ARTICLE

Experimental evidence for scale-induced category convergence across populations

Douglas Guilbeault [1,2], Andrea Baronchelli [3,4] & Damon Centola [2,5,6,7]✉

Individuals vary widely in how they categorize novel and ambiguous phenomena. This individual variation has led influential theories in cognitive and social science to suggest that communication in large social groups introduces path dependence in category formation, which is expected to lead separate populations toward divergent cultural trajectories. Yet, anthropological data indicates that large, independent societies consistently arrive at highly similar category systems across a range of topics. How is it possible for diverse populations, consisting of individuals with significant variation in how they categorize the world, to independently construct similar category systems? Here, we investigate this puzzle experimentally by creating an online "Grouping Game" in which we observe how people in small and large populations collaboratively construct category systems for a continuum of ambiguous stimuli. We find that solitary individuals and small groups produce highly divergent category systems; however, across independent trials with unique participants, large populations consistently converge on highly similar category systems. A formal model of critical mass dynamics in social networks accurately predicts this process of scale-induced category convergence. Our findings show how large communication networks can filter lexical diversity among individuals to produce replicable society-level patterns, yielding unexpected implications for cultural evolution.

[1] The Haas School of Business, The University of California, Berkeley, Berkeley, CA 94720, USA. [2] Network Dynamics Group, The University of Pennsylvania, Philadelphia, PA 19106, USA. [3] Department of Mathematics, City University London, London EC1V 0HB, UK. [4] The Alan Turing Institute, London NW12DB, UK. [5] Annenberg School for Communication, The University of Pennsylvania, Philadelphia, PA 19106, USA. [6] School of Engineering, The University of Pennsylvania, Philadelphia, PA 19106, USA. [7] Department of Sociology, The University of Pennsylvania, Philadelphia, PA 19106, USA. ✉email: dcentola@asc.upenn.edu

People exhibit substantial creativity and variation in how they categorize novel and ambiguous phenomena[1–6]. This observation has led decades of research to argue that category formation in large social groups is unpredictable[7–13]. Larger populations contain a greater diversity of people and thus a greater diversity of categories that can be adopted through communication networks, which are expected to lead to variable and path-dependent cultural trajectories[8–10,14–18]. Meanwhile, there is considerable evidence that independent populations consistently arrive at highly similar category systems across a range of topics[19–21], including flora[22], geometry[23], emotion[24], color[25], and kinship[26]. These findings pose a striking puzzle—how is it possible for separate and diverse populations, composed of individuals with significant variation in how they categorize the world, to independently construct similar category systems[19,27–29]?

One explanation for the observed patterns of category convergence across societies is that there are innate universals in human psychology that arise independently of social interaction[10,19–22,30]. However, because these theories explain similarity across populations in terms of innate human categories, they are limited in explaining how category convergence can emerge when individuals widely vary in their categorization of novel stimuli[1–6,10]. An alternative view holds that stochastic dynamics can lead separate large populations to arrive at similar category systems even when individuals vary in how they categorize the world. Formal models of voting behavior, for instance, show that increasing sample size can increase the likelihood of identifying the most popular choice in a population for both binary[31] and pluralistic choices[32,33]. Similarly, recent findings on critical mass dynamics[28,34] suggest that large populations have the potential to promote the interpersonal spread of popular linguistic conventions. Building on this work, our formal analyses indicate that when the popularity of categories can be described by a hypergeometric distribution (or binomial for infinite populations), then increasing population size can trigger "scale-induced" category convergence, in which a small number of categories are more likely to consistently reach critical mass[34] and spread[35,36] in large populations, resulting in replicable evolutionary trajectories (see Supplementary Information sections 1.1, 1.2, and 1.3 for model specification).

An empirical test of these predictions has not yet been possible because it requires comparing the cultural trajectories of independently evolving small and large populations to observe whether differences in population size directly affect the similarity of the category systems that populations produce. In this study, we developed an online experimental platform called the "Grouping Game" that enabled real-time observation of novel category formation in small and large populations (see 'Methods'). We use the Grouping Game to investigate this puzzle experimentally by examining how small and large populations independently construct category systems for a continuum of novel and ambiguous stimuli. Solitary individuals and small groups produced highly divergent category systems. Yet, across replicated studies with unique subjects, separate large populations converged on highly similar category systems. These findings offer insight into category similarities across societies[19–22], by showing how large communication networks can filter lexical diversity in such a way that leads communities toward convergent and replicable trajectories in category creation.

Results

Figure 1 displays the category systems that emerged in distinct small and large populations. Figure 1a shows that small populations ($N = 2$) produced highly divergent category systems. Only 6% of labels were shared across independent dyads, and there was
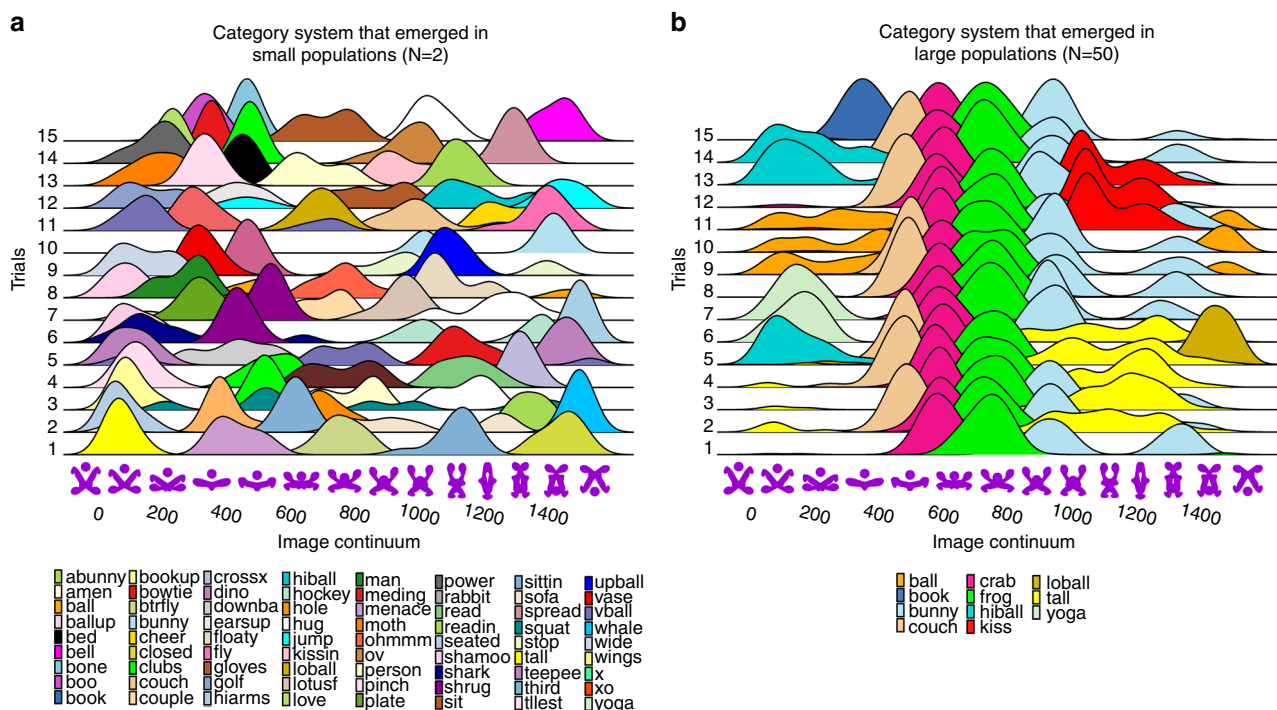

Fig. 1 Larger populations promote category convergence across populations. Comparing the level of convergence in category systems that emerged in small ($N = 2$) (**a**) and large ($N = 50$) (**b**) populations. Each row displays the category system constructed by a single unique population in each condition after 100 rounds of interaction. The horizontal axis displays the image continuum of shapes, consisting of 1500 slices. Density distributions display the frequency of successful coordination for each label, as well as the region of the continuum to which each label referred. Each color indicates a unique label. Similarity in the category systems across independent populations indicates convergence.
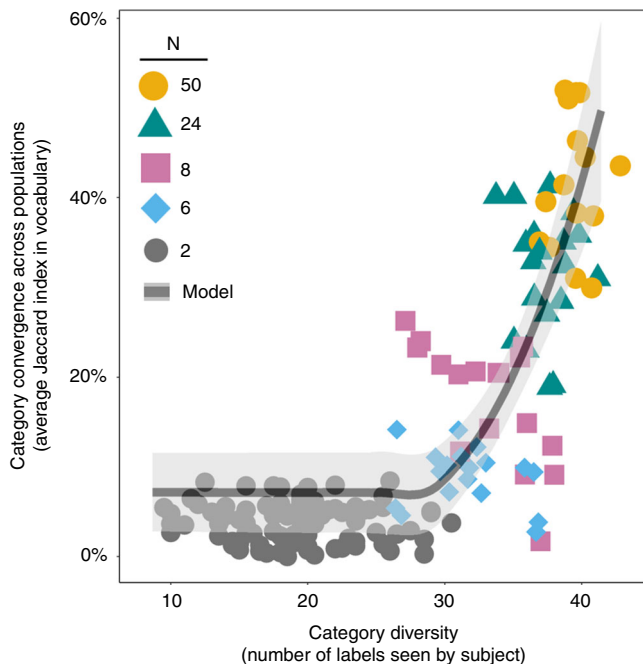
Fig. 2 Convergence in the vocabularies that emerged in populations of different sizes, for N = 2 (black dots), N = 6 (blue diamonds), N = 8 (purple squares), N = 24 (green triangles), and N = 50 (yellow circles). Vertical axis reports the average similarity in vocabulary (average Jaccard Index) between each network trial and all other networks of the same population size. Horizontal axis displays category diversity, measured as the average number of unique labels encountered by subjects in a population. Data points represent experimental results (80 dyads and 15 social networks of each size). Black trend line shows model predictions (averaged over 50 simulated trials; 100 rounds each trial; $d_{min} = 0.01$; $|L| = 5000$; $b = 1$); see Supplementary Information section 1.2 for model specification (equation S1). The measure of center indicated by the model trend line is the mean Jaccard Index among simulated trials of the same population size, ordered by the average category diversity in each trial (Fig. S1). Error bands show 95% confidence intervals.

no consistency in how these dyads partitioned the continuum ($p < 0.001$, $n = 80$, Kruskal–Wallis H Test). As a result, dyads varied not only with respect to the labels they adopted for the same regions of the continuum, but also with respect to the regions of the continuum they successfully categorized. (Complementary analyses showing the same results for the $N = 1$ condition are provided in the Supplementary Information section 1.7; Figs. S6 and S7). By contrast, large populations ($N = 50$) generated remarkably similar vocabularies (50% Jaccard Index, $p < 0.001$, $n = 95$, Wilcoxon Rank Sum Test, two-sided) and similar partitions of the continuum ($p = 0.87$, $n = 15$, Kruskal–Wallis H Test), indicating convergence in how these independent populations categorized the novel stimuli (Fig. 1b).

These findings appear puzzling at first since larger populations are expected to increase the unpredictability of category formation as a result of containing a greater diversity of individuals, and thus a greater number of categories that are introduced and available for adoption. Yet, our results indicate that increasing population size—and thereby increasing the diversity of categories—can counterintuitively lead to convergent trajectories in category formation across populations.

Our theoretical predictions for these convergence dynamics provide an excellent fit with our experimental findings (Fig. 2) (see Supplementary Information section 1.2 for model specification; Fig. S2). Across all experimental conditions, label diversity

significantly increased with population size ($p < 0.001$, $n = 120$, Jonckheere-Terpstra Test). Figure 2 shows that greater label diversity within populations predicts greater similarity in the category systems that emerge between populations of the same size ($p < 0.001$, $n = 120$, Jonckheere-Terpstra Test). We find these convergence dynamics not just for the labels that were used, but also for how participants partitioned the continuum into distinct regions (Supplementary Information section 1.8; Fig. S8).

Robustness experiments (Supplementary Information section 1.9) show that providing more rounds of interaction for the dyads (>125) did not increase their rate of convergence. Instead, it further entrenched their divergent category systems.

We propose a simple mechanism to explain our findings. We suggest that larger populations amplify the spread of initially more frequent labels[37], leading these common labels to reach a "tipping point"[34], after which they diffuse and become widely adopted[35,36]. Figure 3a shows the frequency with which every label was independently suggested by participants across all studies. Consistent with Zipf's law[38], a small number of labels like "crab" and "bunny" were common, meaning they were more likely to arise separately from distinct participants, whereas the vast majority of labels were rare, meaning they were only introduced by a small number of individuals (Fig. 3a).

Figure 3b shows the relationship between population size and critical mass dynamics (formal model and detailed analyses provided in the Supplementary Information section 1.3; Fig. S2). In small populations, common labels were not sufficiently reinforced to reach the tipping point needed to trigger widespread adoption[35,36]. Consequently, small populations ($N = 2$) were significantly more likely to adopt rare labels ($p < 0.001$, $n = 80$, Wilcoxon Signed Rank Test, two-sided), leading these populations to follow divergent evolutionary trajectories. However, increasing population size significantly increased the likelihood that common labels (like "crab" and "bunny") would be reinforced and adopted ($p < 0.001$, $n = 120$, Jonckheere-Terpstra Test), while significantly reducing the likelihood that rare labels would spread ($p < 0.001$, $n = 120$, Jonckheere-Terpstra Test). Our findings indicate a direct relationship between population size and category convergence across independent populations (Fig. 3c). For large populations ($N = 50$), the likelihood of common labels becoming widely adopted approaches unity, leading to consistent and replicable trajectories in collective category formation (Fig. 3b, c and S2).

A crucial implication of our theory is that category similarities across social groups do not solely depend upon cognitively salient features of the labels themselves, but also upon the labels' frequency in the population. An established intuition is that certain categories gain popularity because they have intrinsic appeal (e.g., because of their 'natural' descriptive fit with the stimuli)[39]. However, even when the most popular labels (e.g., "crab" and "bunny") were attempted in dyads, they regularly failed to gain acceptance (Supplementary Information section 1.3; Fig. S2). This suggests that the adoption of these labels is not strictly determined by their cognitive appeal, but rather by the fact that they are more likely to be reinforced and reach critical mass in larger populations.

To evaluate this hypothesis, we experimentally tested the following counterfactual: if we artificially inflated the popularity of infrequent labels to reach critical mass, would this trigger convergence on those labels rather than on more cognitively appealing ones? We conducted six robustness trials ($N = 24$) in which each network contained a minority of confederate subjects (37%) tasked with spreading a novel category system based on infrequent labels (see Supplementary Information section 1.10 for full details on experimental design; Fig. S9). For instance, we trained confederates to use the rare label "sumo" (Fig. 3a) for the same regions of the visual continuum associated with the most popular label in our
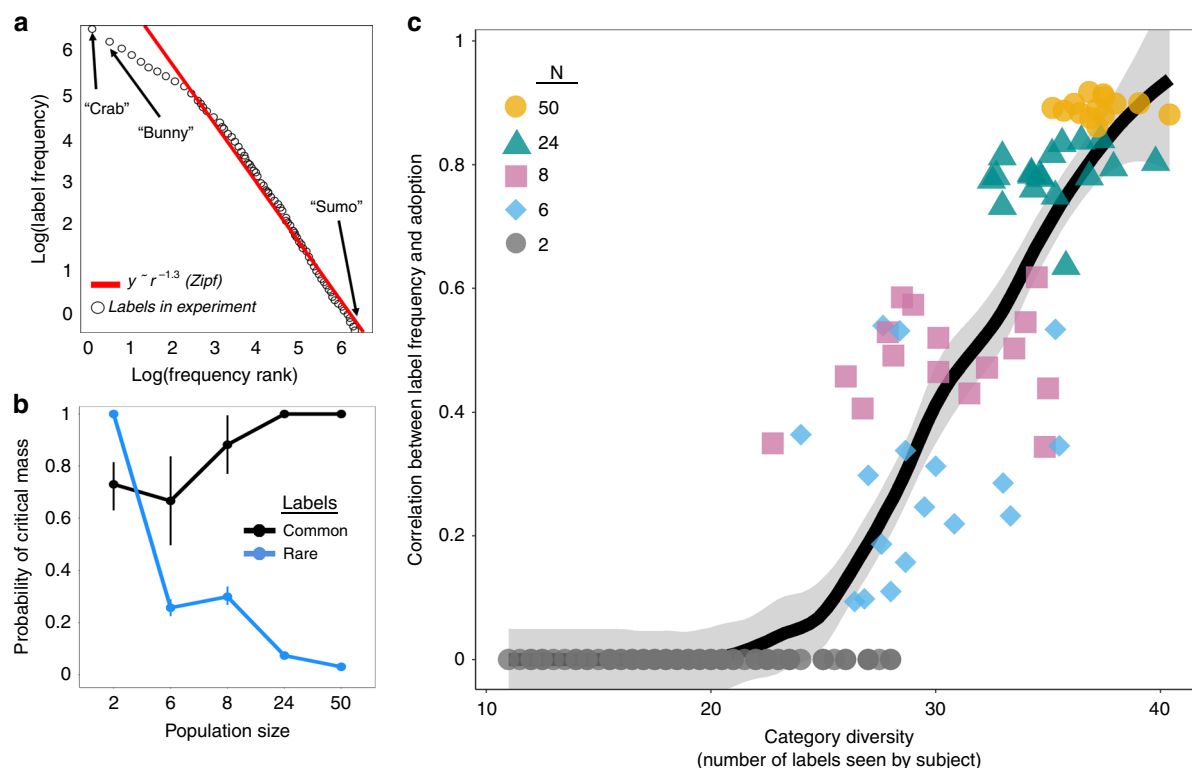
Fig. 3 Larger populations amplify the spread of initially frequent labels. a Using the Zipf distribution to model the initial frequency of labels (including data from all conditions; $N = 2$, $N = 6$, $N = 8$, $N = 24$, and $N = 50$), where initial frequency refers to the number of individuals who introduced a label without any prior exposure to the label in the task. Vertical axis displays the log of each label's initial frequency. Horizontal axis displays the log of each label's frequency rank. **b** Displaying the mean effect of population size on the ability for labels to reach critical mass (when at least 25% of subjects in a network independently introduce a label). Common labels are identified as outliers with high initial frequency (Supplementary Information section 1.3). Data display the proportion of experimental trials in each condition for which each label type reached critical mass. Error bars display 95% confidence intervals. **c** The correlation between the initial frequency of a label in a population and the proportion of subjects in a population who adopted the label (vertical axis), where adopting a label entails that a subject produced a label after being exposed to it. Horizontal axis displays the diversity of categories in each trial, indicated as the average number of unique labels encountered by each subject in a network. Error bands display 95% confidence intervals. All observations are independent and at the network-level. All panels represent data from 80 unique dyads and 15 unique social networks of each size.

initial studies, "crab" (Fig. 1b). Figure 4 shows that although "crab" appeared in each robustness trial, "sumo" consistently out-competed "crab". In every robustness trial, populations adopted the confederates' labels across each region of the continuum, yielding significantly more convergent category systems (58% Jaccard Index) than those that emerged in $N = 24$ populations without confederates (35% Jaccard Index) ($p < 0.01$, $n = 21$, Wilcoxon Rank Sum Test, two-sided; Figs. S10–S12).

Discussion

The "social constructivist" view of cultural evolution suggests that large communication networks contain greater individual variation, which leads to greater divergence and unpredictability in the evolution of category systems[7–15,18]. Here, we show that while increasing the size of communication networks does, in fact, significantly increase the diversity of categories that people encounter, it does not increase divergence. Rather, it increases category convergence across independent populations. Our results suggest that convergence in category formation across independent populations is significantly shaped by the communication networks in which people are embedded.

These findings offer experimental insight into past observational data on category similarities across societies[19–26]. Our findings suggest that communication in large social networks can help filter cognitive and lexical diversity in such a way that promotes the replicable development of similar category systems

across separate communities. Importantly, we observe scale-induced category convergence for an arbitrary and novel continuum of stimuli that lacks pre-existing objective boundaries, whereas some mathematical models assume that well-defined objective boundaries are essential for producing stable convergence dynamics in the emergence of vocabularies[40–42]. We anticipate that future research may extend our findings to study how population dynamics can improve both the stability and accuracy of category systems in domains with objective truth conditions. In particular, we anticipate that future studies may apply our findings to address challenging issues in content moderation and classification, for instance to eliminate individual biases in large-scale citizen science efforts and related human crowdsourcing tasks, such as Galaxy Zoo[43] or Gravity Spy[44], and to improve consistency in the classification of acceptable and unacceptable content on social media[45].

Methods

This research was approved by the Institutional Review Board at the University of Pennsylvania, where the study was conducted, and it included informed consent by all participants.

A total of 1480 subjects were recruited from Amazon Mechanical Turk to participate in an online language game[28,34,46] called "The Grouping Game" (Fig. 5). Each trial consisted of unique individuals, producing independent experimental observations. All subjects were required to live in the U.S. with English as their first language. When logging into the Grouping Game, subjects were randomized into either a dyad, or a network of 6, 8, 24, or 50 people. We conducted additional trials using an alternative version of the Grouping Game constructed for solitary players
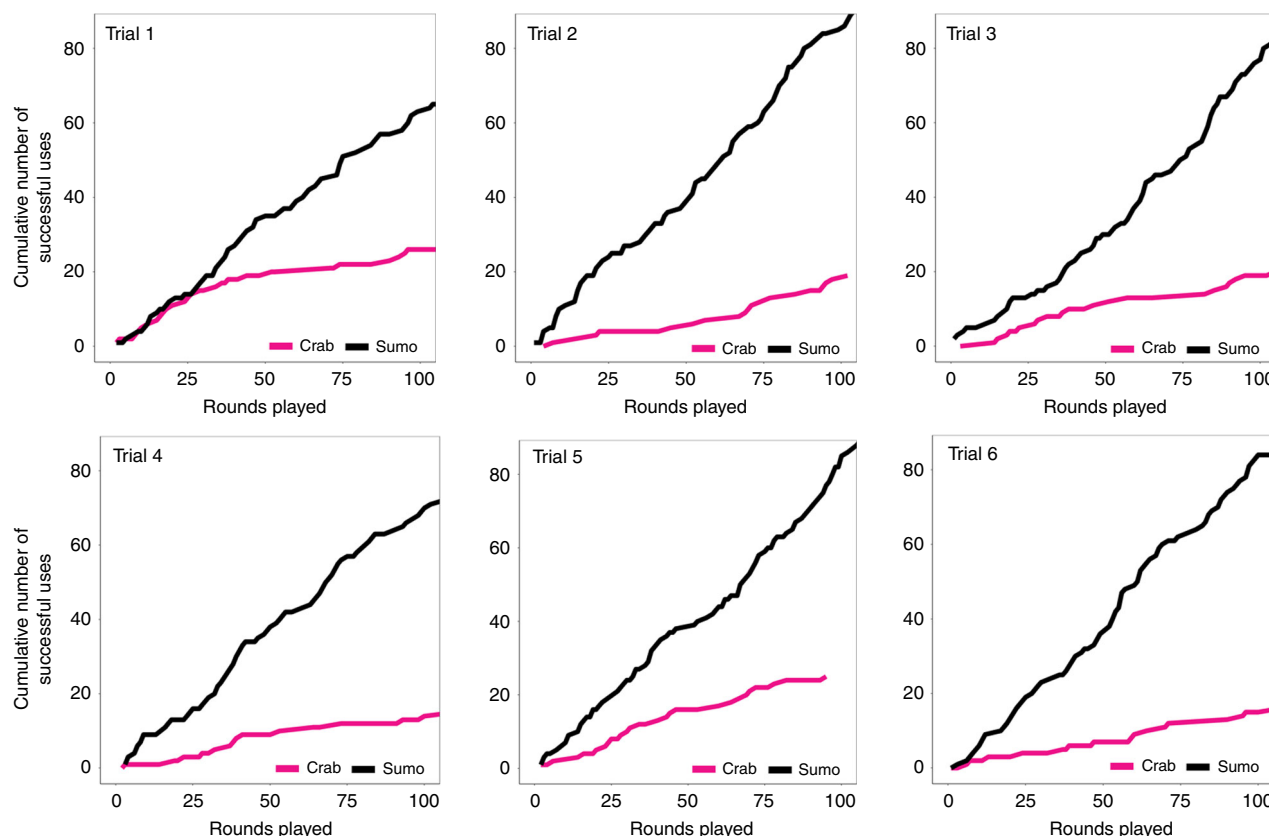
Fig. 4 Time series showing the adoption of the confederates' rare label ("sumo") by experimental subjects (i.e., nonconfederate subjects). Pink lines indicate the cumulative number of successful uses among experimental subjects of the label "crab". Black lines indicate the cumulative number of successful uses among experimental subjects of the label "sumo". Each round is measured as $N/2$ pairwise interactions, such that each player has one interaction per round. The data displayed exclude all interactions between confederates.

($N = 1$), which generated results consistent with our findings for dyads (Figs. S2 and S3). We collected 80 dyads ($N = 2$) and 15 social networks for each population size (i.e., $N = 6$, $N = 8$, $N = 24$, and $N = 50$). There were no differences in the distribution of demographic traits across conditions, in terms of gender ($p = 0.56$), ethnicity ($p = 0.42$), and age ($p = 0.67$) (Kruskal–Wallis H Test). All data were collected between September 2018 and February 2020.

We created a continuum of novel shapes that defined the space of visual stimuli for the Grouping Game (Fig. 5c). Analogous to the visible color spectrum, our continuum was a smooth geometrical progression that was not inherently partitioned[27]. We evenly divided this continuum into 1500 slices. Each slice was a unique shape.

Upon arriving to the study, participants viewed instructions on how to play the Grouping Game. In the game, participants played a series of pairwise one-shot coordination games, where a single coordination game constituted a single round. In each round, participants were randomly paired with another participant in their network. In all conditions, participants could be paired with any other participant, creating fully connected (i.e., homogeneously mixing) populations.

Each round of the game proceeded as follows. First, each subject was randomly paired with another subject in their network (in the dyads, participants were always paired with the same person). Second, in each pair on each round, one subject was randomly assigned to be the "speaker" (Fig. 5a) and the other was the "hearer" (Fig. 5b). Third, the speaker in each pair was shown three randomly selected slices (or shapes) from the visual continuum, which were presented side by side (Fig. 5). One of the three shapes was randomly highlighted only for the speaker. The speaker was given 30 seconds to enter a label of their own creation into a free text-entry window, with the aim of helping their partner to distinguish the highlighted shape from the other two presented shapes. The only restriction on label production was that speakers were not allowed more than six characters to prevent highly detailed sentence-like descriptions that could not fail to coordinate. Even with this character limit, nearly 5000 unique labels were introduced. Fourth, the hearer in each pair was shown the same set of three shapes as the speaker but in an alternate order (Fig. 5b). The hearer was then given 30 seconds to identify the shape corresponding to the speaker's label (Fig. 5b). If the hearer selected the correct shape, both players received a successful payment (10¢). If the speaker failed to select the correct shape, both players were financially penalized (1¢). Every experimental trial of the Grouping Game lasted 60 min. In every trial, each subject played at least 100 rounds.

The image continuum was held constant across conditions. In every trial, every subject was presented with a uniform distribution of images drawn equally from all regions of the continuum (see Supplementary Information section 1.4; Fig. S3). The algorithm that randomly selected three images to display each round was designed so that participants were never shown the same shape twice. All images displayed for a given scene were at least 75 frames apart along the continuum, following prior theoretical models[27,47]. This design induced subjects to categorize the images, because in this environment, subjects would only use the same label on multiple rounds if they were grouping distinct images under a single category. The set of three images displayed on each round were unique to each pairing, such that two separate speaker and hearer pairs interacting at the same time would see distinct image sets on a given round.

Participants had no information about the labels used by other members of the population except for their partner's response in the round in which they were paired[28,34]. In every network ($N = 2$, $N = 6$, $N = 8$, $N = 24$, and $N = 50$), subjects received identical instructions. Subjects did not have information about their partner's identity, nor the size of their network. A manipulation check confirms that subjects' knowledge about their network was held constant across experimental conditions (Supplementary Information section 1.5; Fig. S4). Any differences in the category systems that emerged across experimental conditions can be attributed to the direct effects of population size on the dynamics of category formation.

To identify the categories that emerged in each condition of each trial, we used DBSCAN (Density-based spatial clustering of applications with noise)[48]. A key advantage of DBSCAN is that it does not require one to specify the number of clusters in the data a priori, as opposed to k-means clustering. The DBSCAN algorithm involves two key parameters: *MinPts*, which determines the minimum number of points that must be included in each cluster, and ε, which denotes the radius of the neighborhood around a point x that is used when identifying clusters. For each condition in each trial, we ran DBSCAN to identify clusters of labels based on their values along two features: the total number of successful coordination events associated with a label across the entire population, and the total number of cumulative adopters overtime associated with each label. We ran DBSCAN separately for each condition in each trial because increasing population size significantly increases the number of possible successful coordination events and adopters that can be associated with a label. Following standard methodology, for each application of DBSCAN[48], *MinPts* was set to 3 (the number of dimensions plus one) and ε was chosen by plotting the k-distances among points and using the
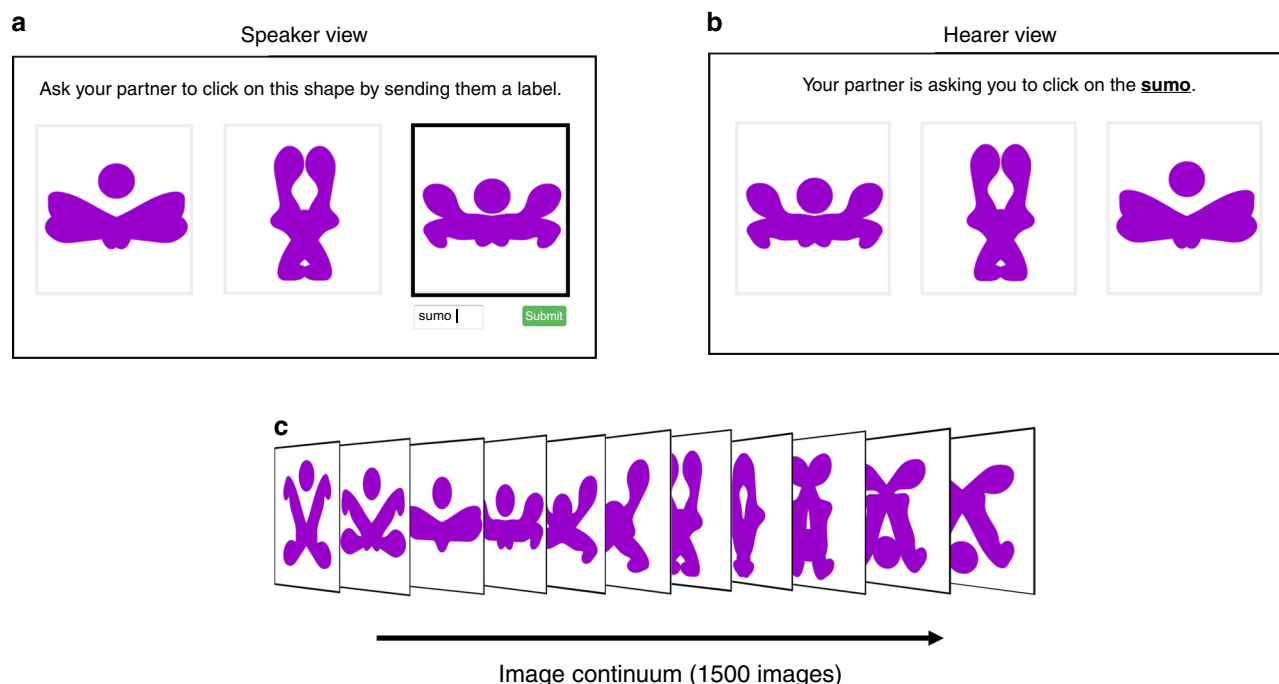
Fig. 5 The Grouping Game. Screenshots of "The Grouping Game" interface from the view of a speaker (**a**) and the view of a hearer (**b**) on a given round. **c** A sample of the continuum of novel shapes used as stimuli in the experiment.

knee of the plot to identify the optimal ε. Emergent categories were identified as the unique cluster of labels with the highest values in terms of their total number of successful uses and their total number of adopters. In practice, DBSCAN identified 3–5 labels as emergent categories. All results are robust to varying vocabulary size across a wide range of fixed sizes (see Supplementary Information Section 1.6; Fig. S5). In cases where two categories were successful for the same region of the continuum, the label with the highest number of coordination successes was deemed the most successful category for this region.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data underlying this study are publicly available at: https://github.com/drguilbe/categories2020; https://ndg.asc.upenn.edu/uncategorized/network-dynamics-of-category-emergence/. Source data are provided with this paper.

Code availability
The source code for this study is publicly available at: https://github.com/drguilbe/categories2020; https://ndg.asc.upenn.edu/uncategorized/network-dynamics-of-category-emergence/.

References
1. Shepard, R. N. & Cermak, G. W. Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cogn. Psychol.* **4**, 351–377 (1973).
2. Spalding, T. & Gregory, M. Effects of background knowledge on category construction. *J. Exp. Psychol.* **22**, 525–538 (1996).
3. Johnson, K. E. & Mervis, C. B. Effects of varying levels of expertise on the basic level of categorization. *J. Exp. Psychol.* **126**, 248–277 (1997).
4. Levinson, S. C. & Wilkins, D. P. *Grammars of Space: Explorations in Cognitive Diversity* (Cambridge University Press, Cambridge, 2006).
5. Ranjan, A. & Srinivasan, N. Dissimilarity in creative categorization. *J. Creat. Behav.* **44**, 71–83 (2010).
6. Lindsey, D. T., Brown, A. M., Brainard, D. H. & Apicella, C. L. Hadza color terms are sparse, diverse, and distributed, and presage the universal color categories found in other world languages. *Iperception* **7**, 1–6 (2016).
7. Berger, P. L. & Luckmann, T. *The Social Construction Of Reality: A Treatise in the Sociology of Knowledge* (Anchor Books, New York, 1967).
8. David, P. A. Path dependence: a foundational concept for historical social science. *Cliometrica* **1**, 91–114 (2007).
9. Fay, N., Garrod, S., Roberts, L. & Swoboda, N. The interactive evolution of human communication systems. *Cogn. Sci.* **34**, 351–386 (2010).
10. Atran, S. & Medin, D. L. *The Native Mind and the Cultural Construction of Nature* (A Bradford Book, New York, 2008).
11. Bowker, G. C. & Star, S. L. *Sorting Things Out: Classification and Its Consequences* (MIT Press, Cambridge, 2000).
12. Salganik, M. J., Dodds, P. S. & Watts, D. J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
13. Macy, M., Deri, S., Ruch, A. & Tong, N. Opinion cascades and the unpredictability of partisan polarization. *Sci. Adv.* **5**, eaax0754 (2019).
14. DiMaggio, P. Classification in art. *Am. Sociol. Rev.* **52**, 440–455 (1987).
15. Reali, F., Chater, N. & Christiansen, M. H. Simpler grammar, larger vocabulary: how population size affects language. *Proc. Biol. Sci.* **285** (2018).
16. Freeberg, T. M., Dunbar, R. I. M. & Ord, T. J. Social complexity as a proximate and ultimate factor in communicative complexity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 1785–1801 (2012).
17. Fay, N. & Ellison, T. M. The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PLoS ONE* **8**, e71781 (2013).
18. Bowern, C. Correlates of language change in hunter-gatherer and other 'small' languages. *Lang. Linguist. Compass* **4**, 665–679 (2010).
19. Malt, B. C. Category coherence in cross-cultural perspective. *Cogn. Psychol.* **29**, 85–148 (1995).
20. Brown, D. E. Human universals, human nature & human culture. *Daedalus* **133**, 47–54 (2004).
21. Youn, H. et al. On the universal structure of human lexical semantics. *Proc. Natl Acad. Sci. USA* **113**, 1766–1771 (2016).
22. Brown, C. H. *Language and Living Things: Uniformities in Folk Classification and Naming* (Rutgers University Press, New Brunswick, 1984).
23. Burris, H. Geometric figure terms: their universality and growth. *J. Anthropol.* **1**, 18–41 (1979).
24. Jackson, J. C. et al. Emotion semantics show both cultural variation and universal structure. *Science* **366**, 1517–1522 (2019).
25. Regier, T., Kay, P. & Khetarpal, N. Color naming reflects optimal partitions of color space. *Proc. Natl Acad. Sci. USA* **104**, 1436–1441 (2007).
26. Kemp, C. & Regier, T. Kinship categories across languages reflect general communicative principles. *Science* **336**, 1049–1054 (2012).
27. Baronchelli, A., Gong, T., Puglisi, A. & Loreto, V. Modeling the emergence of universality in color naming patterns. *Proc. Natl Acad. Sci. USA* **107**, 2403–2407 (2010).
28. Centola, D. & Baronchelli, A. The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl Acad. Sci. USA* **112**, 1989–1994 (2015).

29. Kirby, S., Dowman, M. & Griffiths, T. L. Innateness and culture in the evolution of language. *Proc. Natl Acad. Sci. USA* **104**, 5241–5245 (2007).

30. Hawkins, J. *Explaining Language Universals* (Wiley-Blackwell, Hoboken, 1991).

31. Grofman, B., Feld, S. & Owen, G. Group size and the performance of a composite group majority: statistical truths and empirical results. *Organ. Behav. Hum. Perform.* **33**, 350–359 (1984).

32. Young, P. Condorcet's theory of voting. *Am. Political Sci. Rev.* **82**, 1231–1244 (1988).

33. List, C. & Goodin, R. Epistemic democracy: generalizing the Condorcet Jury Theorem. *J. Polit. Philos.* **9**, 277–306 (2001).

34. Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018).

35. Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734 (2007).

36. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).

37. Pagel, M., Beaumont, M., Meade, A., Verkerk, A. & Calude, A. Dominant words rise to the top by positive frequency-dependent selection. *Proc. Natl Acad. Sci. USA* **116**, 7397–7402 (2019).

38. Adamic, L. & Huberman, B. Zipf's law and the Internet. *Glottometrics* **3**, 143–150 (2002).

39. Winkielman, P., Halberstadt, J., Fazendeiro, T. & Catty, S. Prototypes are attractive because they are easy on the mind. *Psychol. Sci.* **17**, 799–806 (2006).

40. Gerhard, J., Metzger, L. & Riedel, F. Voronoi languages: equilibria in cheap-talk games with high-dimensional types and few signals. *Games Econ. Behav.* **73**, 517–537 (2011).

41. O'Connor, C. Evolving perceptual categories. *Philos. Sci.* **81**, 840–845 (2014).

42. Mani, A., Varshney, L. & Pentland, A. Quantization games on social networks and language evolution. Preprint at http://arxiv.org/abs/2006.00584 (2020).

43. Watson, D. & Floridi, L. Crowdsourced science: sociotechnical epistemology in the e- research paradigm. *Synthese* **195**, 741–764 (2018).

44. Coughlin, S. et al. Classifying the unknown: discovering novel gravitational-wave detector glitches using similarity learning. *Phys. Rev. D* **99**, 082002 (2019).

45. Gorwa, R., Binns, R. & Katzenbach, C. Algorithmic content moderation: technical and political challenges in the automation of platform governance. *Big Data Soc.* **7**, 2053951719897945 (2020).

46. Steels, L. *The Talking Heads Experiment: Origins of Words and Meanings* (Language Science Press, 2015).

47. Puglisi, A., Baronchelli, A. & Loreto, V. Cultural route to the emergence of linguistic categories. *Proc. Natl Acad. Sci. USA* **105**, 7936–7940 (2008).

48. Ester, M., Kriegel, H. P., Sander, J. & Xiaowei, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD-96 Proceedings* (AAAI Press, 1996).

Author contributions

D.G. and D.C. designed the project, D.G. ran the experiments, D.G., A.B., and D.C. conducted the analyses, D.G. and D.C. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-20037-y.

Correspondence and requests for materials should be addressed to D.C.

Peer review information *Nature Communications* thanks Simon Garrod and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.