems was associated with higher communicative success: the correlation between average specificity and communicative success in the last two rounds (where participants were using their final category systems) was small, negative, and not statistically significant, $r = -0.30$ [-0.78, 0.41], $p = .41$.[10] By contrast, alignment within pairs correlated positively with communicative success in the last two rounds, $r = .72$ [.17, .93], $p = .02$.

[10] Confidence intervals on r were calculated using the formula given in Cumming (2012).

Discussion

We conducted Experiment 1 to investigate whether communication alone creates specific, convex, aligned category systems, as compared to baseline category systems created by individuals. For specificity, the answer is yes: communication appears to incentivize more specific category systems. However, surprisingly, more specific category systems do not result in higher levels of communicative success. For convexity, we did not find a clear answer: category systems that come out of communication are not significantly more convex than those created by individuals. Individuals naturally produce category systems that tend towards convexity (about 65% more than would be expected by chance); rather than reinforcing this process, communication if anything adds noise to it, resulting in lower mean convexity and higher variation. For alignment, we also did not find a clear advantage for communication: on average, communication does not boost category system alignment beyond the comparatively low level (33%) attained by individuals via shared perceptual biases. However, communicating pairs who managed to achieve higher alignment were more communicatively successful. We also found that communication leads to more diverse category systems than those produced by individuals, as shown by the higher convergence across all systems in the Individual condition than in the Communication condition. While not predicted in our hypotheses, previous work by Malt et al. (1999) does suggest that communication may prompt more diverse category systems than individual categorization: object categories defined by words in English, Chinese, and Spanish were more diverse than those produced in a card-sorting similarity task done by speakers of the three languages.

Taken together, these results suggest that the incentive for specificity imposed by communication may prevent it from working as hypothesized to boost category system alignment. Communication exerts a pressure for more specific category systems than those spontaneously produced by individuals. From the diversity of systems produced in the Communication condition, we can infer that different participants came up with different ways of making additional distinctions beyond the 6 or so typically made by individuals. This diversity creates a coordination problem for participants in the Communication condition, leading to the otherwise puzzling result that participants who communicated with each other were often less aligned than pseudo-pairs of isolated individuals (who had only their shared perceptual biases to help align their category systems). Thus, increased category system alignment is not a necessary consequence of communication. Similarly, the process of extending conventions through communication does not necessarily lead to convex categories: the high level of variation in convexity in communicators suggests that communication did not consistently lead participants to generalize their labels according to similarity. This suggests that the shareability account cannot fully explain the origins of similarity-based structure in category systems, at least in the case where these systems are built up from scratch during communication.

These results contrast with previous work suggesting that communication leads pairs to align on category systems that are structured by similarity (Markman & Makin, 1998; Voiklis & Corter, 2012). However, participants in these studies used their shared native language; as such, even though they were communicating about novel categories, the culturally transmitted categories of their native language were available to use as a starting point. Thus, communication in these studies was really communication combined with cultural transmission. Cultural transmission may provide a crucial baseline level of alignment that eases the coordination problem posed by jointly innovating a communicative category system.

This leads to a hypothesis: that cultural transmission, by providing a baseline level of alignment on which interacting participants can build, allows communication to have its predicted effect of increasing the structure and alignment of category systems. To test this hypothesis, in Experiment 2 we examined the effect of cultural transmission on category systems, both alone and when combined with communication. We aim to answer our two remaining questions: does cultural transmission alone lead category systems to become more or less effective for communication? and do the separate effects of communication and cultural transmission on category systems change when these pressures are combined?

Experiment 2

Methods

Experiment 2 compared category systems produced by participants in two conditions: a Transmission + Communication condition and a Transmission Alone condition. In the Transmission + Communication condition, pairs of participants first learned a set of labels for the images, then used these labels in a dyadic communication game. The category system defined by the final set of labels produced by one participant (randomly selected from the pair) then became the target for learning by the subsequent pair in a transmission chain. In the Transmission Alone condition, participants did not interact; however, they were assigned into pseudo-pairs who received the same input, so that within-pair alignment could be compared with the Transmission + Communication condition. These pseudo-pairs of participants learned a set of labels for the images and were individually tested; the category system defined by the final set of labels produced by one participant (randomly selected from the pseudo-pair) then became the target for learning by the next pair in a transmission chain.

Where Experiment 1 investigated improvisation of a category system without input, here we wanted to investigate the effect of repeated cycles of learning. To do this, we needed an initial input for participants in the first generation to learn from. Following previous iterated learning experiments (Carr et al., 2017; Kirby et al., 2008, 2015), we designed this input to lack the feature we were interested in: categories. Specifically, we followed Carr et al. (2017) by initially giving each image its own unique label. This provided scope for participants to create categories by generalizing the label for one image to other images, without providing any cues in the input as to how they should generalize. Any generalizations participants did make were hence motivated by pressures coming from cultural transmission and/or communication.

Our predictions are as follows. Based on Carr et al. (2017) and on our results from Experiment 1, we expect specificity to remain higher in the Transmission + Communication condition than in the Transmission Alone condition. Based on Carstensen et al. (2015), we expect convexity to increase in the Transmission Alone condition. Our results from Experiment 1 (where communicators produced less consistently convex systems than individuals) suggest that this increase in convexity may be attentuated in the Transmission + Communication condition, unless the effects of the two pressures differ when combined. Alignment within pairs should increase in the Transmission Alone condition, since iterated learning causes systems to become more learnable (Kirby et al., 2014): pseudo-pairs in later generations should be more successful at learning the target category system and hence more aligned. Again, our results from Experiment 1 (where communicators produced less consistently aligned systems than individuals) suggest that this effect may be attenuated in the Transmission + Communication condition, unless communication and cultural transmission have synergistic effects.

Stimuli

The set of images was the same as in Experiment 1. Labels were again randomly generated CVCV nonsense words.

Variables

Specificity, convexity, and alignment were measured as in Experiment 1.

Communicative success. For the Transmission + Communication condition, communicative success of final category systems was the pair's average score over the last two rounds of the communication game. For the Transmission Alone condition, communicative success of each pseudo-pair was estimated by simulating the communication game using their final category systems. For the simulation, as in the real communication game, each participant alternated between sender and receiver. On each trial, the simulated sender picked the label they used to refer to the target image in their final category system. The simulated receiver then matched this label to the correspondingly labeled category in their own system, and picked from this category the image that was maximally similar on average to all the images in the category. If the sender's label did not appear in the receiver's category system, the receiver picked an image at random from the whole set. Success scores were averaged over four simulated runs through the 25 images, to ensure scores were representative.[11]

Learnability. An additional dependent variable in Experiment 2 was the extent to which participants were able to successfully acquire the input category system. Category system learnability was assessed by measuring the similarity between the category system each participant was trained on and the final category system they produced. Since this is conceptually the same as alignment, similarity was measured using the Adjusted Rand index. A learnability score of 0 therefore meant the participant's reproduction of the target system was no better than chance; a learnability score of 1 meant the system was perfectly reproduced. Note that since the input to generation 1 was not a category system, this score is undefined for generation 1 participants.

[11] To check the validity of the simulation, we also simulated success scores for the Transmission + Communication participants and compared these to their veridical scores. Veridical and simulated per-category success scores had a correlation of $r = .96$, suggesting the simulation was a good fit to communicators' strategies.

Participants

Participants were 170 students at the University of Edinburgh and the University of Chicago (117 female, median age 22). 2 participants were excluded due to experimenter error, 4 due to networking issues, and 4 due to not completing the experiment within the allotted time. This left 160 participants. 80 participants from the University of Edinburgh took part in the Transmission + Communication condition and 80 participants from the University of Chicago took part in the Transmission Alone condition.[12] In each condition, participants were randomly assigned into 8 chains of 5 generations (following Silvey et al. 2015), with each generation consisting of a pair. The Transmission + Communication experiment took 90 minutes. Participants were paid £10, and each member of the pair with the highest communication score was additionally awarded a £10 Amazon voucher. The Transmission Alone experiment took 1 hour. As for Experiment 1, the Transmission Alone condition was shorter than the Transmission + Communication condition because participants did not have to wait for a partner's response on each trial. Participants were paid $10. The study was approved by the Linguistics and English Language Ethics Committee of the University of Edinburgh and the Social and Behavioral Sciences Institutional Review Board at the University of Chicago (IRB15-1364). All participants provided written informed consent.

Procedure

The experiment consisted of two phases. The learning phase was common to both conditions. After the learning phase, Transmission + Communication participants completed a communication phase, while Transmission Alone participants completed a solitary test phase.

Learning phase. During the learning phase, the participant's task was to learn labels that applied to the 25 images in the set. For generation 1 participants, input consisted of 25 labels, with one applying to each image. For subsequent generations, input was defined by the final category system of a participant from the previous generation (see subsection 'Iteration' below).

[12] To check that population differences did not affect the results, we compared accuracy of recalling label/image pairs during the learning phase across the two conditions. Accuracy (adjusted for the number of labels to be learned) was statistically equivalent across the two conditions, $W = 3396, p = .50$, suggesting that population differences did not affect transmission accuracy.

The method for displaying labels on screen was changed from Experiment 1. In Experiment 1, one label was displayed initially, and the participant could click to reveal additional labels. However, we felt that this method would be frustrating for participants during the learning phase of Experiment 2, since they might have to click repeatedly to reveal the correct label. Instead, 30 labels were displayed and remained onscreen for all trials. The number 30 was chosen to make it clear to participants that there were more labels than images and thus they did not have to use all the labels. The order of labels as presented on screen was randomized independently for each participant, but remained constant for a given participant through both phases of the experiment. Participants within a pair in the Transmission + Communication condition had the same wordlist. For comparison, as in Experiment 1, pseudo-pairs of participants within a chain and generation in the Transmission Alone condition also shared this wordlist. This meant there were 40 unique wordlists in total, each used by 2 participants from the Transmission + Communication condition and 2 participants from the Transmission Alone condition. As in Experiment 1, a number of different wordlists were used in order to ensure the results were not dependent on one specific set of labels.

On each learning trial, the participant was presented with the 30 labels and a randomized onscreen array of all 25 images, one of which was selected with a red box to indicate it was the target. The positions of images in the array were randomized independently on each experimental trial. The participant was instructed to click the label for the selected image. Once the participant had clicked a label, they were presented with a screen telling them if their label choice was correct or incorrect. The screen displayed the target image and the correct label for 4 seconds before moving to the next trial. Thus, initially, participants were forced to guess, but they had the opportunity to gradually acquire the label-image pairings via a trial-and-error process.

The learning phase ran for 100 trials, divided into 4 rounds. Each round featured the 25 images as targets, in a randomized order. Halfway through the learning phase, participants had the chance to take an optional break of up to 2 minutes.

Participants in the Transmission + Communication condition completed the learning phase at the same time as their partner, in separate booths. They were informed that their partner was learning the same language as them. At the end of each round, the participant who had finished first was shown a holding screen until their partner had also finished that round, at which point both proceeded to the next round.

The second phase of the experiment differed between the two conditions.

Communication phase (Transmission + Communication condition). Participants were told that they would play a communication game with their partner, using the labels provided to communicate the selected images. The procedure for the communication phase was identical to the Communication condition of Experiment 1, including the provision of full feedback after each trial. The only difference was in the presentation of labels onscreen to the sender. This mirrored the learning phase of Experiment 2: all 30 labels were provided from the start, rather than the participant clicking to reveal new labels as required.

As in Experiment 1, we analyzed the final category system produced by each participant in the pair. Each participant's final category system was defined by the labels that participant used the last time they were the sender for each image (i.e., during the last 2 rounds of the communication phase). Images given the same label were considered to be in the same category.

Test phase (Transmission Alone condition). Participants were instructed that they would be tested on the language they had learned. The test phase mimicked the sender's side of the communication phase. On each test trial, the participant was presented with the 30 labels and a randomized onscreen array of all 25 images, one of which was selected with a red box to indicate it was the target. The positions of images in the array were randomized independently on each experimental trial. The participant was instructed to click the label for the selected image. During the test phase, the participant was given no information on whether their label choices were correct or incorrect in the language they had learned. The test phase ran for 100 trials, divided into 4 rounds. Each round featured the 25 images as targets, in a randomized order. The test phase therefore provided Transmission Alone participants with the same overall amount of experience with the images and labels as Transmission + Communication participants. Halfway through the test phase, participants had the chance to take an optional break of up to 2 minutes.

As in Experiment 1, we analyzed the final category system produced by each participant. This was defined by the labels the participant used the last time they labeled each image.

Iteration. Following Kirby et al. (2015), after each pair had completed the experiment, one of the pair was randomly selected to provide the input for the next generation. The final category system produced by this partic-

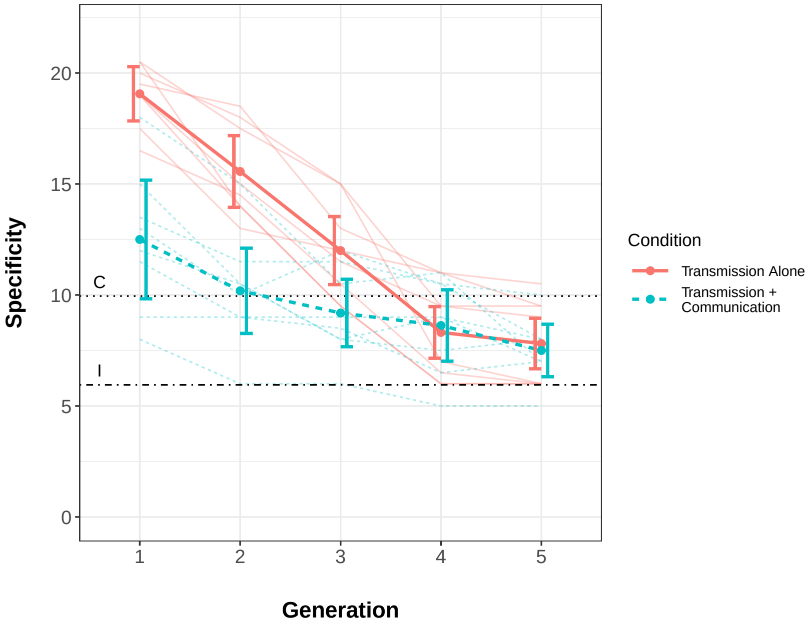


Figure 7: Average specificity (number of categories) of final category systems over generations from the Transmission Alone and Transmission + Communication conditions of Experiment 2. Error bars are 95% confidence intervals. Data from each chain (average for the pair of participants in each chain at each generation) is shown in thinner/paler lines. Reference lines show the average number of categories from the Individual condition (I) and the Communication condition (C) of Experiment 1.

ipant became the target for learning by the next pair or pseudo-pair in the transmission chain. While the groups of images referred to by the same label were preserved, a new wordlist was substituted for the old one. This was intended to prevent iconic associations between particular labels and images having a systematic effect within chains.

Results

Specificity

In generation 1, participants in the Transmission Alone condition had on average 19.1 [17.8, 20.3] categories in their final systems. This was 6.6 [3.8, 9.4] more than participants in the Transmission + Communication condition, who had on average 12.5 [9.8, 15.2] categories in their final systems (Figure 7). The number of categories dropped over generations in both conditions, but this decrease was steeper in the Transmission Alone condition: a loss of 11.3 [9.6, 12.9] compared to 5.0 [2.2, 7.8] in the Transmission + Communication condition. The number of categories in participants' systems in generation 5 was similar in both conditions: 7.8 [6.7, 9.0] in the Transmission Alone condition, and 7.5 [6.3, 8.7] in the Transmission + Communication condition. The final number in the Transmission Alone condition was 1.8 [0.4, 3.2] more than in the Individual condition from Experiment 1. The final number in the Transmission + Communication condition was 2.5 [-0.4, 5.3] less than in the Communication condition from Experiment 1.

A two-way ANOVA found a main effect of Condition, $F(1,110) = 38.0, p < .001$, a linear trend of Generation, $F(1,110) = 220.4, p < .001$, and an interaction, $F(1,110) = 28.9, p < .001$. In both conditions, the number of categories decreased over generations; however, the average number was lower in the Transmission + Communication condition, and the decrease was also less steep over generations. To investigate whether categories that emerge from transmission alone differ from those innovated by individuals, we also compared category systems from generation 5 of the Transmission Alone condition to category systems from the Individual condition of Experiment 1. Category systems emerging from transmission were significantly more specific than those innovated by individuals, $t(30.3) = 2.66, p = .01$.

Convexity

In generation 1, the convexity of participants' category systems in the Transmission Alone condition was 0.32 [0.17, 0.46]. This was 0.15 [-0.04, 0.34] lower than in the Transmission + Communication condition, where participants' systems had an average convexity of 0.47 [0.33, 0.61] (Figure 8). Convexity increased over generations in both conditions, but this increase was larger in the Transmission Alone condition: 0.36 [0.20, 0.52] compared to 0.25 [0.09, 0.41] in the Transmission + Communication condition. Average convexity in generation 5 was similar across the two conditions: 0.68 [0.60, 0.76] in the Transmission Alone condition, and 0.72 [0.61, 0.82] in the Transmission + Communication condition. The final value in the Transmission Alone condition was 0.03 [-0.07, 0.13] higher than in the Individual condition from Experiment 1. The final value in the Transmission + Communication condition was 0.14 [-0.03, 0.32] higher than in the Communication condition from Experiment 1.

A two-way ANOVA found a main effect of Condition, $F(1,110) = 9.6, p = .002$, and a linear trend of Generation, $F(1,110) = 34.9, p < .001$. The inter-

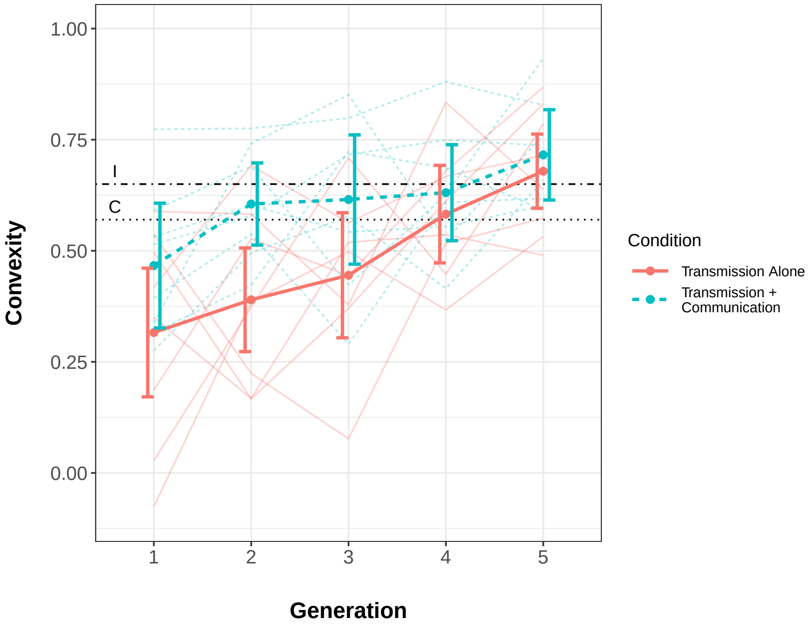
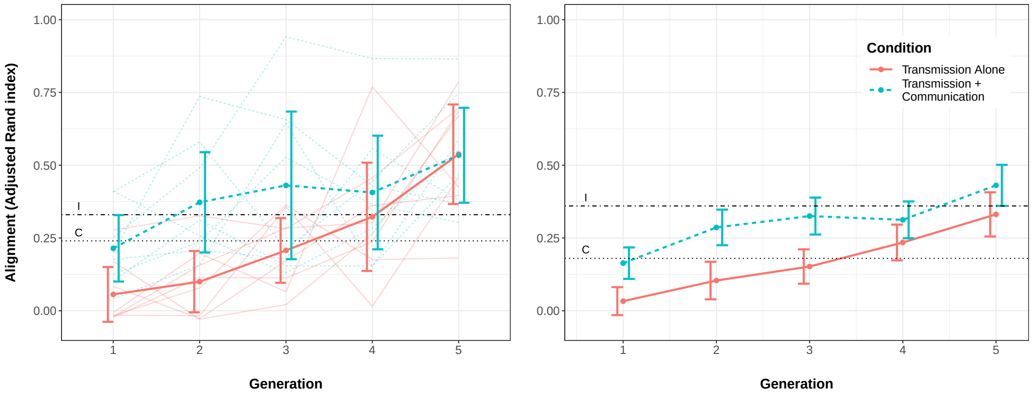


Figure 8: Average convexity of final category systems over generations from the Transmission Alone and Transmission + Communication conditions of Experiment 2. Error bars are 95% confidence intervals. Data from each chain (average for the pair of participants in each chain at each generation) is shown in thinner/paler lines. Reference lines show the average convexity of category systems from the Individual condition (I) and the Communication condition (C) of Experiment 1.



(a) Within pairs

(b) Across all participants

Figure 9: (a) Alignment of final category systems within pairs; (b) convergence across all participants over generations in each condition in Experiment 2. Error bars are 95% confidence intervals (for (a), empirical, and for (b), calculated via Monte Carlo simulation). For alignment, data from each chain is shown in thinner/paler lines. Reference lines show the average alignment and convergence from the Individual condition (I) and the Communication condition (C) of Experiment 1.

action was not statistically significant, $F(1, 100) = 1.95, p = .17$. Convexity was higher on average in the Transmission + Communication condition, and increased over generations in both conditions. Again, we compared category systems from generation 5 of the Transmission Alone condition to category systems from the Individual condition of Experiment 1. Category systems that emerged from transmission were not significantly more convex than those innovated by individuals, $t(29.3) = 0.67, p = .51$.

Alignment and convergence

Figure 9a shows alignment of category systems within pairs over generations in the two conditions. In generation 1 of the Transmission Alone condition, alignment was 0.06 [-0.04, 0.15]; this was 0.16 [0.02, 0.29] lower than in the Transmission + Communication condition. Transmission + Communication pairs in generation 1 were aligned at a level of 0.21 [0.10, 0.33], similar to the value of 0.24 [0.09, 0.39] for the communicators from Experiment 1, despite the Experiment 2 participants having learned identical systems first.

Alignment within pairs increased over generations in both conditions, but this increase was greater in the Transmission Alone condition: 0.48 [0.30, 0.66] compared to 0.32 [0.14, 0.50] in the Transmission + Communication condition. Alignment within pairs was similar across the two conditions in generation 5: 0.54 [0.37, 0.71] for Transmission Alone participants and 0.53 [0.37, 0.70] for Transmission + Communication participants. The final value in the Transmission Alone condition was 0.20 [0.02, 0.39] higher than the alignment of individuals from Experiment 1. The final value in the Transmission + Communication condition was 0.29 [0.08, 0.51] higher than that of communicators from Experiment 1.

A two-way ANOVA found a main effect of Condition, $F(1, 70) = 11.3, p = .001$, and a linear trend for Generation, $F(1, 70) = 36.1.p < .001$. The interaction was not statistically significant, $F(1, 70) = 2.7, p = .10$. Alignment within pairs was higher on average in the Transmission + Communication condition, and increased over generations in both conditions. We also compared category systems from generation 5 of the Transmission Alone condition to category systems from the Individual condition of Experiment 1. Category systems that emerged from transmission were significantly more aligned within pairs than those innovated by individuals, $t(12.3) = 2.38, p = .03$.

Convergence across participants also increased over generations in both conditions (Figure 9b). By generation 5, the convergence of category systems within each condition was equivalent to that of participants in the Individual condition of Experiment 1.

Communicative success

Figure 10 shows communicative success over generations in the two conditions (veridical scores for Transmission + Communication, and simulated scores for Transmission Alone - see Experiment 2 Methods).

In generation 1, average communicative success in the Transmission + Communication condition was 298 [278, 317], compared to 245 [233, 256] in the Transmission Alone condition (chance was 245). Communicative success increased over generations in both conditions, although this increase was larger in the Transmission Alone condition: 79 [57, 100] compared to 27 [4, 50] in the Transmission + Communication condition. By generation 5, success was similar in the two conditions: 323 [305, 341] in the Transmission Alone condition and 324 [314, 334] in the Transmission + Communication condition. This was 31 [5, 56] points more than the average score attained in the final two rounds by participants in the Communication condition of Experiment 1 (dotted line in Figure 10).

A two-way ANOVA found a main effect of Condition, $F(1, 70) = 35.7, p < .001$, a linear trend of Generation, $F(1, 70) = 62.8, p < .001$, and an inter-

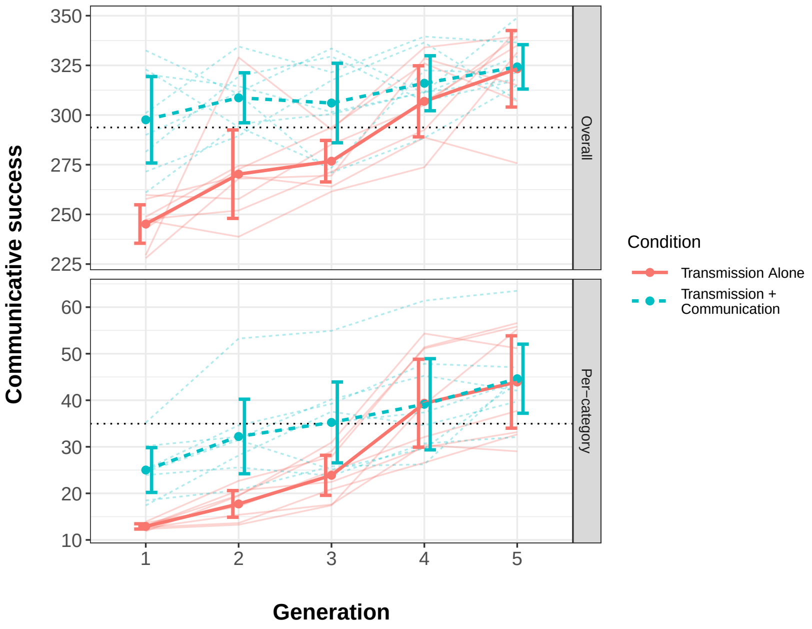


Figure 10: Communicative success of final category systems over generations in the two conditions of Experiment 2. For Transmission + Communication participants, success is the average veridical score over the last two rounds of the communication game. For Transmission Alone participants, success is the simulated score per round, calculated as described in the Methods of Experiment 2. Top panel shows overall success; bottom panel shows success divided by the average number of categories in pairs' or pseudo-pairs' systems, giving a measure of success per category. Reference lines on each panel show the corresponding average from the Communication condition of Experiment 1. Error bars are 95% confidence intervals. Data from each chain is shown in thinner/paler lines.

action, $F(1,70) = 17.3, p = .001$. Communicative success increased over generations in both conditions, but the magnitude of this increase was larger in the Transmission Alone condition, since participants in the Transmission + Communication condition were already achieving relatively high levels of communicative success in generation 1.

The increase in communicative success over generations is surprising given that the specificity of category systems was decreasing (as shown in Figure 7). To measure how these systems are nevertheless becoming more effective for communication, we calculate success per category, dividing overall success by the average number of categories in each pair's systems (bottom panel of Figure 10). Success per category increased linearly in the Transmission +

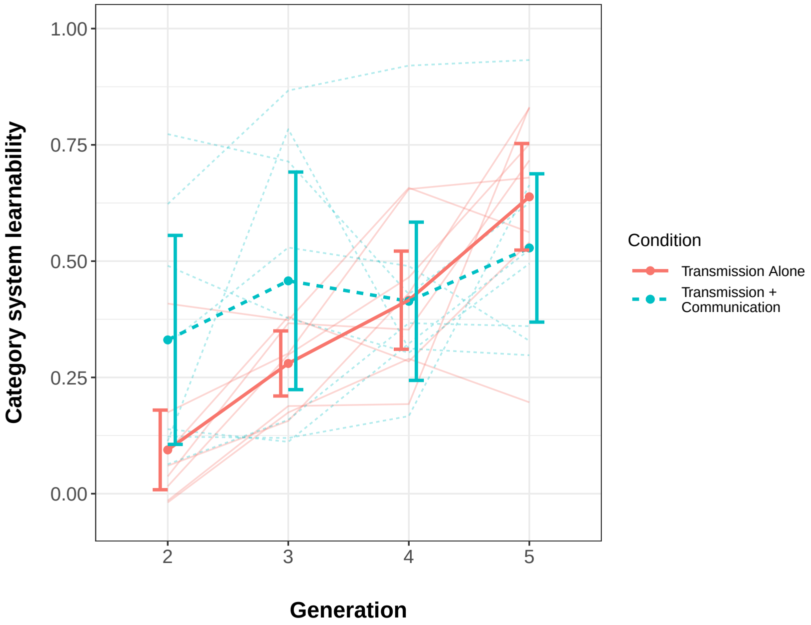


Figure 11: Average learnability of final category systems over generations from the Transmission Alone and Transmission + Communication conditions of Experiment 2. Learnability was defined as the similarity between the category system the participant was trained on and the final category system they produced, measured by the Adjusted Rand Index. Generation 1 is omitted because the input to these participants was not a category system. Error bars are 95% confidence intervals. Data from each chain (average for the pair of participants in each chain at each generation) is shown in thinner/paler lines.

Communication condition, whereas in the Transmission Alone condition, the increase was initially shallow and later steep, again reaching a similar level by generation 5.

Learnability

Figure 11 shows the change in learnability of category systems over generations. In generation 2 (the first generation for which learnability was defined), average learnability was 0.09 [0.01, 0.18] in the Transmission Alone condition. This was 0.24 [0.01, 0.47] less than in the Transmission + Communication condition, where average learnability was 0.33 [0.11, 0.56]. Learnability increased over generations by 0.54 [0.41, 0.68] in the Transmission Alone condition, but only by 0.20 [-0.06, 0.46] in the Transmission + Communication condition. By generation 5, learnability in the two conditions was similar: 0.64 [0.52, 0.75] in the Transmission Alone condition and 0.53 [0.37, 0.69] in the Transmission + Communication condition.

A two-way ANOVA found a linear trend of Generation, $F(1, 88) = 51.9, p < .001$, and an interaction between Generation and Condition, $F(1, 88) = 9.2, p = .003$. The main effect of Condition was not significant, $F(1, 88) = 2.8, p = .10$. Learnability did not change a great deal over generations in the Transmission + Communication condition, whereas it increased steeply in the Transmission Alone condition.

Discussion

In Experiment 1, communication had the predicted effect of increasing the specificity of category systems; however, contrary to our predictions, we did not find evidence that it increased the convexity or alignment of category systems beyond the level attained by individuals. We conducted Experiment 2 to investigate whether these effects of communication would remain when it was combined with cultural transmission, and whether cultural transmission alone would lead to category systems that were more or less effective for communication.

For all three of our main dependent variables, the effect of communication when combined with cultural transmission differed from the effect of communication alone. Category systems that were transmitted and used for communication became less specific earlier on than category systems that were transmitted without any communicative task. Category system convexity was also boosted by communication in early generations, relative to transmission alone. This suggests that in the case of cultural transmission from initially unstructured input, shareability processes working through communication can speed up the emergence of convex categories. Alignment within pairs showed a similar pattern: communication boosted alignment in early generations relative to transmission alone.

However, by generation 5, these differences between the conditions had equalized. Final category systems in the Transmission Alone condition were as specific, as convex and as communicatively effective as final category systems from the Transmission + Communication condition, despite never having been used in a communicative task. Alignment within pairs was also similar across the two conditions by generation 5. While pseudo-pairs from the Transmission Alone condition could not align via communication, the increasing learnability of the input they were trained on allowed them to acquire similar systems without ever interacting.

The convergence of category systems across all participants also increased over generations in both conditions. While not included in our predictions, this fits with previous work showing that cultural transmission causes convergence to the prior, resulting in across-community alignment regardless of whether learners were in the same transmission chain (Xu et al., 2013). The trajectory of convergence in the two conditions illuminates another difference from Experiment 1. There, communication alone appeared to incentivize more diverse systems; here, communication combined with cultural transmission promoted earlier convergence on more globally similar systems (even across pairs who did not learn the same input or communicate together). This was not a result of communicators introducing more errors and thus speeding up convergence to the prior: in early generations, participants in the Transmission + Communication condition were more successful at reproducing the system they were trained on than participants in the Transmission Alone condition (Figure 11). Rather, the changes communicators did make increased the structure of category systems in ways that made them more similar across chains from earlier generations.

Overall, we saw a strikingly similar pattern across all of our dependent variables. The effects of communication combined with cultural transmission differed from the effects of communication alone, and by the final generation, category systems in both conditions were similar in terms of specificity, convexity, and alignment.

General Discussion

The effects of communication depend on cultural transmission

This paper began by asking whether properties of the semantic category systems lexicalized in the world's languages can be explained by adaptation to communication and cultural transmission. Specifically, we aimed to answer three questions: 1) Does communication alone create specific, convex, aligned category systems, as compared to baseline category systems created by individuals? 2) Does cultural transmission alone lead category systems to become more or less effective for communication? 3) Do the separate effects of communication and cultural transmission on category systems change when these pressures are combined?

The answers were as follows. 1) While communication alone creates more specific category systems, these systems are not significantly more convex or aligned than those produced by individuals. 2) Cultural transmission alone eventually leads to systems that are as specific as those that result from transmission and communication, and equivalently convex and aligned in ways that make them communicatively useful: in fact, more communicatively useful on average than category systems that result from communication alone. Our results thus support the conclusion of Carstensen et al. (2015) that cultural transmission is a potential mechanism for the origin of category systems that are simple yet informative. However, as explained in the Introduction, this result rests on the key assumption (implemented in Carstensen et al.'s cost calculation and our communicative feedback function) that the simplest category structures – i.e., those that have a more concise cognitive representation – are also the most functional for communication, at a given level of specificity. Carr et al. (2018) show using agent-based models and human learning experiments that in a situation where simplicity and informativeness can be separated, cultural transmission leads to systems that are simple rather than informative. The communicative effectiveness of the category systems that emerge from cultural transmission alone in Experiment 2 may therefore be a side-effect of their simplicity. We discuss this point further in the section 'Stimuli and task' below. 3) When combined with cultural transmission, the effects of communication change: category systems lose specificity faster, and more rapidly become convex and aligned. Taken together, these results suggests an explanation for the less structured aspects of Nicaraguan Sign Language as used by cohort 1, who improvised the language over communication, compared to NSL as used by cohort 2, who acquired the language by transmission from older children (Silvey et al., 2016). Paradoxically, communication alone does not appear to be the best way to create category systems that are communicatively effective.

These results diverge from previous work that found a trade-off between communication and cultural transmission, with neither pressure sufficient to explain language structure (Kirby et al., 2008, 2015). In the case of category systems, while communication alone does not appear to be sufficient, cultural transmission alone does, if allowed to continue over several generations. In the case of real language, cultural transmission is embedded in communication: while a situation broadly analogous to the Communication condition of Experiment 1 can occur (such as cohort 1 of Nicaraguan Sign Language), a situation analogous to the Transmission Alone condition of Experiment 2 probably could not. However, under the assumptions made in this experimental design, the long-term adaptive pressures of cultural transmission alone on category systems are the same as those of cultural transmission and communication combined: category systems adapted for transmission alone are just as communicatively effective as those adapted for both pressures.

Figure 12 shows the final category systems from each condition of the two experiments. Here we see (A) the relatively low level of alignment within pairs of individuals from Experiment 1, (B) the wide variation in specificity, convexity and alignment of communicators from Experiment 1, and (C and D) the higher consistency of pairs from generation 5 of Experiment 2. This reinforces the central message of the paper: communication without prior learning leads to diverse category systems that vary in their communicative effectiveness; cultural transmission allows category systems to become structured, aligned, and communicatively effective in a way that is robust across individuals.

Culturally transmitted category systems and individual categorization

This paper began by observing that semantic category systems are not innovated by individuals: they are culturally transmitted and used for communication. In hypothesizing that these processes affect category structure, we implied that the category systems that emerge from communication and/or cultural transmission should differ from systems innovated by individuals. How do individuals' category systems from Experiment 1 resemble or differ from the final category systems from Experiment 2? Firstly, the systems from Experiment 2 are slightly but significantly more specific than individuals' systems from Experiment 1. While this could be a lingering effect of the maximally specific input, the trend to lose specificity appears to stabilize over the last two generations (Figure 7). This suggests that a culturally transmitted category system can lead individuals to learn more fine-grained categories that they would spontaneously invent. Secondly, the convexity of final systems in Experiment 2 was similar to that of individuals' systems from Experiment 1. Here, individuals innovate a similar level of convex structure to that converged on by cultural transmission. This suggests that, at least as modeled in this experiment, cultural transmission does not contribute to category convexity beyond the already high level attainable through individual innovation. Thirdly, for alignment, we found that cultural transmission leads to significantly higher levels of category system alignment (around 50% higher than chance) than the relatively low levels (around 33% higher than chance) possible through shared perceptual biases alone. Interestingly, this contrasts with convergence, or the overall similarity of all category systems within a condition (Figure 9b). Category systems within each condition end

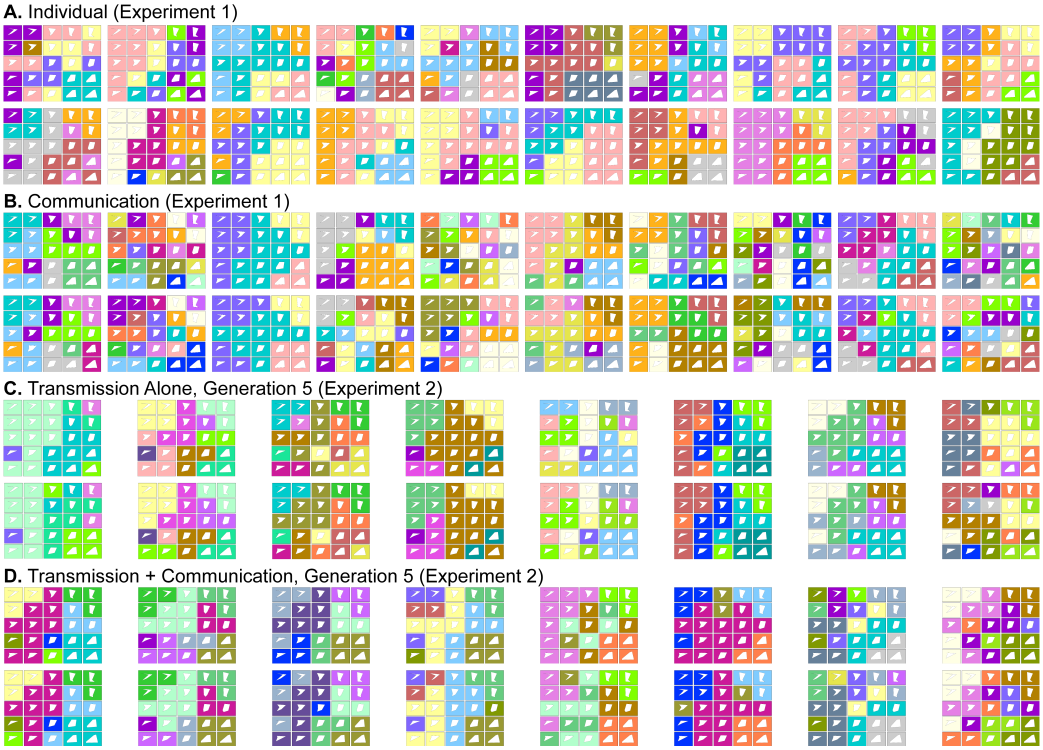


Figure 12: Final category systems from each condition of Experiments 1 and 2. Each grid is the category system of a single participant. Images with the same background color were given the same label, and therefore constitute a category. Vertically adjacent pairs in each experiment and condition shared wordlists (and hence color assignments). A) Final category systems produced by participants in the Individual condition of Experiment 1. B) Final category systems produced by participants in the Communication condition of Experiment 1. C) Final category systems produced by participants in generation 5 of the Transmission Alone condition of Experiment 2. D) Final category systems produced by participants in generation 5 of the Transmission + Communication condition of Experiment 2.

up, by generation 5, about as similar to each other as those of individual innovators. Cultural transmission does not lead to greater convergence across lineages on globally similar systems; however, it does allow for a higher level of category coordination within each lineage.

In terms of specificity and alignment, the category systems that emerge from cultural transmission do not appear to constitute a simple return to individual-level innovation. In terms of convexity, however, cultural transmission does not appear to have an effect beyond that of individual innovation. Convexity may be a more fundamental aspect of human categorization than specificity and alignment, less dependent on the communicative and cultural processes that shape the semantic categories of languages.

Caveats

The design of this experiment involved a number of assumptions which could affect the validity of the results. This section will address each of these in turn.

Number of generations

As the Experiment 2 graphs show, not all variables had stabilized by generation 5 of transmission. This potentially affects the robustness of some of the conclusions. Firstly, we concluded that cultural transmission with or without communication leads to category systems that are similarly specific, convex, and aligned. If we continued the transmission chains for more generations, might we expect category systems in the two conditions to diverge on each of these measures?

Specificity is almost identical across the two conditions in generations 4 and 5; furthermore, the trend appears to be stabilising, suggesting that the similar levels of specificity across conditions may be robust to number of generations in this experiment. However, it is important to note that other experiments have consistently found greater loss in specificity over transmission alone, compared to transmission + communication. In particular, Carr et al. (2017) found this in an experiment with 10 generations; however, in all but one chain, the number of categories at generation 10 in transmission-alone was 6 or 7, comparable to what we find in both conditions. It is possible that this number of categories is sufficiently learnable to be preserved even without a pressure for communication.

Convexity appears less stable from generations 4 to 5; however, the trend is parallel across the two conditions, suggesting that if convexity continued to increase, it would do so equivalently in both conditions.

While alignment is almost identical across the two conditions in generation 5, the trend from generation 4 to 5 is not parallel across the conditions. In this case, it is harder to extrapolate from the data what would happen if more generations were run. One possibility is that the sharper increase in Transmission Alone would continue. Another possibility is that over time, category systems in the Transmission + Communication condition would become more aligned than category systems in the Transmission Alone condition, since participants in this condition have two mechanisms for alignment: the increased learnability of the category systems (also available to participants in the Transmission Alone condition), and alignment through communicative feedback (not available to participants in the Transmission Alone condition). The similarity in alignment between conditions is therefore perhaps the most likely to be overturned if more generations were run.

Secondly, we concluded that the category systems that result from cultural transmission are distinct from those innovated by individuals: in particular, they are more specific and more aligned. This conclusion would change if specificity or alignment were to decrease if the chains were run for more generations. As noted above, a decrease in specificity in the Transmission Alone condition is theoretically possible and should not be ruled out. A decrease in alignment in either condition is unlikely, since the increasing learnability of category systems in both conditions promotes greater alignment. Therefore, the conclusion that cultural transmission leads to more specific category systems than individual innovation may not be as robust as the conclusion that cultural transmission leads to category systems that are more aligned within pairs.

Wordlist availability

A key difference between the current study and that of Carr et al. (2017) is how participants generate labels for communication. In the current study, participants select a label from a wordlist available on screen. In Carr et al.'s study, participants did not have a wordlist available; instead, labels were typed from memory. The wordlist makes all labels equally accessible to participants, whereas retrieving labels from memory boosts the accessibility of some labels at the expense of others (Harmon & Kapatsinski, 2017), increasing the likelihood that some labels will drop out of use entirely. Our finding that cultural transmission alone leads to the same level of specificity as cultural transmission and communication may not hold in the more ecologically valid situation of participants having to retrieve labels from memory. Future experiments should explore this possibility.

Input specificity

The input participants received in generation 1 of Experiment 2 was maximally specific, with a unique label for each of the 25 images. As in previous iterated learning studies, this input was not designed to be a plausible initial state: rather, it was designed to lack categories in order to model their emergence. An alternative starting point without categories would be minimally specific, with every image referred to by the same label. This input would potentially alter the trajectories of category emergence in the two conditions. With no pressure to be specific, Transmission Alone participants might avoid introducing new labels, whereas Transmission + Communication participants might introduce new labels in the aid of better scores in the communication game. Future work should examine the robustness of these results to different inputs.

Stimuli and task

The structure of the communicative task in these experiments encodes four main assumptions, each of which could affect the validity of the results.

The first assumption is that communication involves making perceptual distinctions. In real language use, communication is more often concerned with events, relations, and functional or social properties. Our findings may not generalize to the structure of relational categories, which vary more across languages (Bowerman, 1996) and are more likely than perceptual categories to depend on language for their acquisition (Gentner, 2016).

The second is the structure of the feedback function, which encodes the assumption that being almost right about the identity of a target is better than being very wrong. While we justify this decision in the Methods, it may not be appropriate for all communicative situations. The convexity results, in particular, likely depend on this property of the feedback function, since it makes the same structures simple and informative in this stimulus space. It is important to note that this does not hold for all stimulus spaces: Carr et al. (2018) show that in a stimulus space with two clearly separable dimensions, the simplest category structures are one-dimensional, whereas the most informative are two-dimensional. By contrast, in a stimulus space like the one used in these experiments, or the spatial scenes used by Carstensen et al. (2015), the dimensions of variation are not so clearly separable. Here, the simplest category structures (those with the most concise cognitive representation) are arguably the same as the most informative (those that allow the most accurate reconstruction of the speaker's intended referent): multidimensional convex categories whose members cluster tightly together in similarity space. Supporting this, Vong et al. (2019) find that even for stimuli with only four dimensions, categories with family resemblance structure are more learnable than "simpler" categories based on a single dimension. Future work should unpack the extent to which the same structures are simple and informative in different real-world domains.

The third assumption is that communication can be modeled as taking place on a single-word basis. This assumption does not generalize beyond very early child language. In adult language use, semantic categories are used in combination, and these combinations interact in ways that alter the content of the categories involved (e.g., 'red wine' involves different possible reds than 'red hair': Gärdenfors, 2000). The effect of combination on the structure of semantic categories may alter the dynamics of communication and cultural transmission considerably, and is a key area to be explored in future research.

The fourth assumption is that communication is dyadic. Relaxing this simplifying assumption could potentially provide a fairer test of the shareability hypothesis in the case of communication alone; Freyd (1983) hypothesized that shareability effects on semantic structures should become more pronounced as the size of the community increases. Supporting this, Fay et al. (2008) found that symbol systems created by interacting communities of participants were more effective than those created by pairs. Having participants communicate with more than one partner could reduce the amount of between-pair variation, leading to more structured and aligned category systems across participants without the need for cultural transmission.

Conclusion

The semantic categories labeled by words in human languages are culturally transmitted and used for communication. We set out to investigate whether these pressures affect their specificity, convexity, and alignment across speakers. To do this, we examined the effect of these pressures separately and together on participants' categorization of images drawn from a continuously varying similarity space. We found that communication alone led to more specific category systems, but did not raise convexity and alignment beyond the relatively low levels achieved by individual innovation. However, when combined with cultural transmission, the effects of communication changed: category systems became less specific, more convex, and more aligned within fewer generations than under cultural transmission alone. Category systems in the final generation were similar across the two conditions in all three properties. Under the assumptions made in this experiment, communication is neither necessary nor sufficient to create category systems that are robustly effective for communication. Category systems that resulted from cultural transmission were distinct from those innovated by individuals: they were more specific and more aligned within pairs. The fact that semantic category systems are culturally transmitted may allow individuals to learn more category distinctions than they would spontaneously innovate, and also to coordinate their semantic systems more effectively with others in their speech community than would be possible on the basis of shared perceptual biases alone. The results offer insight into how language as a culturally transmitted system allows humans to go beyond individual biases and learn shared category systems enriched by the knowledge and communicative precedents of previous generations.

Acknowledgements

Thanks to Alex Carstensen for help with informativeness/communicative cost, and to Christos Christodoulopoulos for help with the communication experiment. This work was supported by the Arts and Humanities Research Council (PhD studentship 1085982 to CS) and by discretionary research funds from the University of Chicago.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429. doi:`10.1037/0033-295X.98.3.409`.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science*, *302*, 1234–1236. doi:`10.1126/science.1087513`.

Bowerman, M. (1996). The origins of children's spatial semantic categories: cognitive versus linguistic determinants. In J. J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–176). Cambridge: Cambridge University Press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493. doi:10.1037/0278-7393.22.6.1482.

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*, 892–923. doi:10.1111/cogs.12371.

Carr, J. W., Smith, K., Culbertson, J., & Kirby, S. (2018). Simplicity and informativeness in semantic category systems. URL: psyarxiv.com/jkfyx. doi:10.17605/OSF.IO/JKFYX.

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 Meeting of the Cognitive Science Society.* (pp. 303–308). Austin, TX: Cognitive Science Society.

Chemla, E., Buccola, B., & Dautriche, I. (2019). Connecting content and logical words. *Journal of Semantics*, . doi:10.1093/jos/ffz001. Advance online publication.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–508. doi:10.1017/S0140525X08004998.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Deacon, T. W. (1997). *The symbolic species: The coevolution of language and the brain*. New York: Norton.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3553–3561. doi:10.1098/rstb.2008.0130.

Freyd, J. (1983). Shareability: The social psychology of epistemology. *Cognitive Science*, *7*, 191–210. doi:10.1207/s15516709cog0703_2.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in Human Neuroscience*, *5*, 11. doi:10.3389/fnhum.2011.00011.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2014). *Geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181–218. doi:10.1016/0010-0277(87)90018-7.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*, 292–304. doi:10.1111/j.1756-8765.2009.01020.x.

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*, *71*, 650–657. doi:10.1037/amp0000082.

Gentner, D., Özyürek, A., Gürcanli, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, *127*, 318–330. doi:10.1016/j.cognition.2013.01.003.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22–44. doi:10.1016/j.cogpsych.2017.08.002.

Hermer-Vazquez, L., Moffet, A., & Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition*, *79*, 263–299. doi:10.1016/S0010-0277(00)00120-7.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218. doi:10.1007/BF01908075.

Jäger, G., & van Rooij, R. (2007). Language structure: Psychological and social constraints. *Synthese*, *159*, 99–130. doi:10.1007/s11229-006-9073-5.

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

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 United States of America*, *105*, 10681–10686. doi:10.1073/pnas.0707835105.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. doi: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. doi:10.1016/j.cognition.2015.03.016.

Landau, B., & Shipley, E. (2001). Labelling patterns and object naming. *Developmental Science*, *4*, 109–118. doi:10.1111/1467-7687.00155.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262. doi:10.1006/jmla.1998.2593.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*, 331–354.

Pothos, E. M., & Chater, N. (1997). Classification and prior assumptions about category "shape": New evidence concerning prototype and exemplar theories of categorization. In M. G. Shafto, & P. Langley (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 620–625). Mahwah, NJ: Erlbaum.

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 12116–12120. doi:10.1073/pnas.0914044107.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850. doi:10.2307/2284239.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1436–1441. doi:10.1073/pnas.0610341104.

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). Malden, MA: Wiley-Blackwell. doi:10.1002/9781118346136.ch11.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences*, *4*, 321–354. doi:10.1207/s15327809jls0403_3.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323. doi:10.1126/science.3629243.

Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures*. (5th ed.). Boca Raton, FL/London: Chapman & Hall/CRC.

Silvey, C., Flaherty, M., Goldin-Meadow, S., Kirby, S., & Smith, K. (2016). Communication without prior learning inhibits the emergence of systematic structure. In *Proceedings of EvoLang XI, Language Adapts to Interaction Workshop, 21 March, 2016*. URL: http://evolang.org/neworleans/workshops/papers/LATI_8.html.

Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science*, *39*, 212–226. doi:10.1111/cogs.12150.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, *28*, 469–489. doi:10.1017/S0140525X05000087.

Theiler, J., & Gisler, G. (1997). A contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation. In B. Javidi, & D. Psaltis (Eds.), *Proceedings of SPIE 3159: Algorithms, Devices, and Systems for Optical Information Processing* (pp. 108–118). doi:10.1117/12.279444.

Voiklis, J., & Corter, J. E. (2012). Conventional wisdom: Negotiating conventions of reference enhances category learning. *Cognitive Science*, *36*, 607–634. doi:10.1111/j.1551-6709.2011.01230.x.

Vong, W. K., Hendrickson, A. T., Navarro, D. J., & Perfors, A. (2019). Do Additional Features Help or Hurt Category Learning? The Curse of Dimensionality in Human Learners. *Cognitive Science*, *43*, e12724. doi:10.1111/cogs.12724.

Warglien, M., & Gärdenfors, P. (2011). Semantics, conceptual spaces, and the meeting of minds. *Synthese*, *190*, 2165–2193. doi:10.1007/s11229-011-9963-z.

Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, *63*, 165–174.

Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, *280*. doi:10.1098/rspb.2012.3073.