

Reconciling long-term cultural diversity and short-term collective social behavior

Luca Valori^a, Francesco Picciolo^a, Agnes Allansdottir^b, and Diego Garlaschelli^{c,d,1}

^aDepartment of Chemistry, University of Siena, Via A. De Gasperi 2, 53100 Siena, Italy; ^bUniversity of Siena, 53100 Siena, Italy; ^cLorentz Institute for Theoretical Physics, University of Leiden, Niels Bohrweg 2, NL-2333 CA Leiden, The Netherlands; and ^dCABDyN Complexity Centre, University of Oxford, Park End Street OX1 1HP, Oxford, United Kingdom

Edited* by Robert Axelrod, University of Michigan, Ann Arbor, MI, and approved November 22, 2011 (received for review June 15, 2011)

An outstanding open problem is whether collective social phenomena occurring over short timescales can systematically reduce cultural heterogeneity in the long run, and whether offline and online human interactions contribute differently to the process. Theoretical models suggest that short-term collective behavior and long-term cultural diversity are mutually excluding, since they require very different levels of social influence. The latter jointly depends on two factors: the topology of the underlying social network and the overlap between individuals in multidimensional cultural space. However, while the empirical properties of social networks are intensively studied, little is known about the large-scale organization of real societies in cultural space, so that random input specifications are necessarily used in models. Here we use a large dataset to perform a high-dimensional analysis of the scientific beliefs of thousands of Europeans. We find that interopinion correlations determine a nontrivial ultrametric hierarchy of individuals in cultural space. When empirical data are used as inputs in models, ultrametricity has strong and counterintuitive effects. On short timescales, it facilitates a symmetry-breaking phase transition triggering coordinated social behavior. On long timescales, it suppresses cultural convergence by restricting it within disjoint groups. Moreover, ultrametricity implies that these results are surprisingly robust to modifications of the dynamical rules considered. Thus the empirical distribution of individuals in cultural space appears to systematically optimize the coexistence of short-term collective behavior and long-term cultural diversity, which can be realized simultaneously for the same moderate level of mutual influence in a diverse range of online and offline settings.

collective social phenomena | homophily | social influence | sociophysics | ultrametricity

How a society spontaneously organizes macroscopically from the microscopic, uncoordinated behavior of individuals is one of the most studied and exciting problems of modern science (1–3). Collective social phenomena are systematically observed in several aspects of everyday life, including the onset of large-scale popularity and fashion [both offline (4) and online (5–7)], the existence of large speculative bubbles and herding behavior in financial markets (8–10), the spontaneous emergence of order in traffic and crowd dynamics (11), the properties of voting dynamics (4, 12), the structure of countrywide communication networks (3, 13, 14), individual specialization and the spreading of cooperation, habits, fear, gossip, rumors, etc. (2, 11, 15). When collective phenomena occur, large parts of a population turn out to be globally correlated or synchronized as a result of the combination of many local interactions, even if no centralized mechanism takes place. Importantly, the collective outcome is different from a mere superposition of noninteracting individual behaviors, and contrasts the “representative agent” scenario often postulated in economic theories (16). The different characteristics, choices and behaviors of individuals, rather than being ‘averaged out’ in the long run and at a large scale, may in some circumstances become amplified at the societal level (1).

The substrate for collective phenomena is being expanded by the increasing connectedness and integration of people across the world (17, 18), made possible on one hand by the steady enhancement of communication technologies (18–21) and on the other hand by the intensification of global trade, worldwide travel, and international education (22–24). All these factors virtually reduce physical and communication distances among people, irrespective of their cultural traits. In particular, large-scale electronic platforms where people can exchange information with unprecedented speed and breadth are triggering a major shift toward effectively infinite-ranged social interactions (see *SI Text, Discussion*). This naturally leads to the question of whether the diversity of behaviors, attitudes, and opinions is destined to be progressively reduced in the long run. Naively, one expects that stronger and more frequent bursts of collective social phenomena taking place on short timescales may gradually result in more homogeneous cultural traits in the long term. However, the emergence of group boundaries has preserved cultural diversity for millenia across populations (25–27) and survived many technological revolutions that extended the length and range of social interaction (18–21). Moreover, recent studies of online behavior suggest that, even in the virtual world, social influence massively emerges (6) and determines large-scale collective phenomena (7), but at the same time is observed to coexists with the systematic presence of long-lived communities of people sharing similar traits (28–30).

The above paradox is reinforced by the fact that similar mechanisms are believed to be among the key driving forces of both collective social behavior (1, 6) and cultural convergence (31). Various simplified models have been introduced to quantitatively simulate the fate of cultural diversity and the dynamics of opinions in large groups (2, 11). Rather than conveying a detailed picture of reality, models of social dynamics aim at understanding the effects that different mechanisms proposed in social science may have when combined together and when taking place at a large-scale level. In particular, the popular model proposed by Axelrod (31) studies the combined effects of two key phenomena: *social influence* [the observed tendency of social interactions to favor convergence and consensus (32)] and *homophily* [the observed tendency of culturally similar individuals to interact more than dissimilar ones (33–35)]. The importance of the model resides in showing that social influence and homophily do not necessarily reinforce each other and determine a culturally homogeneous society. In fact, the model suggests that the persistence of cultural diversity in the long term is ensured by the inhibition of influence among dissimilar individuals, even if socially tied. These results are enhanced by the introduction of an impor-

Author contributions: D.G. designed research; A.A. and D.G. performed research; L.V., F.P., and D.G. analyzed data; and D.G. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹To whom correspondence may be addressed. E-mail: garlaschelli@lorentz.leidenuniv.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1109514109/-DCSupplemental.

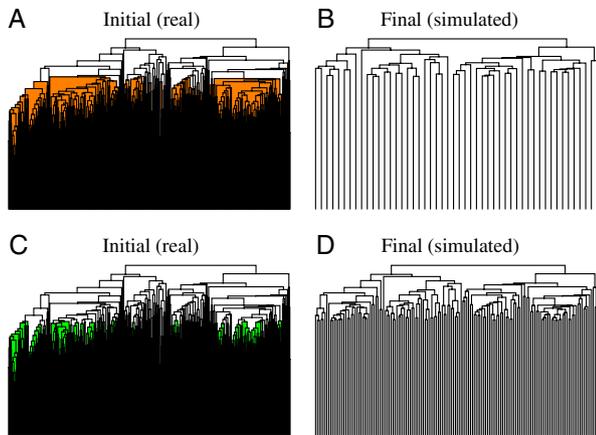


Fig. 1. The hierarchical structure implied by interopinion correlations, and its constraining effects on cultural evolution. (A) Dendrograms resulting from the application of an average linkage clustering algorithm to the cultural vectors $\{\vec{v}_i\}$, represented as leaves of the tree along the horizontal axis, for the Germany group; when we simulate the modified Axelrod model (see text), real data are considered as the initial state, and a confidence level (corresponding to the horizontal line below which the shaded region originates) is imposed. (B) Due to ultrametricity, in the corresponding final state of the model all the individuals within a common shaded branch in the initial dendrogram collapse to the same cultural vector, with negligible effects on the upper part of the dendrogram. (C) The same initial state as (A) is considered, but a lower confidence level is imposed. (D) Correspondingly, the final state of the model consists of a larger number of distinct cultural vectors, each containing on average less individuals with collapsed vectors. Thus the number of distinct final vectors (the leaves of the final dendrogram) coincides with the number of branches intersecting the horizontal line in the initial dendrogram. If shuffled or random opinions are taken as the initial state of the model (not shown), this is no longer true since the convergence of cultural vectors also affects the dendrogram's structure above the horizontal line, signalling a lack of ultrametricity.

branches, indicating that cultural space is heterogeneously populated by dense communities of similar individuals, separated by sparsely occupied regions. The hierarchical character of this distribution shows that denser regions are iteratively fragmented into denser regions nested within them. This peculiar organization indicates that the original distances are approximately *ultrametric* (54); i.e., the tree-like representation rendered by the dendrogram is not just an artifact of the clustering algorithm, but a natural property of the data. This means that the height of the first branching point connecting two individuals i and j approximately corresponds to the original distance d_{ij} between \vec{v}_i and \vec{v}_j . By contrast, the dendrograms for shuffled and real data are trivially structured (see Fig. S2) with no well-defined internal separation between different hierarchical levels. In this case the dendrogram is not representative of the original distribution of vectors, and is merely an uninformative outcome of the algorithm which is forcing nonultrametric data into a tree-like description. In such a situation, the vertical dimension of the dendrogram loses its correspondence with the original intervector distances, and provides a highly distorted image of the latter.

The Effects of Ultrametricity. The ultrametric hierarchy discussed above has important static and dynamic consequences. As we show in the *SI Text*, the branches of the dendrogram “cut” horizontally at a distance ω coincide with the connected components of the ω -dependent cultural graph we defined in the beginning. The (normalized) size of the largest connected component represents a global measure of influence among individuals as a function of ω , while the density of links in the cultural graph represents (for the same value of ω) a purely local measure of influence. We found that real data are always characterized by

higher levels of both local and global influence than shuffled and random data (see *SI Text*). Different groups also show differences among themselves, and their global levels of influence remain significantly different even after standardizing (controlling for) their local influence level. Therefore any process which depends on the cultural distance between individuals might have very different global outcomes even when taking place on locally identical structures.

All the above results show that even randomly sampled individuals are not characterized by uniformly random cultural vectors. While this is not surprising, the peculiar hierarchical distribution implied by empirical interopinion correlations is highly nontrivial, and unpredictable a priori. As we show in what follows, the dynamics of opinions and culture is strongly dependent on initial conditions, and the observed ultrametricity dramatically alters the predictions of models simulating both short- and long-term dynamics. In particular, it determines a condition for the coexistence of cultural heterogeneity and collective social phenomena, thus solving the apparent paradox. Moreover, as we discuss in detail in the *SI Text*, it constrains the dynamics so severely that its outcome is nearly insensitive to the details of the model considered. This result inverts the frequent expectation that initial conditions are less relevant than model specifications, and actually makes the latter less important (with the advantage of reducing the arbitrariness that, in principle, underlies the mathematical definition of any model). Quite unexpectedly, a range of more complicated scenarios [such as models with different interaction probabilities (2, 31, 38, 39), different network topologies (2, 31, 37, 42), coevolution of networks and opinions (43–47), cultural drift (40, 41, 48), higher-order interactions (2, 38, 49), and external sources of information (50)] lead to results that are essentially equivalent (or directly mapped) to what is obtained simply assuming that individuals are subject to dyadic interactions on a complete graph (see *SI Text*). We will therefore illustrate our results in this simple case, which corresponds to the assumption that social interactions are infinite-ranged, and only limited by the bounded confidence hypothesis.

Short-term Collective Social Behavior. We first study the effects of the empirical structure of real opinions on short-term collective social behavior. We consider a simple prototypic model where, on short timescales, cultural vectors do not evolve but nonetheless determine the choices that individuals make under the influence of each other. To this end, we extend the Cont-Bouchaud (CB) model (9), originally proposed to model herding effects in financial markets, to a more general “coordination model” which incorporates a dependence on real cultural vectors $\{\vec{v}_i\}$ (see *SI Text*). Individuals are assumed to make binary choices such as yes/no, buy/sell, approve/reject, etc. We can represent the choice expressed by the i th individual as $\phi_i = \pm 1$. The effects of mutual influence and bounded confidence are modeled by allowing pairs of individuals whose cultural distance d_{ij} is smaller than a threshold ω (which is the only parameter of the model) to exchange information before making their choices. As a result of this information exchange, all the agents belonging to the same connected component of the resulting ω -dependent cultural graph (see *SI Text*) collectively “agree” on the choice to make. If A labels a connected component of the graph, the choice of all agents belonging to A is the same ($\phi_i = \phi_A \forall i \in A$), while different connected components make statistically independent choices. We can imagine, even if this is not strictly necessary, that the process is repeated several times and with no memory: ϕ_i may first represent whether i liked or disliked an item, then whether i liked or disliked a different item, etc. The key property is that ϕ_i is short-lived; i.e., it represents any choice made over a timescale much shorter than that required to modify the cultural vector \vec{v}_i . It is a spurious variable representing action, not culture.

The overall outcome of a single run of the process (e.g., the net demand for a specific item) is the sum of individual preferences, and can be quantified by the average choice

$$\Phi = \frac{1}{N} \sum_{i=1}^N \phi_i = \frac{1}{N} \sum_A S_A \phi_A \quad [1]$$

where the second sum runs over all connected components, and S_A is the size of component A . The sign of Φ reflects the choice of the majority, and a key property characterizing the outcome of the model is the probability $P_\omega(\Phi)$ that the average choice takes the particular value Φ , for a given value of ω . Following the procedure described in the *SI Text*, we computed $P_\omega(\Phi)$ for various values of ω (from $\omega = 0$ to $\omega = 1$ in increments of 0.01) and for all the 13 groups in our dataset (both real and shuffled), plus the completely random set. In Fig. 2A we report the results for real Germany data. As can be seen, there exists a critical value ω_c (in the case shown, $\omega_c = 0.14 \pm 0.01$) such that, for $\omega < \omega_c$, $P_\omega(\Phi)$ is symmetric about zero (as for $\omega = 0$) and, for $\omega > \omega_c$, $P_\omega(\Phi)$ has two symmetric peaks. Right at $\omega = \omega_c$, $P_\omega(\Phi)$ displays a flattened region. This behavior is typical of symmetry-breaking phase transitions. Here the order parameter of the transition is the most probable value(s) Φ_\pm of Φ . This is shown in Fig. 2B, where we also plot the behavior for shuffled and random data. As we show in the *SI Text*, we found that the critical thresholds for shuffled data (which are equal for all groups) are always larger than those

for real data (which are different across groups), and that the ones for random data are even larger. This shows that empirical interopinion correlations, which are responsible for the ultrametric distribution of individuals in cultural space, strongly facilitate collective social behavior by systematically lowering the resistance to coordination.

This simple model indicates that, depending on the local interaction range, different individual choices or actions can either be “averaged out” and disappear at the macroscopic level or give rise to a collectively coordinated behavior. Understanding this transition in real societies is one of the fundamental open questions of modern social science (1). In economics, this problem is related to whether it is legitimate to use the concept of “representative agent” as an idealized individual that makes the average choice of the society (16). Our simplified model, when simulated on real data, suggests that both regimes are possible, and that a simple local parameter can trigger very different global outcomes. In particular, the variance of the collective outcome, which grows with the separation between the peaks of $P_\omega(\Phi)$, can either decay to zero or be amplified macroscopically. These considerations indicate that a good measure of the level of collective behavior achievable for a given value of ω is the width of $P_\omega(\Phi)$. Thus we can define

$$C(\omega) \equiv \sigma_\omega(\Phi) = \sqrt{\sum_A \left(\frac{S_A}{N} \right)^2 \omega} \quad [2]$$

as a measure of short-term social *coordination*. The latter equality in the above formula (see *SI Text* for a rigorous proof) states that, intriguingly, $C(\omega)$ is uniquely determined by the sizes $\{S_A\}$ of the connected components of the underlying cultural graph obtained for that particular value of ω , and is therefore actually independent of the dynamical model considered. This quantity, which ranges between $C(0) = 0$ (no coordination) and $C(1) = 1$ (complete coordination), will be useful in what follows.

Long-term Cultural Diversity. We now take a dynamic perspective and focus on a longer temporal scale over which the cultural vectors themselves can change. In this case we use a modified version (38, 39) of Axelrod’s model (31), which is designed to simulate the evolution of vectors of cultural traits on social networks, again in a way that real data can enter into the model. In an elementary time-step, two individuals i and j are randomly selected. If the normalized overlap $o_{ij} \equiv 1 - d_{ij}$ (where d_{ij} is the above-defined distance between the vectors \vec{v}_i and \vec{v}_j) is smaller than or equal to θ , no interaction takes place. Otherwise, with probability equal to o_{ij} , the two individuals interact: a component $v_i^{(k)}$, chosen randomly among the components where \vec{v}_i and \vec{v}_j differ, is changed and set equal to i ’s corresponding component: $v_j^{(k)} = v_i^{(k)}$. Otherwise nothing happens, and two other individuals are selected. In the allowed final configurations, any two cultural vectors are either completely identical or separated by a distance larger than $\omega \equiv 1 - \theta$, and the average $\langle N_D \rangle_\omega$ (over many realizations) of the number N_D of distinct vectors in the final stage, or equivalently the fraction

$$D(\omega) \equiv \frac{\langle N_D \rangle_\omega}{N} \quad [3]$$

is a convenient way to measure the long-term cultural *diversity* as a function of ω . A more detailed description of the model can be found in the *SI Text*.

We ran several realizations of the model, by taking both real and randomized cultural vectors $\{\vec{v}_i\}$ as the starting configuration. As we show in the *SI Text*, we find that real data are those that achieve the largest level of long-term cultural heterogeneity. Indeed, for real data the realized value of $\langle N_D \rangle$ is the largest pos-

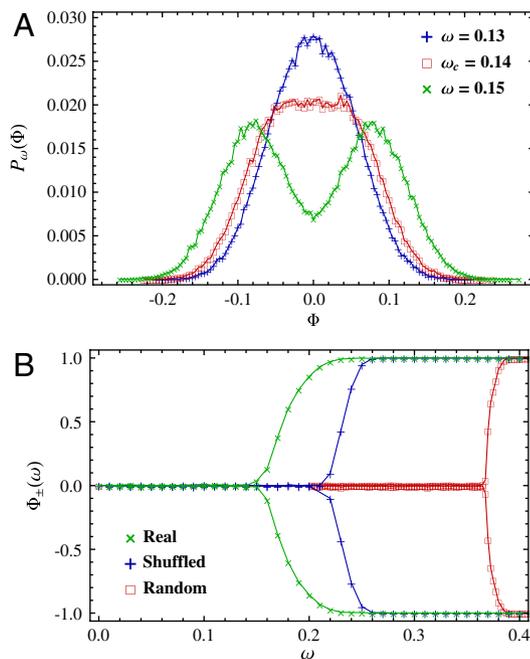


Fig. 2. At a critical confidence level, a spontaneous breaking of choice symmetry occurs. (A) When our “coordination model” is simulated on real data (in the example shown, the Germany group), we observe an abrupt change in the probability $P_\omega(\Phi)$ of a collective outcome at a critical confidence value ω_c . For $\omega < \omega_c$ individual choices are uncorrelated and sum up to a vanishing global outcome $\Phi = 0$, at which $P_\omega(\Phi)$ has a single peak. For $\omega > \omega_c$ local interactions result in global correlations that spread across the entire system, and a macroscopically coordinated output, whose probability is peaked about the two nonzero values $\Phi_\pm(\omega)$, emerges. Right at $\omega = \omega_c$, $P_\omega(\Phi)$ displays a flat region typical of critical phenomena. (B) The most probable value of Φ is the order parameter of the phase transition. For $\omega < \omega_c$ it is vanishing, while for $\omega > \omega_c$ it branches into the two symmetric values $\Phi_\pm(\omega)$. In addition to the results for real Germany data, here we also show the results for shuffled and random data. Real data always have a lower critical threshold than randomized data, indicating an enhanced possibility to behave collectively. All the other groups show the same behavior.

sible ($\langle N_D \rangle \approx N_C$) indicating that cultural convergence is confined within the initial connected components, each of which eventually becomes a single cultural domain. By contrast, in randomized data there are less final cultural domains than initial connected components, indicating that the latter often “merge” into larger cultural domains. The reason for the remarkably different behavior of real and randomized data is, once again, the ultrametric character of the former. As illustrated in Fig. 1, ultrametricity implies that the branches obtained cutting the real-data dendrogram at some value of ω will collapse into a single cultural vector. This means that the initial structure of the dendrogram above ω will be “frozen” and unaffected by cultural evolution. This confines cultural convergence locally within the lower branches. By contrast, in randomized data the lack of ultrametricity implies that branches are not well separated, so that the local convergence of vectors within a branch reduce the separation of the branch itself from nearby branches. Thus in this case branches are unstable, and often merge modifying the entire structure globally.

Combining the above findings with our previous results about collective behavior highlights the remarkable effects of ultrametricity. Given a group of individuals, we measured both the (initial) short-term social coordination $C(\omega)$ defined in eq. 2 and the (final) long-term cultural diversity $D(\omega)$ defined in Eq. 3 for various values of ω . Then we plotted $D(\omega)$ versus the value $C(\omega)$ obtained for the same ω , as in Fig. 3. If we look at random data, we retrieve the naive result that the coexistence of cultural heterogeneity and social collective behavior is impossible, since we have either $D \approx 0$ or $C \approx 0$. By contrast, real cultural vectors simultaneously allow high levels of both short-term coordination and long-term diversity, including the approximately balanced regime $C \approx D \approx 1/2$ which in the case shown is achieved for the moderate influence level $\omega \approx 0.17$. Shuffled data follow an intermediate curve, showing that both the heterogeneous frequencies of real opinions and the correlations among the latter play a significant role in enhancing the coexistence of diversity and coordination.

Note that, for real data, as the branches (below the threshold ω) of the dendrogram merge under the effect of cultural convergence, the cultural graph acquires more and more links until its initial connected components eventually become fully connected cliques. However, the number and sizes of such connected components (obtained for that value of ω) will remain unchanged[†]. Therefore, if we repeat the short-term coordination process at any frequency during the long-term cultural process, we will always obtain the same value of $C(\omega)$: the amplitude of collective social phenomena will not be decreased by cultural convergence. Seen from another perspective, this indicates that the persistence, or even an increased frequency, of collective behavior does not directly imply a reduction of cultural diversity. Thus, surprisingly, we find that empirical hierarchical correlations simultaneously enhance collective behavior and sustain cultural heterogeneity. The incompatibility of these two phenomena, which holds for random data, breaks down for real-data.

The above results were obtained assuming infinite-ranged social interactions. Online platforms bypassing offline social ties are modifying the traditional scenario and leading us closer to this extreme setting. Nonetheless, as we discuss in the *SI Text*, our results are surprisingly robust to changes in the dynamical rules of the model considered. This robustness suggests that, even if online interactions are tremendously expanding the substrate for

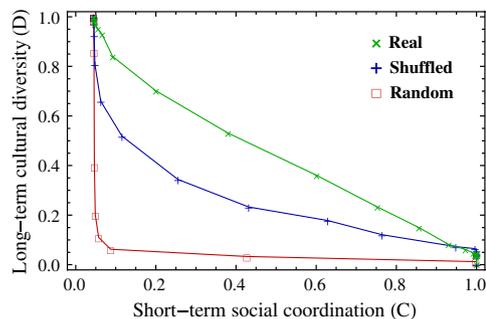


Fig. 3. Phase diagram summarizing our results. The long-term cultural diversity D is shown as a function of short-term social coordination C for real, shuffled and random data. If random cultural vectors are considered, cultural heterogeneity and collective behavior are mutually excluding: one has either $D \approx 0$ or $C \approx 0$. This approximately corresponds to the traditional situation explored when considering a random graph of interaction among individuals. By contrast, real cultural vectors simultaneously allow high levels of both short-term coordination and long-term diversity, including the approximately balanced regime $C \approx D \approx 1/2$. Shuffled data follow an intermediate curve, showing that both the heterogeneous frequencies of real opinions and the correlations among the latter play a significant role in enhancing the coexistence of diversity and coordination in the real world.

short-term collective social behavior, they are not necessarily going to suppress long-term cultural diversity. It also indicates that, despite the fact that the structure of both offline (55–57) and online (28–30) social networks is intensively studied, our understanding of cultural dynamics is still severely limited by our little knowledge of the large-scale organization of real societies in multidimensional cultural space. These findings highlight the scarce predictive power of models that consider random specifications, and show the unique importance of empirical analyses of high-dimensional cultural vectors offering the possibility to explore cross-correlations among opinions and their consequences.

Conclusions

By using a large and detailed dataset, we have characterized the empirical properties of the large-scale distribution of individuals in multidimensional cultural space. We found that real interopinion correlations organize individuals hierarchically and ultrametrically in cultural space, a result which is not retrieved when randomized or one-dimensional opinions are considered. By using simple models, we showed that ultrametricity has profound and nontrivial consequences on short- and long-term cultural dynamics. In the short term, we found the existence of a symmetry-breaking phase transition where collective behavior arises out of purely local interactions. The critical threshold of this transition is remarkably lower in real data than in randomized cases, indicating that ultrametricity enhances short-term collective behavior. However, in the long term the same ultrametric property suppresses cultural convergence by restricting it within disjoint domains, implying a strong sensitivity to initial conditions. These opposite effects imply that, whereas in random data the coexistence of short-term coordination and long-term diversity is unfeasible, in real data it is strongly enhanced and can be achieved in a broad region of parameter space. Indeed, real data appear to optimize the coexistence of both phenomena. As another effect of ultrametricity, these results are surprisingly robust to changes in the details of the particular model considered: even more complicated theoretical scenarios would lead to a similar outcome. Thus the apparent paradox of the coexistence of short-term collective social behavior and long-term cultural diversity might have, as a simple and parsimonious explanation, the empirically observed hierarchical distribution of individuals in cultural space.

ACKNOWLEDGMENTS. All authors thank Maria I. Loffredo for her support and guidance. A.A. thanks all her colleagues who have contributed to the series

[†]Since we expect ultrametricity to be a robust property of real cultural traits also in other situations, it would be very interesting to test this prediction on real data reporting the evolution of culture. If we imagine to track the dynamics backwards in time, we also expect that the number and sizes of connected components will be preserved. Since social influence only modifies the lower branches of the dendrogram, the upper structure must be inherited from previous stages. This structure is exactly what we interpret as the residual cultural correlations that are entangled with social groups from the beginning of human history (see *SI Text*).

of Eurobarometer surveys on science and technology over the years. D.G. warmly thanks Felix Reed-Tsochas for stimulating discussions at the CABDyN Complexity Centre (University of Oxford) and acknowledges financial support through the European Commission FP7 Future and Emerging Technol-

ogies Open Scheme Project ICTeCollective (Contract 238597). This work was also supported by the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, Netherlands) with funds from Duyfken Trading Knowledge BV, Amsterdam, Netherlands.

1. Buchanan M (2007) *The Social Atom* (Bloomsbury).
2. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81:591–646.
3. Lazer D, et al. (2009) Life in the network: The coming age of computational social science. *Science* 323:721–723.
4. Fortunato S, Castellano C (2007) Scaling and universality in proportional elections. *Phys Rev Lett* 99:138701.
5. Ratkiewicz J, Fortunato S, Flammini A, Menczer F, Vespignani A (2010) Characterizing and modeling the dynamics of online popularity. *Phys Rev Lett* 105:158701.
6. Onnela J-P, Reed-Tsochas F (2010) Spontaneous emergence of social influence in online systems. *Proc Natl Acad Sci USA* 107:18375–18380.
7. Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proc Natl Acad Sci USA* 105:15649–15653.
8. Mantegna R, Stanley HE (2000) *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, U.K.).
9. Cont R, Bouchaud J-P (2000) Herd behavior and aggregate fluctuations in financial markets. *Macroecon Dyn* 4:170–196.
10. Sinha S, Chatterjee A, Chakraborti A, Chakraborti B (2010) *Econophysics: An introduction* (Wiley-VCH Verlag GmbH, New York).
11. Chakraborti B, Chakraborti A, Chatterjee A (2006) *Econophysics and Sociophysics: Trends and Perspectives* (Wiley-VCH Verlag GmbH, New York).
12. Galam S (2007) From 2000 Bush-Gore to 2006 Italian elections: Voting at fifty-fifty and the contrarian effect. *Qual Quant* 41:579–589.
13. Onnela J-P, et al. (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA* 104:7332–7336.
14. Bohannon J (2006) Tracking people's electronic footprints. *Science* 314:914–916.
15. Fowler J, Christakis N (2010) Cooperative behavior cascades in human social networks. *Proc Natl Acad Sci USA* 107:5334–5338.
16. Kirman AP (1992) Whom or what does the representative individual represent? *J Econ Perspect* 6:117–136.
17. Greig JM (2002) The end of geography? Globalization, communications, and culture in the international system. *J Conflict Resolut* 46:225–243.
18. Brady D, Beckfield J, Zhao W (2007) The consequences of economic globalization for affluent democracies. *Annu Rev Sociol* 33:15. 1–15.22.
19. Rogers EM (1971) *Communication of Innovations: A Cross-Cultural Approach* (Free Press, New York).
20. Rogers EM (1995) *Diffusion of Innovations* (Free Press, New York).
21. Blossfeld HP, Klijzing E, Mills M, Kurz K (2005) *Globalization, Uncertainty and Youth in Society* (Routledge, New York).
22. McNeill W (1988) Diffusion in history. *The Transfer and Transformation of Ideas and Material Culture*, eds Peter J Hugill and D. Bruce Dickson (Texas A&M Press, College Station), pp 75–90.
23. Foster R (1991) Making national cultures in the global ecumene. *Annu Rev Anthropol* 20:235–260.
24. Elkins D (1997) Globalization, telecommunications, and virtual ethnic communities. *Int Polit Sci Rev* 18:139–52.
25. Barth Fredrik (1969) *Ethnic and Group Boundaries* (Waveland, Long Grove, IL), Reprint 1998.
26. Durkheim E (1912) *The Elementary Forms of Religious Life* (Oxford University Press, New York), Reprint 2001.
27. Boyd R, Richerson PJ (2005) *The Origin and Evolution of Cultures* (Oxford University Press, New York).
28. Backstrom L, Huttenlocher D, Lan X, Kleinberg J (2006) Group formation in large social networks: Membership, growth, and evolution. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery in Data* (ACM Press, New York), pp 44–54.
29. Szell M, Thurner S (2010) Measuring social dynamics in a massive multiplayer online game. *Soc Networks* 32:313–329.
30. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *Proc Natl Acad Sci USA* 107:13636–13641.
31. Axelrod R (1997) The dissemination of culture. *J Conflict Resolut* 41:203–226.
32. Festinger L, Schachter S, Back K (1963) *Social Pressures in Informal Groups: a Study of Human Factors in Housing* (Stanford University Press, Palo Alto, CA).
33. Homans GC (1962) *Sentiments and Activities* (Free Press, New York).
34. McPherson JM, Smith-Lovin L (1987) Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *Am Sociol Rev* 52:370–79.
35. McPherson JM, Smith-Lovin L, Cook J (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–44.
36. Sherif M, Hovland CI (1961) *Assimilation and Contrast Effects in Communication and Attitude Change* (Yale University Press, New Haven, CT).
37. Weisbuch G (2004) Bounded confidence and social networks. *Eur Phys J B* 38:339–343.
38. Flache A, Macy M (2007) Local Convergence and Global Diversity: The Robustness of Cultural Homophily. Arxiv preprint physics/0701333.
39. De Santis L, Galla T (2009) Effects of noise and confidence thresholds in nominal and metric Axelrod dynamics of social influence. *Phys Rev E* 79:046108.
40. Klemm K, Eguiluz VM, Toral R, San Miguel M (2003) Global culture: A noise-induced transition in finite systems. *Phys Rev E* 67:045101(R).
41. Klemm K, Eguiluz VM, Toral R, San Miguel M (2005) Globalization, polarization and cultural drift. *J Econ Dyn Control* 29:321–334.
42. Guerra B, Poncela J, Gómez-Gardenes J, Latora V, Moreno Y (2010) Dynamical organization towards consensus in the Axelrod model on complex networks. *Phys Rev E* 81:056105.
43. Centola D, González-Avella JC, Eguiluz VM, San Miguel M (2007) Homophily, cultural drift and the co-evolution of cultural groups. *J Conflict Resolut* 51:905–929.
44. Zanette DH, Gil S (2006) Opinion spreading and agent segregation on evolving networks. *Physica D* 224:156–165.
45. Santos FC, Pacheco JM, Lenaerts T (2006) Cooperation prevails when individuals adjust their social ties. *PLoS Comput Biol* 2:e140.
46. Garlaschelli D, Capocci A, Caldarelli G (2007) Self-organized network evolution coupled to extremal dynamics. *Nat Phys* 3:813–817.
47. Kozma B, Barrat A (2008) Consensus formation on adaptive networks. *Phys Rev E* 77:016102.
48. Mas M, Flache A, Helbing D (2010) Individualization as driving force of clustering phenomena in humans. *PLoS Comput Biol* 6(10):e1000959.
49. Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Sociol* 113:702–734.
50. González-Avella JC, Cosenza MG, Tucci K (2005) Nonequilibrium transition induced by mass media in a model for social influence. *Phys Rev E* 72:065102(R).
51. INRA, Brussels (Europe) and Report International for the European Commission, Directorate-General XII Science Research and Development. *Europeans, Science and Technology—Public Understanding and Attitudes*, Available at http://ec.europa.eu/public_opinion/archives/ebs/ebs_076_en.pdf.
52. Gaskell G, et al. (2000) Biotechnology and the European public. *Nat Biotechnol* 18:935–938.
53. Gaskell G, et al. (2011) The 2010 Eurobarometer on the life sciences. *Nat Biotechnol* 29:113–114.
54. Rammal R, Toulouse G, Virasoro M (1986) Ultrametricity for physicists. *Rev Mod Phys* 58:765–788.
55. Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78:1360–80.
56. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, U.K.).
57. Caldarelli G (2007) *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, New York).

Supporting Information

Valori et al. 10.1073/pnas.1109514109

SI Text

Framing our Case Study. In this section we make general considerations that help framing our case study correctly. We first discuss how social influence, homophily, and bounded confidence are believed to affect social dynamics, and how offline and on-line environments are expected to contribute differently to the process. Then, we explain why these considerations lead us to the choice of the Eurobarometer data (1–3), and in particular its 1992 edition (Eurobarometer 38.1: “Europeans, Science and Technology—Public Understanding and Attitudes” (1)), as an appropriate dataset for our analysis.

Social networks and cultural graphs: the on-line shift. As we mentioned in the main article, the bounded confidence hypothesis states that two individuals are potentially influenced by each other only if the distance between their cultural variables (or vectors) is smaller than a certain threshold ω (4–6). Thus, while social influence takes place on a social network connecting individuals, bounded confidence involves a different graph, that we denote as the *cultural graph*, where pairs of individuals separated by a distance smaller than the confidence ω are connected by “cultural channels,” irrespective of whether they are neighbors in a social network. When combined together, these hypotheses imply that the actual network of interactions is given by the overlap between social ties and (confidence-dependent) cultural channels. If s_{ij} and $c_{ij}(\omega)$ denote the elements of the adjacency matrix of the social and cultural network respectively (equal to 1 if a link between vertices i and j is there, and 0 otherwise), the overlap network is described by an adjacency matrix with entries $a_{ij}(\omega) = s_{ij}c_{ij}(\omega)$, where $c_{ij}(\omega) = 1$ if $d_{ij} < \omega$ and $c_{ij}(\omega) = 0$ if $d_{ij} > \omega$.

Models that simulate the evolution of societies must therefore be complemented by empirical analyses characterizing both social networks and cultural graphs. However, while a huge literature is devoted to the study of real networks formed by social ties (using both traditional small-scale surveys (7) and more recently large-scale communication data (8, 9)), little is known about the empirical properties of those formed by cultural channels. As a consequence, when considering models of opinion dynamics, social networks have been so far considered as proxies for the actual interaction graphs; i.e., $a_{ij} \approx s_{ij}$, which amounts to assume $c_{ij} \approx 1$ (or equivalently an unbounded confidence $\omega = +\infty$) for all pairs of individuals. This assumption is helpful in the traditional offline situation where social ties are expected to be dominant over cultural channels, for instance when social groups are formed independently of the cultural traits of people (e.g., acquaintances made in public schools). However, it prevents our understanding of the opposite extreme, i.e. when people look for culturally similar peers, that they do not know initially, to interact with (for instance by joining an open discussion group on a focused topic of interest). In such a situation the empirical knowledge of $c_{ij}(\omega)$ may become even more important than that of s_{ij} . While these nonsocial interactions used to represent only a secondary means of opinion diffusion up to some years ago, they are now becoming more and more pervasive as novel electronic platforms are being developed and used at a large scale (10–16). Indeed, in recent years many new possibilities of interaction have rapidly emerged, such as the exchange of opinions through the WWW (on-line forums, blogs, discussion groups, etc.) and other media. In these platforms, people already interested in a topic search new peers (who they may never meet physically afterwards) to discuss with, experiencing novel ways by which opinions can interact. If the

opinions under investigation regard technical or specialized topics, on-line interactions can be even stronger than those occurring through direct knowledge, as virtual communities gather together people with oriented and focused interests, who often recognize each other as the most natural and qualified peers to share ideas with, even if no direct knowledge exists between them.

The above picture suggests an extension of the idea of homophily, according to which “likes attract,” to an on-line setting (17). Indeed, recent empirical studies of on-line communities (12) confirm that even in such context “social” interactions produce stable groups which are strongly connected internally and sparsely connected across, as predicted by the “strength of weak ties” hypothesis (18) and confirmed empirically in the offline case even at a society-wide level (8). When additional on-line possibilities of interaction are considered, cultural channels may become the dominant means of interaction, as people with similar interests are more likely to access the same platform(s) and exchange opinions more frequently than culturally dissimilar individuals. In this opposite extreme, the social network through which opinions can in principle interact is virtually extended to a complete graph (i.e., $s_{ij} = 1 \forall i, j$) where everybody is connected to everyone else (in social space, the interaction becomes infinite-ranged). Cultural graphs therefore become a natural proxy for the actual interaction graph: $a_{ij}(\omega) \approx c_{ij}(\omega)$. Despite its increasing importance, our understanding of this novel type of long-range opinion dynamics, which bypasses social ties and is dominated by the structure of real opinions and cultural graphs, is incomplete.

Our analysis bridges this gap by using the Eurobarometer (1–3) data which reports the opinions, beliefs and attitudes of thousands of Europeans. These data allowed us to produce cultural graphs where we investigated the outcome of (socially) infinite-range, on-line-like dynamics on both short and long timescales. An unexpected outcome of our analysis is that the empirically observed ultrametricity of the distribution of individuals in cultural space poses severe constraints on the dynamics of opinions. In particular, besides making short-term collective social behavior compatible with long-term cultural diversity (as we discuss at length in the main article, with additional details in this SI file), ultrametricity also implies that a number of more complicated scenarios, going beyond our initial assumption of a complete graph of social interaction and introducing more sophisticated dynamical mechanisms, can be easily mapped to the results we find in our simple setting. Therefore our choice of simple models and infinite-ranged social interactions actually encompasses a more diverse range of interesting situations. This is discussed in detail in the section *More Complicated Scenarios* at the end of this SI file.

The Eurobarometer dataset. Our aim is to understand the effects of realistic, rather than completely random, cultural traits on the outcome of social and opinion dynamics. Since neighbors in a social network tend to have similar characteristics, and this may partially be already a product of homophily, a problem of circularity arises whenever one tries to assess whether such similarity is the cause or the consequence of a social tie being there. For this reason, in our analysis we need to consider, as the initial state of dynamical models, a realistic cultural specification which is neither completely random, nor already trivially generated by the mechanism we want to study. We therefore need to access the opinions of a set of real individuals who do not know each other and are well separated socially.

For this reason, we focused on the Eurobarometer dataset (1–3), a large project that is specifically designed to survey, through questionnaires filled during face-to-face interviews, several beliefs, attitudes, and opinions across Europe. The Eurobarometer survey makes use of sampling protocols ensuring that individuals are selected avoiding the bias due to (among other factors) social closeness. It is also designed in such a way that the target population is a representative sample in terms of individual characteristics such as age, gender, SES (socioeconomic status), etc. These conditions guarantee that we can regard the Eurobarometer data as a stochastic realization of cultural specifications which, rather than being completely random as ordinarily happens in models, reflect the cultural characteristics of real societies through observed statistical weights. In other words, the data are maximally random under the constraint of reflecting a country's overall cultural characteristics, and unbiased by social influence effects (producing convergent cultural traits) that might be already present if the individuals were sampled nearby in a social network.

Of course, simply because the sampled individuals are part of the same society and share basic habits and attitudes, correlations between the same cultural traits across different individuals are still present in the data. These dependencies include similarities induced by potential indirect interactions; i.e., paths of social ties connecting individuals through intermediate neighbors not present in the data. But such residual correlations are exactly what we desire as a starting configuration, and are the main object of our empirical analysis. Since humans belong to social groups from the very beginning of the history of mankind, trying to get rid of this unavoidable source of “baseline” similarity would imply another circular argument leading us to the distant (in time and relatedness to the present study) problem of the origin of culture itself. This observation also conclusively shows that completely random specifications of cultural traits are in any case inappropriate as seeds in models because we cannot even regard such specifications as a sort of “primordial” features that precede human interactions.

Several editions of the Eurobarometer survey exist, which differ in the nature and subject of the questions, in the target population (the sampled individuals are different at each round), and in the number of target countries (due to the temporal increase of the number of European member states). For our purposes, these major structural changes between successive snapshots impede the use of different time points as a way to consistently track the evolution of opinions in a well-defined set of individuals. This limitation highlights the desirability of designing future surveys and/or experiments that allow us to explicitly keep track of individual changes in a fixed set of opinions for the same group of people. However, as discussed, it is possible for us to use a single snapshot if the data as the initial condition for dynamical models, and study how the empirical properties of real cultural data change the results that are obtained when standard random specifications are used. This leaves us the freedom to select, among the available snapshots, a particularly convenient one. We chose the 1992 snapshot for two reasons. First, this early year in the dataset rules out the possibility of on-line interactions among the sampled individuals, since at that time the Internet was still in its infancy, and no extensive use of the WWW was possible. Thus, besides offline social interactions, we can also exclude on-line contacts. Second, the subject of the 1992 survey [Eurobarometer 38.1: “*Europeans, Science and Technology—Public Understanding and Attitudes*” (1)] minimizes, as we now show, the effects of selective exposure to media and leaders presenting similar patterns of opinions (such as political parties, nonscientific articles in newspapers, etc.), which are deliberately beyond the scope of our analysis.

The expectation that the scientific opinions surveyed in 1992 are relatively independent of political beliefs or social identities (with respect to other Eurobarometer editions such as “European

Parliamentary Elections,” “Relations with Third World Countries and Energy Problems,” “Immigrants and Outgroups in Western Europe,” etc.) is confirmed in the official report (1) summarizing the results of the analysis of the dataset. Citing from that source: “This report presents the results of a major Community-wide opinion survey carried out in 1992 on behalf of the European Commission concerning the Public Understanding of, and attitudes towards, Science and Technology. [...] Eurobarometer is a regular Community-wide survey of opinion on a large number of issues of interest to the European Union. A sample of over 13,000 Community citizens, representative of the total population of the Member States, were interviewed on subjects of science and science policy with some emphasis on environmental issues. [...]” The main findings are summarized below:

- A large majority of EC citizens (70 per cent) watches the news on television on a daily basis. While less popular, the news in daily papers and on the radio still attracts the daily interest of nearly half of the Community population.
- Compared with the results of 1989, over 50% (or very close to it) of the populations of the same six Member States consult all three news sources on a daily basis: Denmark, Germany, Ireland, Luxembourg, the Netherlands, and the United Kingdom.
- Scientific news items turn out to be the ones in which EC citizens are most interested. With the exception of “medical discoveries,” this interest diminishes as people get older.
- It turns out that Europeans in general are convinced that the level of information they have reached with regard to a number of issues in the news is no match for their interest in these same issues. Particularly where scientific themes are concerned, very few people consider themselves to be very well informed.
- A strong positive link exists, however, between people's level of interest in scientific issues and their self-estimated level of informedness on these issues.
- As to Europeans' media use which is specifically geared toward scientific information, the television again plays a major role. Taking the regular and occasional viewers together, 59% of Europeans do watch television programs on science and technology. Almost half (45%) read articles on science in newspapers, and only one out of every five (21%) is to be found reading scientific magazines. There seems to be a direct link between the easiness of the medium and the respondents paying attention to science and technology related stories on it.”

The above findings have two important consequences for our analysis. First, they suggest that, for the individuals in the data, the main sources of scientific information are mass media which are uniformly accessed on a regular basis by the entire population (mainly television and newspapers), rather than specialized media accessed heterogeneously by diversely interested niches. This implies that all the individuals in the sample are directly exposed to a common source of scientific knowledge, which obviously determines some level of cultural convergence not due to actor-actor interactions. Even if this type of convergence is not explicitly modeled in our study, we can reasonably expect that scientific sources in mass media tend to attract all opinions in a common direction, and that this effect prevails on the opposite tendency of media to polarize around multiple centers, which is usually dominant for other topics such as politics. Thus all our results have a clear (and more correct) interpretation in terms of the effects of actor-actor interactions that take place on top of a common uniform “background” convergence (see a more rigorous discussion in the last subsection of this SI file). Second, it seems that the Europeans in the survey are all interested in scientific topics, but on the other hand think that their level of informedness on the latter is not sufficient. A high level of interest and a limited confidence in one's knowledge suggest that, on the topics considered, these individuals are sufficiently open to the influence of other people's opinions. This indicates that one of the implicit

assumptions in models of opinion dynamics, i.e., that people can be potentially influenced on the topic considered, is realized in this case study.

As we discuss below and in the main article, we used the multiple-choice answers to the Eurobarometer questionnaire as a way to reconstruct the “cultural vectors” of the individuals in the sample. From such vectors, we computed the cultural distances for all pairs of individuals. This procedure leads us again to discuss the nature of residual correlations in the data, this time in terms of the correlations between different traits within each cultural vector. That is, we have to make some considerations regarding the cultural “dimensions” covered by the Eurobarometer survey. Our definition of cultural distance (see below) implicitly assumes that different components of the vectors are a priori independent, or in other words that, irrespective of possible ex-post correlations between answers to different questions (that we observe and study), different questions are independent of each other in their ex-ante formulation. As a counterexample of highly (but trivially) correlated data, imagine a maximally redundant questionnaire made by many replicas of the same question: in such a case, the vectors are effectively unidimensional (unless individuals give different answers when the same question is asked two or more times). Our unavoidable assumption of independence, while not easily verifiable, is however largely justified by the fact that the Eurobarometer data are collected through face-to-face interviews. This type of survey, when conducted on a large scale, is very expensive and time consuming. Therefore, unlike other sources of data (such as automated downloads of behavioral or cultural information from the Internet), face-to-face interviews require that questionnaires are carefully designed in such a way that, given the total number of questions (i.e., the total cost of the survey), the maximum possible amount of information is obtained from the answers. This leads us to the expectation that the questions in the survey are maximally nonredundant and independent of each other, and motivates us to proceed further with our analysis.

More Details on the Analysis of Data. In this section we provide additional details, with respect to the main article, about how we reconstructed real multidimensional opinions from empirical data, how we measured interopinion correlations, and what is implied by the ultrametric structure we observed.

Definition of cultural vectors and distances. As we mentioned, the 1992 dataset reports the results of face-to-face interviews where more than 13,000 individuals (from the 12 European countries* that were part of the European Community in that year) were asked to fill a questionnaire containing several multiple-choice questions. Raw data were originally arranged in a SPSS spreadsheet file[†] of $N \approx 13,000$ rows, where each row i reported the (numerically coded) answers of the i th individual to all the multiple-choice questions. A small number of such questions regarded demographic data (such as gender, age, etc.), while the remaining $F = 161$ questions concerned the true topic of the survey. We therefore discarded demographic data, and transformed the “cultural” data contained into the F answers given by each of the N individuals into N F -dimensional vectors $\{\vec{v}_i\}$, using the same representation as in Axelrod’s model (19). Each component $v_i^{(k)}$ of these vectors was given a value such that the corresponding con-

tribution $d_{ij}^{(k)}$ to the overall distance

$$d_{ij} = \frac{1}{F} \sum_{k=1}^F d_{ij}^{(k)} \quad [\text{S1}]$$

was in the range $[0,1]$, with $d_{ij}^{(k)} = 0$ representing i and j giving an identical answer to question k , and $d_{ij}^{(k)} = 1$ representing i and j giving opposite answers. For “metric” questions (6), where answers were possible in an equally spaced scale of Q_k possibilities, the maximum information is retained by mapping the original answers to the possible values

$$v_i^{(k)} = 0, \frac{1}{Q_k - 1}, \frac{2}{Q_k - 1}, \dots, 1$$

and defining $d_{ij}^{(k)} \equiv |v_i^{(k)} - v_j^{(k)}|$. For nonmetric questions (associated to Q_k unordered possible alternatives), “opposite answers” simply means “different answers,” and we therefore mapped the possible values of $v_i^{(k)}$ to Q_k arbitrary symbols and defined $d_{ij}^{(k)} \equiv 1$ if $v_i^{(k)} \neq v_j^{(k)}$ and $d_{ij}^{(k)} \equiv 0$ otherwise. If all questions were nonmetric, this choice would be equivalent to $d_{ij} = 1 - o_{ij}/F$ where o_{ij} is the overlap (number of components with identical value) between \vec{v}_i and \vec{v}_j , a commonly used notion of cultural similarity (19) as mentioned in the main article. Note that for binary answers (such as “yes”/“no”) the metric and nonmetric definitions coincide.

Each individual in the data belongs to one of the 12 Member States. This information allowed us to generate groups of individuals sampled either from the same country, or from different ones. We labeled different groups of individuals with different Greek letters (α, β, \dots), and used the notation $i \in \alpha$ to indicate an individual i belonging to group α . In order to deal with samples of equal size, we selected $N = 500$ individuals per country[‡] and generated 12 groups accordingly. This reproduces a situation where, for instance, people with the same language join a medium-sized on-line discussion group mediating electronic, nonsocial interactions mentioned above. Similarly, it mimics individuals that are physically “put together” to participate in a discussion group or to a social experiment. Finally, this choice allows us to establish an upper bound for the cultural homogeneity predicted by the “traditional” dynamics taking place on any possible social network connecting the same individuals. We also generated an additional thirteenth group with $N = 500$ individuals sampled from all the 12 European countries, and denoted it as the *Europe* group. This reproduces a situation analogous to the one described above, but where several individuals across Europe can form a group irrespective of their nationalities—e.g., using a common language such as English.

Measuring interopinion correlations. Our multidimensional data allowed us to investigate cross-correlations among opinions, an information which is not available in one-dimensional studies. We measured the covariance matrix between $d_{ij}^{(k)}/F$ and $d_{ij}^{(l)}/F$, whose entries read

$$\sigma_{\alpha}^{(k,l)} \equiv \frac{\langle d_{ij}^{(k)} d_{ij}^{(l)} \rangle_{ij \in \alpha} - \langle d_{ij}^{(k)} \rangle_{ij \in \alpha} \langle d_{ij}^{(l)} \rangle_{ij \in \alpha}}{F^2} \quad [\text{S2}]$$

*In alphabetical order: Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, UK.

[†]Citing from the official report (1): “... like other Eurobarometer surveys, the raw data are available in machine-readable form at the Zentralarchiv für Empirische Sozialforschung, Köln Universität, Bachemer Strasse 40–5000 Köln-41, Germany (fax: (49) 221.476.94.44) and the Institute for Social Research, University of Michigan, Ann Arbor—Michigan 48104-1248, USA (Fax : (1) 313.763.13.46). The Commission warmly invites researchers to take the opportunity to explore this data”. On-line sources of the data also exist, such as <http://einstein.library.emory.edu/eurolinks> and <http://www.cessda.org/index.html>.

[‡]The dataset contains a sample of approximately 1,000 individuals per country, except Luxembourg (500 individuals) and Germany (2,000 individuals aggregating two sets of 1,000 individuals, one for East Germany and one for West Germany according to the division existing in 1992). Thus 500 is the maximum size we can take if we want samples of equal size without discarding any country.

where $\langle \cdot \rangle_{ij \in \alpha}$ denotes an average over all pairs of individuals in group α . From the above matrix it is possible to obtain the interopinion correlation matrix, whose entries read

$$\rho_{\alpha}^{(k,l)} \equiv \frac{\sigma_{\alpha}^{(k,l)}}{\sigma_{\alpha}^{(k,k)} \sigma_{\alpha}^{(l,l)}} \quad [\text{S3}]$$

and range between -1 (perfect anticorrelation) and 1 (perfect correlation). The correlation $\rho_{\alpha}^{(k,l)}$ tells us whether small (large) differences between the answers given to question k imply small (large) differences between answers to question l , for a given group α . A color plot of the correlation matrix is shown in Fig. S1A for the group sampled from Germany data, which is a reference case throughout our analysis (similar results are found for all the other sampled groups). As can be seen, there are several pairs of opinions (k,l) characterized by strong positive correlations ($\rho_{\alpha}^{(k,l)} \approx 1$), but only a few pairs with weak negative correlations ($\rho_{\alpha}^{(k,l)} \lesssim 0$). The predominance of positive correlations indicates that, in general, two individuals i and j giving similar/different answers to question k (small/large $d_{ij}^{(k)}$) tend to give similar/different answers to question l as well (small/large $d_{ij}^{(l)}$), while the opposite outcome (small $d_{ij}^{(k)}$ and large $d_{ij}^{(l)}$) occurs much less frequently. As we show in Fig. S1B, the interopinion correlation matrix for shuffled data lacks any structure, indicating the absence of statistically significant correlations (obviously, the same is true for random data).

The observed correlations have important effects on the distribution of individuals, i.e. of the vectors $\{\vec{v}_i\}$, in cultural space. We find (see Table S1) that the average interindividual distance $\mu_{\alpha} \equiv \langle d_{ij} \rangle_{ij \in \alpha}$ of random data is larger than real data, while it is easy to show theoretically (and confirm by looking at the measured values) that real and shuffled data always have the same value of μ_{α} , i.e., $\mu_{\alpha, \text{real}} = \mu_{\alpha, \text{shuffled}}$. This means that the observed “attraction” among opinions does not imply, as one would naively expect, that the empirical vectors $\{\vec{v}_i\}$ are closer to each other in cultural space than shuffled data.

However, real and shuffled data differ significantly in other properties of the distribution of vectors in cultural space. A first difference is that real distances are much more broadly distributed than shuffled ones. This can be inspected by measuring the intragroup variance

$$\sigma_{\alpha}^2 \equiv \langle d_{ij}^2 \rangle_{ij \in \alpha} - \langle d_{ij} \rangle_{ij \in \alpha}^2 = \sum_{k,l} \sigma_{\alpha}^{(k,l)} \quad [\text{S4}]$$

Note that for shuffled data we have

$$\sigma_{\alpha, \text{shuffled}}^2 = \sum_k \sigma_{\alpha, \text{shuffled}}^{(k,k)} = \sum_k \sigma_{\alpha, \text{real}}^{(k,k)} \quad [\text{S5}]$$

since $\sigma_{\alpha, \text{shuffled}}^{(k,k)} = \sigma_{\alpha, \text{real}}^{(k,k)}$ and $\sigma_{\alpha, \text{shuffled}}^{(k,l)} = 0$ for $k \neq l$, while for real data we have

$$\sigma_{\alpha, \text{real}}^2 = \sigma_{\alpha, \text{shuffled}}^2 + \sum_{k \neq l} \sigma_{\alpha, \text{real}}^{(k,l)} > \sigma_{\alpha, \text{shuffled}}^2 \quad [\text{S6}]$$

where the last inequality comes from the observed positivity of $\sum_{k \neq l} \sigma_{\alpha, \text{real}}^{(k,l)}$. As can be seen from Table S1, $\sigma_{\alpha, \text{real}}$ is roughly twice as large as, and more variable than, $\sigma_{\alpha, \text{shuffled}}$. Therefore, even if the distribution of intervector distances has the same average value in real and shuffled data, real distances are more broadly distributed than shuffled ones. A second difference is that, as shown in Fig. S2 and discussed in the main article, real vectors are approximately ultrametrically distributed in cultural space, while

shuffled and random ones are not. This means that the broader distribution of real distances (with respect to shuffled ones) is implied by the ultrametric structure, characterized on one hand by an increased frequency of both nearby vectors (representing individuals within the same branch of the dendrogram), and on the other hand by an increased frequency of distant ones (representing individuals belonging to different branches). By contrast, shuffled data generate vectors with the same average distance but more uniformly and nonultrametrically distributed in cultural space.

Link density and largest connected components. The ultrametricity of real intervector distances (documented in the main article) implies that if we “cut” the dendrogram of cultural vectors at some height ω we obtain a set of disconnected branches, within which individuals are separated by a cultural distance smaller than ω , and across which individuals are separated by a distance larger than ω . In other words, we obtain the connected components of the cultural graph defined by linking pairs of individuals separated by a distance lower than a certain threshold ω . The concept of bounded confidence implies that individuals belonging to different connected components of the cultural graph, even if linked by a social tie, cannot interact (neither directly nor indirectly). Therefore, in a fragmented cultural graph information (intended as mutual influence) can only diffuse locally. A necessary condition in order to have a global spread of information is that cultural channels form a giant connected component spanning (a finite fraction of) the N individuals in a given group. This implies that the fraction $s_{\alpha}(\omega)$ of vertices spanned by the largest connected component for a given value of ω represents the largest fraction of individuals in group α that can potentially influence each other, through either direct or indirect interactions, if the confidence threshold is set to ω . This lead us to choose $s_{\alpha}(\omega)$ as a global measure of potential influence, capturing how local interactions combine together at a large-scale level. It is also important to consider a purely local measure of influence, i.e., the average probability that any two individuals can influence each other through a direct interaction. To this end, we also considered the density $f_{\alpha}(\omega)$ of realized cultural channels, i.e., the fraction of pairs of individuals in group α closer than ω [which is nothing but the cumulative density function (CDF) of the distance distribution]. The resulting curves are shown in Fig. S3 A, B. For both quantities, we observe large differences between real and randomized data. In particular we find that, for a given value of α , real data are characterized by higher levels of local and global influence than shuffled and random data. In order to understand whether the differences among the curves in Figs. S2A and S2B can be simply traced back to overall differences in the average values (μ_{α}) and variances (σ_{α}^2) of the intervector distances, in Fig. S3 C, D we show f_{α} and s_{α} when plotted as a function of the standardized parameter $z \equiv (\omega - \mu_{\alpha})/\sigma_{\alpha}$. We find (see Fig. S3C) that all the $f_{\alpha}(z)$ plots collapse onto a single universal curve, which is indistinguishable from the cumulative density function (CDF) of the standard Gaussian distribution $f_0(z) \equiv \int_{-\infty}^z dx (e^{-x^2/2})/\sqrt{2\pi}$. In other words, in each group α the distances are normally distributed, with group-specific mean μ_{α} and variance σ_{α}^2 . This means that all the empirical differences in link density among groups are taken care of after rescaling the distance, so that the control parameter z completely specifies the density of realized cultural channels of any group. Thus, if cultural channels were placed uniformly among individuals as in a homogeneous random graph, one would also observe an analogous universal collapse of the $s_{\alpha}(z)$ curves and of any other topological property. By contrast, as can be seen in Fig. S3D, this approximately occurs only in the shuffled case, but not for real data. This result indicates again a nontrivial distribution of real cultural vectors, and singles out differences across the sampled groups that are not simply explained in terms of an overall variability. In particular,

the universality of the link density function observed in Fig. S3C does not imply a universal structure of connected components, due to correlations between pairs of edges generated by the correlations between distances. In other words, even after standardizing the local level of mutual influence, real data continue to differ significantly in their global level of influence. Therefore any process which depends on the cultural distance between individuals might have very different global outcomes even when taking place on locally identical structures.

It is important to note that, for ultrametric data, $f_\alpha(\omega)$ coincides with the fraction of pairs of cultural vectors within the same connected branches, when the dendrogram is cut at the height ω . By contrast, for randomized data both $f_\alpha(\omega)$ and $s_\alpha(\omega)$ are no longer in relation with the structure of the dendrogram cut at a given point in the vertical dimension, since the latter does not represent the original distances, due to the lack of ultrametricity. This important property has remarkable consequences for the outcome of dynamical models defined on real data, as discussed in detail in the main article and in the following sections of this SI file.

More Details on the Dynamical Models Considered. In this section we provide additional details about the two dynamical models we used in order to simulate short-term collective social behavior and long-term cultural convergence.

The modified Cont-Bouchaud model. In its simplest formulation, the Cont-Bouchaud (CB) model (20) considers a population of individuals (in the financial jargon, agents) that can make a binary choice between buying or selling an asset traded in the market. We can represent the choice expressed by the i th agent as $\phi_i = \pm 1$. Binary choices are the simplest possibility considered also in other models of social processes, such as voting dynamics (21). For our purposes, we interpret ϕ_i as a short-lived variable which represents the choice or action of agent i (e.g., whether i likes or dislikes an item). It is not a cultural variable (e.g., a unidimensional opinion), because we assume culture and opinions to evolve over a longer time scale (and this will be modeled using the modified Axelrod model). In the CB model, the effects of mutual influence are modeled by introducing a random graph through which agents can exchange information before making their choices. As a result of this information exchange, all the agents belonging to the same connected component are assumed to collectively agree on the choice to make. Therefore, if A labels a connected component of the graph, the choice of all agents belonging to A is the same ($\phi_i = \phi_A \forall i \in A$), while different connected components make statistically independent choices. The key result is that the probability distribution of the aggregate choice of all individuals (which in the CB model is the aggregate demand determining the price change of the asset) crucially depends on the topology of the interaction graph. In particular, if the connection probability p is set at the critical value $p_c \sim N^{-1}$ (for $N \rightarrow \infty$) of the phase transition giving rise to the giant connected component, the distribution of the sizes of connected components acquires a power-law form, which in turn implies a power-law distribution of price returns similar to the empirically observed ones. Other values of p yield different outcomes. In the limit $p = 0$ (empty graph) all agents make independent choices and the distribution becomes Gaussian. By contrast, for $p = 1$ (complete graph) all agents always make the same choice and the distribution is double-peaked.

In order to study the effects of the nontrivial distribution of individuals in cultural space, we extended the CB model to a more general “coordination model” which incorporates a dependence on real data. In particular, we introduce a more realistic mechanism allowing culturally similar agents to make similar choices. To take this aspect into account, we assume that each agent is described by a cultural vector $\{\vec{v}_i\}$ and that agents inter-

act, rather than on a random graph defined by a value of p , on the cultural graph defined by a value of the confidence ω . Again, agents within the same connected component are assumed to make collectively the same choice, while agents belonging to different components make statistically independent actions.

The overall outcome of the process (e.g., the result of the survey/referendum/election) is the sum of individual preferences, and can be quantified by the average choice

$$\Phi = \frac{1}{N} \sum_{i=1}^N \phi_i \quad [\text{S7}]$$

whose sign reflects the action of the majority. If the choices of all agents are independent ($\omega = 0$), Φ is the sum of N uncorrelated random variables with finite variance and, as follows from the Central Limit Theorem, normally distributed. In such a case, if the individual binary probabilities are equal (i.e., the events $\phi_i = +1$ and $\phi_i = -1$ are equiprobable), then the probability $P_\omega(\Phi)$ that the average choice takes the particular value Φ is symmetric about the most probable value $\Phi = 0$. If $\omega > 0$, Φ can be rewritten as the sum over different connected components:

$$\Phi = \frac{1}{N} \sum_A S_A \phi_A \quad [\text{S8}]$$

where now A labels the components, $\phi_A = \pm 1$ is the choice of all actors in component A , and S_A is the size of A . Now, a crucial point is that even if each connected component makes one of the two choices $\phi_A = \pm 1$ with equal probability, and hence the distribution $P_\omega(\Phi)$ is still symmetric about $\Phi = 0$, the symmetry breaks spontaneously at some critical threshold ω_c .

To see this, we computed $P_\omega(\Phi)$ in the following manner. For each group of individuals in our data, and for a given value of ω (from $\omega = 0$ to $\omega = 1$ in increments of 0.01), we identified the connected components, assigned each of them one of the two choices randomly, and computed the resulting value of Φ . We repeated this procedure 500,000 times on each sampled group and measured $P_\omega(\Phi)$ as the normalized histogram of the values obtained. As reported in the main article, we found a critical value ω_c marking the transition from a single-peaked to a double-peaked shape of the distribution $P_\omega(\Phi)$. The above analysis provides a method to determine with small indeterminacy ($\Delta\omega = 0.01$) the threshold value ω_c for each particular, finite group under study. We repeated our analysis on each of the 13 sampled groups (real and randomized) and obtained the corresponding critical thresholds $\omega_c \pm \Delta\omega$ (see Fig. S4). Note that smaller (larger) values of ω_c require smaller (larger) levels of influence between individuals in order to trigger collective behavior. Therefore ω_c represents a novel measure of the resistance of a social group to act collectively. Importantly, we found that the thresholds for shuffled data are always larger than those for real data, and that the ones for random data are even larger. This shows that empirical interopinion correlations, which are responsible for the ultrametric distribution of individuals in cultural space, strongly facilitate collective social behavior by systematically lowering the resistance to coordination. While for shuffled and random data all thresholds are equal within errors, the entity of the enhancement of collective behavior in real data varies significantly across the sampled groups and determines nonuniversal values of the thresholds. This implies that even two randomly sampled social groups (for instance one in Italy and one in Portugal) with the same size and under the same level of mutual influence may evolve to opposite collective states (coordination or heterogeneity) if the critical thresholds for the two groups differ. In general, ω_c also represents a fundamental threshold for any dynamical process dependent on cultural data. Any mechanism taking place within a cultural distance smaller than ω_c will not

propagate through the whole network, while if the interaction range is larger than ω_c the information can percolate the entire system.

We now prove rigorously the equality

$$C(\omega) \equiv \sigma_\omega(\Phi) = \sqrt{\sum_A \left(\frac{S_A}{N}\right)_\omega^2} \quad [\text{S9}]$$

which establishes a tight relation between network topology and the level of collective social behavior in the model. If, for a given value of ω , we denote the expected value of Φ as $\langle \Phi \rangle_\omega = \sum_\Phi P_\omega(\Phi) \Phi$ and its second moment as $\langle \Phi^2 \rangle_\omega = \sum_\Phi P_\omega(\Phi) \Phi^2$, the variance $\sigma_\omega^2(\Phi)$ is defined as

$$\sigma_\omega^2(\Phi) \equiv \langle \Phi^2 \rangle_\omega - \langle \Phi \rangle_\omega^2. \quad [\text{S10}]$$

For simplicity, in what follows we drop the dependence of all quantities on ω . For a fixed value of ω , the sizes of the connected components of the network are given by $\{S_A\}$, and determine the aggregate choice Φ through Eq. S8. Since Φ is a sum of the random variables $\{S_A \phi_A / N\}$, its variance $\sigma^2(\Phi)$ can be easily expressed as

$$\sigma^2(\Phi) = \sum_A \sum_B \frac{S_A S_B}{N^2} \sigma_{AB} = \sum_A \frac{S_A^2 \sigma_A^2}{N^2} + \sum_A \sum_{B \neq A} \frac{S_A S_B}{N^2} \sigma_{AB} \quad [\text{S11}]$$

where σ_{AB} denotes the covariance between the choices ϕ_A and ϕ_B of two different connected components A and B , and σ_A^2 is the variance of ϕ_A . Now, since

$$\sigma_A^2 \equiv \langle \phi_A^2 \rangle - \langle \phi_A \rangle^2 = 1 \quad [\text{S12}]$$

and since different connected components make statistically independent choices, it follows that

$$\sigma_{AB} \equiv \langle \phi_A \phi_B \rangle - \langle \phi_A \rangle \langle \phi_B \rangle = \delta_{AB} \sigma_A^2 = \delta_{AB} \quad [\text{S13}]$$

where $\delta_{AB} = 1$ if $A = B$ and $\delta_{AB} = 0$ if $A \neq B$. Thus the variance of Φ is simply

$$\sigma^2(\Phi) = \sum_A \frac{S_A^2}{N^2} \quad [\text{S14}]$$

which proves the last equality in Eq. S9.

Note that, technically, a critical threshold is defined only for infinite systems, and most of the methods available to measure its exact value in other models make use of an ideal thermodynamic limit. However, here we are interested in determining the parameter value that drastically changes the behavior of an intrinsically finite system. An extrapolation to an infinite number of individuals would make any reference to the outcome of real social processes, and to empirical data, vanish. By contrast, the method we use has an easily interpretable meaning even in the finite case. Moreover, as we now show, in the infinite size limit it would yield the exact value of the critical threshold. Note that for $\omega = 0$ there are N connected components of size $S_A = 1$ (all vertices are isolated) and therefore $\sigma_0^2(\Phi) = 1/N$ (the results of the Central Limit Theorem are recovered). In the opposite limit $\omega = 1$, the network is a single connected component of size $S_A = N$, which yields $\sigma_1^2(\Phi) = 1$. Thus the social coordination $C(\omega) \equiv \sigma_\omega(\Phi)$ varies from $C(0) = 1/\sqrt{N} \rightarrow 0$ (no collective behavior) to $C(1) = 1$ (perfect collective behavior). For generic values of ω note that, since $\sum_A S_A = N$, the expression for $\sigma^2(\Phi)$ in Eq. S14 has the form of an inverse participation ratio. This means

that $\sigma^2(\Phi) \simeq 1/n$ if the sum over A is dominated by n terms of approximately equal size. In particular, $\sigma^2(\Phi) \simeq 1$ if there is one dominant connected component, while $\sigma^2(\Phi) \simeq 1/N$ if each connected component trivially contains only one vertex. This result rephrases the connection between the shape of $P(\Phi)$ and the underlying network topology: when there is no giant component, the width of $P(\Phi)$ is $\sigma(\Phi) \simeq 1/\sqrt{N} \rightarrow 0$, while when the giant component is there the width of $P(\Phi)$ has the finite value $\sigma(\Phi) \simeq 1$. Thus when there is no giant component there must be a single peak, while the presence of two peaks at finite distance necessarily implies the presence of the giant component. In such a case, $\sigma(\Phi)$ gives an estimate of the separation between the peaks. Importantly, these results remain valid as N goes to infinity, therefore our method to compute ω_c as the value marking a spontaneous symmetry breaking (from single-peaked to double-peaked) in the probability $P_\omega(\Phi)$ provides a consistent way to define a ‘‘critical’’ value, which technically is defined only for infinite systems, even for our inherently finite data. For infinite systems, our method would yield the correct value of the critical threshold. Other finite-size techniques would require assumptions about how topological quantities scale with network size. While for network models [such as the Erdős-Rényi random graph (22)] it is possible to derive these assumptions, for real systems this is not possible.

The modified Axelrod model. In the original version of the Axelrod model (19), N individuals (in social science jargon, ‘‘actors’’) sitting at the vertices of a social network are represented as vectors of cultural traits (or features), that evolve through discrete steps. In an elementary time-step, an individual i and one of his neighbors (say j) are selected. Then the normalized overlap $o_{ij} \in [0, 1]$ between their cultural vectors is computed as the fraction of identical components (note that the overlap is related to the cultural distance d_{ij} through $o_{ij} = 1 - d_{ij}$). With probability equal to o_{ij} , the two actors interact: one of j ’s traits, chosen randomly among the set of traits where i and j differ, is changed and set equal to the corresponding trait of i . Otherwise nothing happens, and two other actors are selected. These rules implement the two basic mechanisms of social influence (neighboring actors tend to converge culturally) and homophily (similar individuals interact more frequently). The Axelrod model leads to the important conclusion that these two mechanisms do not necessarily reinforce each other leading to a culturally homogeneous society. In fact, the model predicts that diversity is preserved: when two individuals become completely different (zero overlap), they no longer interact. Thus in the allowed final configurations, two neighboring actors are either completely identical or completely different, and the society is split into *cultural domains* of identical vectors, with no overlap between adjacent domains. The average $\langle N_D \rangle$ (over many realizations) of the number N_D of different domains in the final stage, or equivalently the fraction $\langle N_D \rangle / N$, is a convenient way to measure the predicted cultural diversity as a function of the model parameters.

As for the CB model, for our purposes it is important to incorporate real data into the Axelrod model. In this case, convenient generalizations have already been proposed. While in the original model traits were nonmetric, a modification bringing us closer to real data is the introduction of metric features and the consequent redefinition of cultural distance d_{ij} as a metric distance between cultural vectors (6). Another important variant introduces the effect of bounded confidence: Flache and Macy (5) introduced a threshold θ such that, if the overlap is smaller than or equal to θ , no interaction takes place. Otherwise, it takes place with probability o_{ij} (the original version of the model is recovered if $\theta = 0$). Clearly, θ has exactly the same meaning as $1 - \omega$, where ω is the confidence we introduced above. The threshold compensates the effect that, as the number

of features grows, the presence of completely different pairs of agents becomes unlikely. It also allows us to reproduce more realistically the fact that individuals are uninfluenced by each other if they are different enough, not necessarily on each and every opinion they have[§].

The above modifications allowed us to use the empirical cultural vectors (rather than commonly assumed uniformly random vectors) as the starting configuration, and study the final diversity predicted by the model. In order to simulate long-range on-line-like dynamics, we assumed that the social network is a complete graph. However, as discussed in the last section of this SI file, many more complicated choices either reduce or map to this simple case. Since we never found more than one individual with exactly the same cultural vector, the initial number of cultural domains was always $N = 500$ for each sampled group. Using the same definition of cultural distance as above, we simulated the modified Axelrod dynamics on the real cultural vectors as a function of the threshold θ . In Fig. S5A we report the average number $\langle N_D \rangle_\omega$ of different cultural domains in the final state of the dynamics, as a function of $\omega = 1 - \theta$ for real, shuffled and random opinions. As can be seen, for large values of θ (small ω) the final fraction of culturally homogeneous domains is finite (at the extreme $\theta = 1$ there is no evolution from the starting configuration and the initial vectors remain all distinct), while for small values of θ (large ω) the same fraction is of order $1/N$ (and vanishes at the extreme $\theta = 0$ corresponding to the ordinary Axelrod model). This means that the final cultural diversity decreases as ω increases. With respect to real data, the curves for shuffled and random data are moved rightward. Naively, if combined with our previous results, this finding appears to confirm the expectation that larger cultural diversity is only reached in a regime (small ω) where short-term collective behavior is weak or absent, and conversely strong collective behavior can only exist for large values of ω which suppress cultural heterogeneity in the long run. Moreover, this appears to apply equally to real, shuffled, and random data.

However, this conclusion is incorrect. In Fig. S5B we show the average number $\langle N_D \rangle$ of final different cultural domains versus the number N_C of initial connected components in the underlying cultural network, both obtained for various values of the threshold ω and for the three usual cases of real, shuffled, and random data. We find that, for a given value of N_C , real data are those that achieve the largest level of long-term cultural heterogeneity (value of $\langle N_D \rangle$). Indeed, for real data the realized value of $\langle N_D \rangle$ is the largest possible[¶] ($\langle N_D \rangle \approx N_C$) indicating that cultural convergence is confined within the initial connected components, each of which eventually becomes a single cultural domain. By contrast, in randomized data there are less final cultural domains than initial connected components, indicating that the latter often “merge” into larger cultural domains.

More Complicated Scenarios. When using real data as seeds for the models, we assumed that each individual can interact with everyone else, if sufficiently similar culturally. As we mentioned, this corresponds to the assumption that the social network (combining on-line and offline interactions) connecting the N individuals

of a given group α is a complete graph. Also, we assumed that direct actor-actor interactions are the only possible source of changes in opinions, and that they obey specific mathematical rules dictated by the dynamical model considered. However, in both respects our results are actually more general, since it turns out that many other model assumptions would lead to similar outcomes. To show this, we conclude with a discussion of several more complicated scenarios that one could introduce, including models with different interaction probabilities (5, 6, 21), different network topologies (4, 19, 21–23), coevolution of networks and opinions (17, 24–27), cultural drift (28, 29), higher-order interactions (5, 21), and external sources of information (30). Despite their sophistication, in many cases these models can be easily related to our simple approach. The following discussion shows that the observed approximate ultrametricity of cultural data has the surprising effect of constraining the dynamics of even quite complicated models so dramatically that their outcome becomes essentially equivalent, or directly mapped, to what we obtained using simple dyadic interactions on a complete graph.

Different interaction probabilities. The rules of the modified Axelrod model described above, when defined on a complete graph, can be compactly rephrased in terms of the following probability that at a given time step two actors i and j interact:

$$p_{ij} = o_{ij}\Theta(o_{ij} - \theta) \tag{S15}$$

where $o_{ij} = 1 - d_{ij}$ is the normalized distance-based overlap between the cultural vectors \vec{v}_i and \vec{v}_j , and $\Theta(x)$ is the Heavyside step function defined as $\Theta(x) = 0$ if $x \leq 0$ and $\Theta(x) = 1$ if $x > 0$. Now, let us still make the complete-graph assumption, but let us wonder to what extent the results we obtain using the modified Axelrod model are dependent on the above specific form of the probability p_{ij} (5, 6, 21). Indeed, any modified probability \tilde{p}_{ij} of the form

$$\tilde{p}_{ij} = f(o_{ij})\Theta(o_{ij} - \theta) \tag{S16}$$

[where $f(o_{ij})$ is an increasing function of o_{ij}] would equally implement the bounded confidence hypothesis ($\tilde{p}_{ij} = 0$ if o_{ij} is smaller than a threshold) and the homophily mechanism (above threshold, a larger o_{ij} implies a larger \tilde{p}_{ij}). The effect of changing the form of $f(o_{ij})$ on the results is a different relative rate of interaction between actors with $o_{ij} > \theta$ (even if pairs with larger/smaller o_{ij} will still interact at a higher/lower rate, and the ones below threshold will still not interact). For a generic distribution of vectors in cultural space, this may change the shape and composition of the cultural clusters that gradually coalesce in the system, and in the end this may have an uncontrolled effect on the final average number of cultural domains $\langle N_D \rangle_\omega$ obtained for a given value of $\omega = 1 - \theta$. However, if the distribution of cultural vectors is ultrametric as we approximately find empirically, then the different rate of interaction will only affect the pairs of actors found within the same branch of the dendrogram, if the latter is cut at the “height” $1 - \theta$. Now, remember that for real data we found that the average final number of cultural domains is $\langle N_D \rangle \approx N_C$, i.e., different such branches collapse to the same cultural state, but do not influence each other. If the same initial ultrametric configuration is considered, a modification of the relative rates of interaction may change the temporal order with which actors within a given branch collapse to the same final state, but will still not mix different branches. Therefore, since the initial number N_C of such branches is independent of $f(o_{ij})$ and only depends on θ , the same is true for the final number $\langle N_D \rangle$ of cultural domains. This shows that, if the initial cultural specification is ultra-

[§]We also note that it can compensate for the arbitrariness of the number of features appearing in the cultural vectors: two identical cultural vectors may result actually different if additional features are considered, making the concept of zero overlap not well defined. This is particularly important when dealing with real-world opinions such as our questionnaire data: adding just one more question to the questionnaire cannot change the nature of cultural evolution.

[¶]As we mentioned, in real data ultrametricity implies that the initial connected components are entire branches of the dendrogram, and are therefore completely connected cliques. Since in a clique everyone can interact with everyone else, all individuals in the same component will eventually converge to the same cultural vector. Also note that it is extremely unlikely, for high-dimensional vectors such as the ones in our analysis ($F = 161$), that two distinct connected components will end up with the same cultural vector by chance. Thus, for real data, the maximum value of $\langle N_D \rangle$ is N_C .

metric, the outcome of the modified Axelrod model is surprisingly robust to changes in the definition of p_{ij} .

Social networks with given topology. We now consider the effect of specifying, instead of a complete graph, a given social network of interaction (4, 19, 21–23) with adjacency matrix elements s_{ij} (consistently with our notation in the first section of this SI file). With this choice, the probability of interaction in the modified Axelrod model becomes

$$p_{ij} = s_{ij} o_{ij} \Theta(o_{ij} - \theta) \quad [\text{S17}]$$

(since only connected actors, for which $s_{ij} = 1$, can interact). In general, a specific choice of the social network can change the results significantly. However, in our case we have no indication about the structure of the network (recall that the individuals are carefully chosen in such a way that they do not know each other, and our aim is to study what happens if they start interacting either on-line or offline). Therefore we must consider networks generated according to some model. This leads us to two possibilities. As a first choice, we could assume that the network is formed according to some stochastic rule which is completely independent of the cultural vectors [e.g. a regular lattice, a random graph, a small-world model, a scale-free network, etc. (22)]. As a second choice, we could assume that the network is formed with a dependence on the cultural vectors. In particular, we could exploit the well documented result that neighbors in a social network tend to be culturally similar: homophily is found not only to *strengthen* the existing social contacts (giving higher probability of social influence to more culturally similar actors as in the Axelrod model), but also to give a higher probability of *formation* of social contacts.

In both cases, the correct way to proceed would be to average the results over several realizations of the same network model. This means that there are now two averages to perform, one with respect to the opinion dynamics of the Axelrod model (as we did so far on the complete graph) and one with respect to the network formation rule. There are two possible ways, which in principle can lead to different results, to combine these averages. One is by considering network formation as what is technically called a *quenched disorder*: for a given network realization, the average over the dynamics is performed; then a different network realization and the corresponding dynamical average are obtained, and so on until the dynamical averages for all network realizations are performed, and finally averaged at the end. Another way is by considering network formation as an *annealed disorder*, i.e., by performing the average over network realizations first. The latter choice is simpler to characterize mathematically, because it removes the correlations induced by a fixed network structure. It represents a mean-field approximation to the results obtained in the quenched case. As we now show, many of the differences (with respect to our complete-graph choice) that would arise for a single realization of a network model are washed away after averaging (in the annealed sense) over network realizations.

In the annealed case, the probability p_{ij} appearing in Eq. S17 is effectively replaced by the quantity $\tilde{p}_{ij} = \langle p_{ij} \rangle$ where $\langle \cdot \rangle$ denotes an ensemble average over the possible realizations of the network. The ensemble average has effect only on s_{ij} and leads us to the probability

$$\tilde{p}_{ij} = \langle p_{ij} \rangle = \langle s_{ij} \rangle o_{ij} \Theta(o_{ij} - \theta) \quad [\text{S18}]$$

where $\langle s_{ij} \rangle$ is the probability that, under the network model considered, the vertices i and j are connected. If we assume that the network model has no dependence on the cultural vector (the first of the two options discussed above), $\langle s_{ij} \rangle$ is nothing but the average link density s produced by the network model con-

sidered, and is therefore a constant (uniquely determined by the choice of parameters in the network model) independent of i and j . For instance, for the Erdős-Rényi random graph model, $s = \langle s_{ij} \rangle$ is simply the probability that any two vertices are connected. Therefore Eq. S18 becomes equivalent to Eq. S16, with

$$f(o_{ij}) = s \cdot o_{ij} \quad [\text{S19}]$$

which is an increasing function of o_{ij} . Therefore we can replicate our arguments above to conclude that, for ultrametric data, our results would be unaffected by this choice. If we consider the second option, i.e., a network model that depends on the cultural vectors and gives higher connection probability to pairs of actors with larger overlap, then $\langle s_{ij} \rangle$ is exactly such probability and is therefore an increasing function $g(o_{ij})$ of o_{ij} . This implies that, again, Eq. S18 becomes equivalent to Eq. S16 where now

$$f(o_{ij}) = g(o_{ij}) o_{ij} \quad [\text{S20}]$$

which is still an increasing function of o_{ij} . Therefore the above arguments hold in this case as well, and we can conclude that in the annealed case the specification of different network topologies will leave the complete-graph results unaffected. With quenched disorder, we may obtain different results which would require a more detailed investigation, but this goes beyond the scope of the present work. Simulations of a different model of unidimensional opinion dynamics with bounded confidence (4) have revealed a high degree of robustness for various quenched network topologies, as we found here in the annealed case.

Networks coevolving with opinions. One can regard annealed disorder as a scenario where the social network evolves much faster than cultural dynamics (so that the average over different networks is carried out first), and quenched disorder as the opposite case where the cultural process is much faster than the evolution of the social network (and the average over cultural dynamics is carried out first). A very interesting intermediate case, which has been suggested to be more realistic than both extremes (17, 24–27), is when cultural dynamics and network evolution take place at a comparable rate. This case leads to a mechanism of coevolution between cultural traits and social network structure (17, 24, 25, 27). According to such mechanism, social ties between actors who become culturally dissimilar are constantly removed and randomly replaced, so that a feedback between dynamics and topology (26) is introduced. Usually, this is implemented by assuming that actors redirect their ties—from former neighbors that are eventually found at an exceedingly long cultural distance, to new neighbors that are selected randomly among all actors. This process implies that, apart from the total number of ties, the network progressively loses memory of its initial condition (e.g., a lattice or other model specification as above) and becomes entirely determined by cultural traits (17) [or any other variable responsible for the creation of links (26)]. This leads exactly to our results on a complete graph, the only quantitative differences being the density of links (which in our case is not constrained by the initial topology) and the overlap threshold θ (which in our case spans the full range $0 \leq \theta \leq 1$, while in some models is fixed to zero as in the original Axelrod model). The second difference simply implies that other parameter choices are automatically included in our model as particular cases. The first difference implies that, if the coevolution process starts from a network with link density $s < 1$ (instead of $s = 1$ as in our case), then Eq. S15 must be effectively (in a sense similar to the annealed-disorder case) replaced by Eq. S16 with $f(o_{ij})$ given by Eq. S19. Again, this means that our results are robust to this modification. So we find that, after controlling for the overall link density, our choice of a complete graph for the

social network can also be viewed as a particular asymptotic (i.e., after the memory of the initial state is lost) implementation of coevolving models with rewiring of ties.

Cultural drift. It is also interesting to discuss the stability of our results after the introduction of *cultural drift*, which is usually modeled as a spontaneous and random change in the cultural traits of individuals, taking place continuously with a given noise rate (28, 29). While cultural drift usually has counterintuitive and destabilizing consequences (17, 28, 29), we expect that in this case ultrametricity will severely suppress such effects. To see this, we note that it is unreasonable to assume that, on the same time scale over which influence and homophily take place, cultural drift can move the opinion vectors by more than the confidence threshold ω . This would imply that actors continuously and spontaneously perform a “schizophrenic” change to a previously unaccepted opinion. Therefore the only reasonable range of noise rate must be such that, between two interactions among actors, the average distance by which cultural vectors have drifted is much smaller than ω . But such process would affect the dendrogram only below the height ω , and we therefore predict no substantial effects of cultural drift, as in the other cases we have discussed. By protecting the dynamics from noise, the introduction of the threshold $\theta = 1 - \omega$ appears again to reduce the effects of additional model parameters and specifications, and to ensure that our results are directly interpretable in terms of the empirical properties of cultural data.

Higher-order interactions. We now discuss qualitatively the effect of adding higher-order (nondyadic) interactions (5, 21) to the actor-actor cultural dynamics. Such interactions require that groups of three, four, or more individuals interact simultaneously and collectively converge culturally. The dynamics of the Axelrod model, if extended to triadic interactions, would prescribe that groups of three actors (say i, j, k) interact with a probability p_{ijk} which is some function $f(\vec{v}_i, \vec{v}_j, \vec{v}_k)$ of the vectors $\vec{v}_i, \vec{v}_j, \vec{v}_k$. After this interaction, the three actors will have a cultural feature in common, i.e., $v_i^{(l)} = v_j^{(l)} = v_k^{(l)}$, for some component l where their vectors previously differed. Adding the bounded confidence mechanism to this dynamics implies that the interaction is possible only if the three vectors are closer to one another than a certain threshold. More specifically, if $\theta_2 \equiv \theta$ denotes the usual (overlap) threshold for dyadic interactions, triadic interactions introduce an additional threshold θ_3 such that they take place with probability

$$p_{ijk} = f(\vec{v}_i, \vec{v}_j, \vec{v}_k) \Theta(o_{ij} - \theta_3) \Theta(o_{jk} - \theta_3) \Theta(o_{ki} - \theta_3). \quad [\text{S21}]$$

Similarly, fourth-order interactions would involve a threshold θ_4 which only allows convergence if all pairs within a group of four actors mutually overlap by more than θ_4 , and so on for fifth- and higher-order interactions.

In principle, all such thresholds $\theta_3, \theta_4, \dots$ represent additional free parameters that might expand the phase space into regions with very different outcomes. However, to be realistic, higher-order interactions must be *individually* weaker than lower-order ones. To see this, consider again the case of triadic interactions. Their introduction adds an *extra* triadic probability of interaction p_{ijk} on top of the combination of three dyadic interactions with probabilities p_{ij}, p_{jk}, p_{ki} . As the order of the interaction increases, the extra probabilities should be smaller and smaller, i.e. they should be corrections to the lower-order terms, otherwise one would be conceiving a system where it is increasingly more probable to engage in larger and larger (culturally influential) group interactions. Similarly, the confidence threshold θ_k for k th order interactions should not be such that individuals become paradoxically more and more open to an opinion change simply because the size of the group they are temporarily interacting with in-

creases. This means that one should choose the values of the overlap thresholds in such a way that

$$\theta_2 \geq \theta_3 \geq \theta_4 \geq \dots \quad [\text{S22}]$$

In terms of the distance thresholds $\omega_k = 1 - \theta_k$, this reads

$$\omega_2 \leq \omega_3 \leq \omega_4 \leq \dots \quad [\text{S23}]$$

If we interpret bounded confidence as an intrinsic tendency of individuals to be open to influences only within an absolute cultural distance, irrespective of the order of the interactions where it is involved, then we could set $\omega_2 = \omega_3 = \omega_4 = \dots$ in the above equation. However, since this is quite open to different interpretations, in what follows we will not make this additional assumption and we will consider the more general case. Note that for ultrametric cultural vectors Eq. S23 implies that, if ω_2 is our usual height at which we cut the dendrogram into disconnected branches, the introduction of the entire hierarchy of higher-order interactions will only affect cultural convergence within such branches, again with no mixing across different branches. In more detail, let us denote as ω -branch a branch produced by cutting the dendrogram at the height ω . An argument similar to the one we have used in the case of dyadic interactions implies that, if only k th order interactions were present and if the vectors were ultrametric, then the number of final cultural domains would be essentially the initial number of ω_k -branches, since each of the latter would eventually merge into a single cultural vector. Now, Eq. S23 implies that, if $k > h$, then the ω_k -branches (affected by higher-order interactions) are nested within the ω_h -branches (affected by lower-order interactions). This shows that even the simultaneous introduction of all higher-order interactions can only affect the order and speed of convergence of subbranches nested within the ω_2 -branches. Surprisingly, this would leave our original result, obtained with only dyadic interactions, unaffected. As all other mechanisms we have considered so far, this remarkable stability of the outcome of cultural dynamics is entirely due to the ultrametricity of real data.

External sources of information. We conclude our list of more sophisticated scenarios with a discussion of the effects of an exposure of individuals to external sources of information (such as media, newspapers, etc.) (30). Such effects are more difficult to control, due to a greater freedom in the definition of models capturing the relevant mechanism. We can however tremendously simplify the discussion by focusing on a particular case which we expect to be the relevant one for the present case study. In the subsection “The Eurobarometer dataset” of this SI file we argued that the main external sources of information are in our case mass media equally accessible by the entire population, rather than by specialized niches. This leads us to the expectation that any further convergence, on top of actor-actor interactions, of the cultural vectors can be to a large extent approximated by a sort of background convergence reducing all cultural distances gradually and homogeneously. We can model this additional process as follows. Imagine that we “switch off” (dyadic or higher-order) actor-actor interactions, and let the vectors gradually converge under the only effect of external information. In our assumptions, this occurs homogeneously in cultural space, so that the net approximate effect is a temporal reduction of all distances d_{ij} : if $d_{ij}(0)$ denotes the initial distance between vectors \vec{v}_i and \vec{v}_j , after a time t all distances will be

$$d_{ij}(t) \approx c(t) d_{ij}(0) \quad \forall i, j \quad [\text{S24}]$$

where the positive function $c(t)$, which embodies the temporal dynamics, must decrease as t increases, and be such that

$c(0) = 1$. Therefore $0 \leq c(t) \leq 1$. We do not need any further assumption on the form of $c(t)$, since what is important is that it produces a background convergence that can be easily controlled for as follows. If we “switch on” actor-actor interactions, then the background convergence has the effect, at any time t , of replacing $o_{ij}(t) = 1 - d_{ij}(t)$ with $\tilde{o}_{ij}(t) = 1 - c(t)d_{ij}(t) = 1 - c(t) + c(t)o_{ij}(t)$ into Eq. S15. Dropping the time-dependence in the notation, the effective probability of interaction would become

$$\begin{aligned} \tilde{p}_{ij} &= \tilde{o}_{ij}\Theta(\tilde{o}_{ij} - \theta) = (1 - c + co_{ij})\Theta(1 - c + co_{ij} - \theta) \\ &= f(o_{ij})\Theta(o_{ij} - \tilde{\theta}) \end{aligned} \quad [\text{S25}]$$

where

$$f(o_{ij}) = 1 - c + co_{ij} \quad [\text{S26}]$$

is an increasing function of o_{ij} , and

$$\tilde{\theta} = \frac{\theta + c - 1}{c} \quad [\text{S27}]$$

is a transformed threshold. Equation S25 is equivalent to Eq. S16, apart from the presence of $\tilde{\theta}$ instead of θ . This simply means that

the results one would recover in this case are equivalent to the results we obtained with a different value of the threshold. In particular, the quantities we obtain for a given value ω of the overlap threshold would be replaced by the quantities we obtain for the transformed threshold

$$\tilde{\omega} = 1 - \tilde{\theta} = \frac{\omega}{c} \quad [\text{S28}]$$

as simple intuition would suggest. Therefore, the addition of external information can, under our assumptions motivated by this specific case study, be simply reabsorbed in a redefinition of the only parameter of the modified Axelrod model.

In all the extensions we have considered in this concluding section, it appears that the introduction of bounded confidence in terms of a cultural threshold, besides being important in itself for the theoretical reasons we discussed, makes the surprising effects of ultrametricity particularly evident. In particular, it has the advantage of freeing the results from otherwise arbitrary specifications (and additional free parameters) of the dynamical model. This allows us to obtain clearly interpretable outcomes in terms of the single parameter θ , even when considering quite complicated settings.

- INRA Brussels (Europe) and Report International for the European Commission, Directorate-General XII, Science Research and Development, *Europeans, Science and Technology—Public Understanding and Attitudes* (1993) Available at http://ec.europa.eu/public_opinion/archives/ebs/ebs_076_en.pdf.
- Gaskell G et al. (2000) Biotechnology and the European public. *Nat Biotechnol* 18:935–938.
- Gaskell G et al. (2011) The 2010 Eurobarometer on the life sciences. *Nat Biotechnol* 29:113–114.
- Weisbuch G (2004) Bounded confidence and social networks. *Eur Phys J B* 38:339–343.
- Flache A, Macy M (2007) Local convergence and global diversity: The robustness of cultural homophily. Arxiv preprint physics/0701333.
- De Sanctis L, Galla T (2009) Effects of noise and confidence thresholds in nominal and metric Axelrod dynamics of social influence. *Phys Rev E* 79:046108.
- Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge, U.K., Cambridge University Press).
- Onnela J-P et al. (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA* 104:7332–7336.
- Ebel H, Mielsch L, Bornholdt S (2002) Scale-free topology of e-mail networks. *Phys Rev E* 66: 35103.
- Blossfeld H P, Klijzing E, Mills M, Kurz K (2005) *Globalization, Uncertainty and Youth in Society* (New York, Routledge).
- Brady D, Beckfield J, Zhao W (2007) The consequences of economic globalization for affluent democracies. *Annu Rev Sociol* 33:15.1–15.22.
- Backstrom L, Huttenlocher D, Lan X, Kleinberg J (2006) Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery in Data* 44–54 (New York, ACM Press).
- Onnela J-P, Reed-Tsochias F (2010) Spontaneous emergence of social influence in online systems. *Proc Natl Acad Sci USA* 107:18375–18380.
- Ratkiewicz J, Fortunato S, Flammini A, Menczer F, Vespignani A (2010) Characterizing and modeling the dynamics of online popularity. *Phys Rev Lett* 105:158701.
- Bohannon J (2006) Tracking people’s electronic footprints. *Science* 314:914–916.
- Lazer D et al. (2009) Life in the network: The coming age of computational social science. *Science* 323:721–723.
- Centola D, González-Avella J C, Eguiluz V M, San Miguel M (2007) Homophily, cultural drift and co-evolution of cultural groups. *J Conflict Resolut* 51:905–929.
- Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78:1360–80.
- Axelrod R (1997) The dissemination of culture. *J Conflict Resolut* 41:203–226.
- Cont R, Bouchaud J-P (2000) Herd behavior and aggregate fluctuations in financial markets. *Macroecon Dyn* 4:170–196.
- Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81:591–646.
- Caldarelli G (2007) *Scale-Free Networks: Complex Webs in Nature and Technology* (New York, Oxford University Press).
- Guerra B, Poncela J, Gómez-Gardenes J, Latora V, Moreno Y (2010) Dynamical organization towards consensus in the Axelrod model on complex networks. *Phys Rev E* 81:056105.
- Zanette D H, Gil S (2006) Opinion spreading and agent segregation on evolving networks. *Physica D* 224:156–165.
- Santos F C, Pacheco J M, Lenaerts T (2006) Cooperation prevails when individuals adjust their social ties. *PLoS Comput Biol* 2:e140.
- Garlaschelli D, Capocci A, Caldarelli G (2007) Self-organized network evolution coupled to extremal dynamics. *Nat Phys* 3:813–817.
- Kozma B, Barrat A (2008) Consensus formation on adaptive networks. *Phys Rev E* 77:016102.
- Klemm K, Eguiluz V M, Toral R, San Miguel M (2003) Global culture: A noise-induced transition in finite systems. *Phys Rev E* 67:045101(R).
- Klemm K, Eguiluz V M, Toral R, San Miguel M (2005) Globalization, polarization and cultural drift. *J Econ Dyn Control* 29:321–334.
- González-Avella J C, Cosenza M G, Tucci K (2005) Nonequilibrium transition induced by mass media in a model for social influence. *Phys Rev E* 72:065102(R).

