LETTER

Experimental evidence for the influence of group size on cultural complexity

Maxime Derex[1], Marie-Pauline Beugin[1], Bernard Godelle[1] & Michel Raymond[1,2]

The remarkable ecological and demographic success of humanity is largely attributed to our capacity for cumulative culture[1–3]. The accumulation of beneficial cultural innovations across generations is puzzling because transmission events are generally imperfect, although there is large variance in fidelity. Events of perfect cultural transmission and innovations should be more frequent in a large population[4]. As a consequence, a large population size may be a prerequisite for the evolution of cultural complexity[4,5], although anthropological studies have produced mixed results[6–9] and empirical evidence is lacking[10]. Here we use a dual-task computer game to show that cultural evolution strongly depends on population size, as players in larger groups maintained higher cultural complexity. We found that when group size increases, cultural knowledge is less deteriorated, improvements to existing cultural traits are more frequent, and cultural trait diversity is maintained more often. Our results demonstrate how changes in group size can generate both adaptive cultural evolution and maladaptive losses of culturally acquired skills. As humans live in habitats for which they are ill-suited without specific cultural adaptations[11,12], it suggests that, in our evolutionary past, group-size reduction may have exposed human societies to significant risks, including societal collapse[13].

The accumulation of socially learned information over many generations has enabled humans to develop powerful technologies that no individual could have invented alone[14]. Cumulative culture is most likely to be restricted to the *Homo* genus and remains an evolutionary puzzle[15]. Several hypotheses have been proposed to explain this explosion in cultural complexity, with a recent emphasis on social-learning mechanisms specific to humans, such as teaching, language or imitation[16,17]. These mechanisms of faithful transmission stabilize cultural knowledge, thus enabling successive improvements, as has been previously shown theoretically[18] and empirically[19,20]. However, perfect transmission is most probably unrealistic, as for any given transmission event, an information loss is expected, particularly for complex tasks[4,21]. Moreover, transmission is only one aspect of the problem, as cumulative cultural evolution also requires the creation of new knowledge; that is, innovation.

The determinants of technological regression—the opposite situation—have been studied in Tasmanian aboriginals. It was argued that cultural losses were associated with population-size reduction[22]. A general model of cultural evolution that links demographic factors to psychological aspects of social learning has been proposed by Henrich[4]. Considering that transmission events for complex tasks are generally imperfect, with a large variance in fidelity, a learner could acquire by chance greater skill than the demonstrator if the number of transmission events (that is, the population size) is sufficiently large. As there is a psychological propensity to imitate successful individuals (prestige bias), this individual becomes the new demonstrator, driving cultural evolution. A decrease in population size makes such events unlikely, making cultural regression unavoidable. Analytical modelling shows that, as the population size increases, the combination of imperfect learning and prestige bias can lead to cumulative evolution, even if transmission is generally inaccurate. Bursts of cultural complexity during

the Palaeolithic era (2.6 million years ago to 10 thousand years ago) and particularly during the Upper Palaeolithic transition (45 thousand years ago) may illustrate demographic processes, rather than changes in cognitive abilities[5,23]. However, factors favouring the ability to develop complex culture will most probably also have a positive effect on population size, thus limiting causal assessments using correlative studies. Furthermore, studies using anthropological data produced mixed results[6–9]. The only experimental study to investigate how group size influences cumulative cultural evolution reported no relationship[10]. However, only one cultural task was considered, and the larger group size was limited to three individuals. More parameters must be explored experimentally to investigate the effect of group size on cultural complexity.

Following Henrich's analysis, the maintenance of a cultural task within a group should depend on group size and task complexity. Specifically, within a group of a particular size, greater loss of information is expected for a more complex task. Alternatively, for a task of a given complexity, greater loss of information is expected in a smaller group. Thus, when considering two improvable tasks, one simple and one complex, artificially introduced into groups of different sizes, we predict that the simple task will be better conserved than the complex task (prediction 1); the probability of conserving the complex task will increase with group size (prediction 2); and better performance will be observed in the larger groups for both tasks (prediction 3).

To study the effect of group size on cultural complexity, 366 men participated in a dual-task computer game. Players had to collect resources individually to improve their 'health'. A cultural package composed of two demonstrations, one concerning a simple task and one concerning a complex task, was introduced within groups of different sizes (2, 4, 8 or 16 players). The players were told that each item in the cultural package could be improved. During each of the 15 trials of the game, each player had to build an arrowhead (simple task) or a fishing net (complex task) to collect 'life units' (see Extended Data Fig. 1). The cultural trait diversity of the group thus consisted of some players building one artefact, while the remaining players built the other; diversity was lost when all individuals built the same object.

As expected from prediction 1, the simple task was more likely than the complex task to be maintained for all group sizes ($\chi^2 = 3.83$, d.f. = 1, $P = 0.05$; Fig. 1a, b). For each task, the probability of being lost (none of the individuals of the group exploited the task at the end of the game, see Methods) by a group decreased with increasing group size ($\chi^2 = 7.62$, d.f. = 1, $P = 0.006$), as expected from prediction 2 (Fig. 1a, b). Interestingly, the increased probability of maintaining the complex task in large groups did not reduce the probability of maintaining the simple task (type of task × group size interaction $\chi^2 = 0.85$, d.f. = 1, $P = 0.36$). Indeed, the probability of maintaining cultural diversity (that is, observing both tasks in the group) increased with group size ($\chi^2 = 16.3$, d.f. = 1, $P < 0.0001$; Fig. 1c).

For each group size, the performances of the best within-group artefacts (simple and complex) at the fifteenth trial were compared to the score of the equivalent artefact from the cultural package. The simple task was stable in the smaller groups and improved in the larger groups (Fig. 2). A linear model was used to investigate the effect of

[1]University of Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. [2]CNRS, Institute of Evolutionary Sciences, CC 065, Place Eugène Bataillon, Montpellier, France.
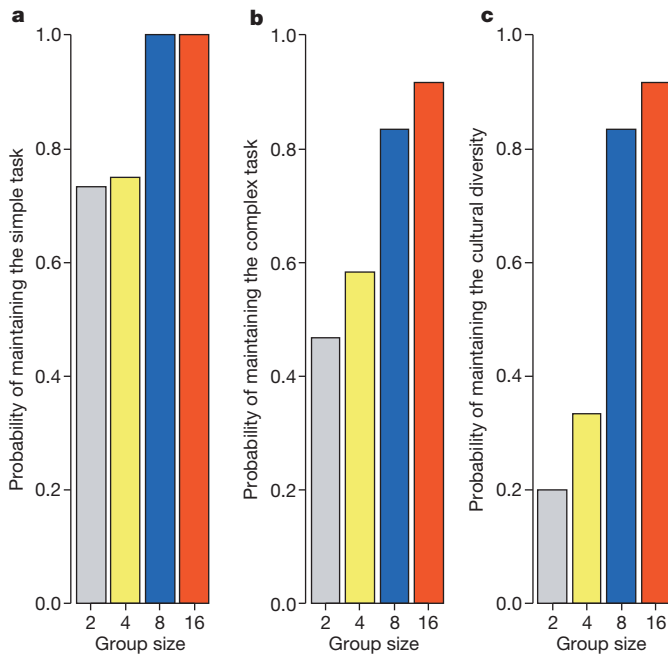
Figure 1 | Group size affects the maintenance of cultural tasks.
a–c, Probability of at least one observation of the simple task (**a**), the complex task (**b**) or both (that is, cultural diversity) (**c**) among the three last trials, for group size of 2 ($n = 15$ replicates), 4 ($n = 12$), 8 ($n = 12$) and 16 ($n = 12$) players.

group size and shows that group size had a linear effect on the performance of the best within-group arrowhead, suggesting that cultural evolution was enhanced in larger groups, consistent with prediction 3 ($F_{1,48} = 10.2$, $P = 0.003$; Fig. 2 and Extended Data Fig. 2). Performance of the complex task deteriorated in the smaller groups and remained stable in the larger groups (Fig. 3). Group size had a linear and quadratic effect on the performance of the best within-group fishing net ($F_{1,47} = 7.12$, $P = 0.01$ and $F_{1,47} = 4.22$, $P = 0.05$, respectively; Fig. 3). Among groups maintaining the complex task, only the 8- and 16-player groups improved it compared to the original cultural package (see Extended Data Figs 3 and 4).

The improvement of both tasks was linked to group size, suggesting that refinement of pre-existing technology is facilitated by increasing group size. The link between innovation rate and group size is not
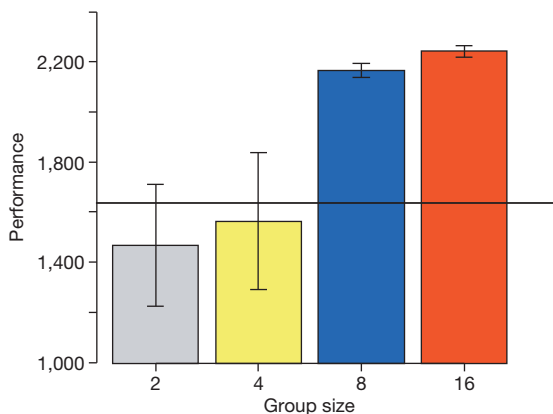

Figure 2 | Larger groups favour improvements to the simple cultural trait.
The horizontal line shows the arrowhead performance from the cultural package. Performance is measured using arbitrary life units. Plotted are the mean values ± s.e.m. The simple task was stable in the smaller groups (mean performance: 2-player groups = 1,466, $t = -0.71$, d.f. = 14, $P = 0.49$; 4-player groups = 1,563, $t = -0.27$, d.f. = 11, $P = 0.79$) and improved in the larger groups (8-player groups = 2,166, $t = 18.84$, d.f. = 11, $P < 0.0001$; 16-player groups = 2,242, $t = 27.57$, d.f. = 11, $P < 0.0001$).
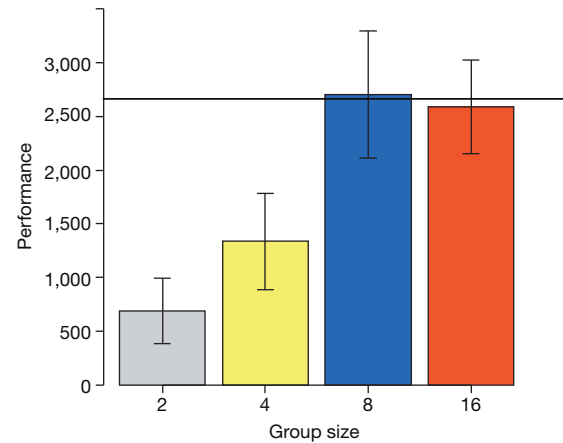

Figure 3 | Larger groups prevent degradation of the complex cultural trait.
The horizontal line shows the fishing-net performance from the cultural package. Performance is measured using arbitrary life units. Plotted are the mean values ± s.e.m. The complex task deteriorated in the smaller groups (mean performance: 2-player groups = 685, $t = -6.50$, d.f. = 14, $P < 0.0001$; 4-player groups = 1,334, $t = -2.99$, d.f. = 11, $P = 0.01$) and remained stable in the larger groups (mean performance: 8-player groups = 2,706, $t = 0.07$, d.f. = 11, $P = 0.95$; 16-player groups = 2,590, $t = -0.17$, d.f. = 11, $P = 0.87$).

surprising, as the combination of inter-individual variance in cognitive abilities and sampling effect increase the probability of observing high performers within a large group. Furthermore, a group can collectively achieve a solution to a cognitive problem that is not available to an individual through 'swarm intelligence'[24]. Whatever the mechanism, the best within-group artefacts drove the performance of the entire group, as shown by the correlation between best within-group artefacts and other within-group artefacts at the final trial (arrowhead, Pearson correlation = 0.39, $t = 5.53$, d.f. = 167, $P < 0.0001$; fishing net, Pearson correlation = 0.29, $t = 2.78$, d.f. = 87, $P = 0.007$).

When technological complexity is measured by the number of existing tools in the cultural repertoire, archaeological data produce mixed results[6–9]. The occurrence of new tools is poorly understood, but individuals rarely invent new tools from scratch; pre-existing technologies should have a role through combination; that is, bringing together two established cultural traits to generate a new trait[18,25,26]. Interestingly, this game suggests that increasing group size favours the maintenance of cultural diversity, a prerequisite for subsequent innovation through combination. It is worthy of note that the aim of the game was to maximize the player's 'health'. Thus, a player not able to perform the complex task (for example, lacking good visual memory) could perform better by efficiently repeating the simple task than by trying the complex one. It suggests that the individual diversity associated with larger group size could be pivotal to the maintenance of cultural trait diversity. By facilitating the maintenance of cultural diversity, increasing group size could also favour the emergence of division of labour at the group level. Such conditions pave the way for the emergence of inter-individual collaborations and group-level organization, some of the most important properties of human groups[27].

At the individual level, results also show that complex-task (fishing net) copying was most of time associated with a loss of skill, whereas simple-task copying was not (see Supplementary Information). This confirms that greater loss of information is expected for a more complex task, as suggested by Henrich[4]. At the group level, the maintenance of the complex task observed in large groups is thus explained by an increased probability to observe rare events directly linked to group size, such as a perfect copy or even an innovation, rather than overall better individual copying abilities. Following an innovation, prestige bias leads individuals to shift, and copy a new model. Even if copying deteriorates information, the mean group performance can increase, allowing cultural evolution to operate[4]. Accordingly, cultural

complexity—as measured in the archaeological record, for example—is most probably not a direct marker of the mean cognitive ability, as an ecological increase in population size could trigger the onset of a cumulative cultural evolution. Such an event may subsequently lead to the evolution of advanced copying ability, as this trait will most probably be an advantage in such a cultural environment. The players' difficulty in properly copying the fishing net from the cultural package (100% of fishing-net builders failed at the first trial) also illustrates the importance of multiple demonstrations and multiple attempts in the acquisition process[28]. In our game, players acquired the correct skill over several trials. In large groups, high-performing copiers (more likely to be observed as group size increases) can prevent the skill from disappearing, enabling players who lack good copying ability to benefit from more demonstrations.

Our results support Henrich's hypothesis: changes in group size can generate both adaptive cultural evolution and maladaptive losses of culturally acquired skills[4]. In our evolutionary past, group-size reduction may have exposed human societies to notable risks, as humans live in many habitats to which they are ill-suited without specific cultural adaptations[11,12]. Indeed, the more that we depend for our survival on large bodies of culturally transmitted knowledge, the more we rely on living in large groups. Under such conditions, group-size reduction could have triggered important loss of skills, leading to societal collapse[13], particularly in challenging environments. Interestingly, some cumulative cultural innovations, such as writing, printing and various forms of long-term data storage, allow the preservation of information outside of individuals, such that it is unknown whether the maintenance of current cultural complexity is nowadays similarly dependent on group size.

METHODS SUMMARY

Each player was randomly assigned to a group of 2, 4, 8 or 16 players, and all groups started the game by benefiting from the same cultural package (composed of an arrowhead and a fishing net, see Methods section for the complete details of the game). The simple task involved drawing an arrowhead, for which the performance evaluation depended only on its shape. The arrowhead demonstration in the cultural package involved 15 steps and provided 1,638 life units. The complex task involved building a fishing net, for which the performance evaluation depended on its shape and the procedure used to build it. The fishing-net demonstration in the cultural package involved 39 steps (the sequence of which mattered) and provided 2,665 life units. The starting individual life level was 3,400 units, and 1,000 units (daily needs) were subtracted at each trial. The task difficulties were designed so that, for a non-experienced player, the probability of scoring below their daily needs (and thus having a negative score) was low when choosing the arrowhead task and high when choosing the fishing-net task. Each trial was followed by an information period during which players could choose a single demonstration to observe (ranked by their performance), from one of their group members or the cultural package. The cultural package was available up to the third trial: from the fourth trial and after, social information came only from players' group members. A total of 366 male students (mean age = 24.1 years, s.d. = 4.4) played this game only once, in groups of 2 (15 replicates), 4, 8 or 16 (12 replicates each) players.

1. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (Univ. of Chicago Press, 1985).
2. Tomasello, M., Kruger, A. C. & Ratner, H. H. Cultural learning. *Behav. Brain Sci.* **16,** 495–511 (1993).
3. Boyd, R. & Richerson, P. J. *Not by Genes Alone* (Univ. of Chicago Press, 2005).
4. Henrich, J. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *Am. Antiq.* **69,** 197–214 (2004).
5. Powell, A., Shennan, S. & Thomas, M. G. Late Pleistocene demography and the appearance of modern human behavior. *Science* **324,** 1298–1301 (2009).
6. Kline, M. A. & Boyd, R. Population size predicts technological complexity in Oceania. *Proc. R. Soc. B* **277,** 2559–2564 (2010).
7. Read, D. Population size does not predict artifact complexity: analysis of data from tasmania, arctic hunter-gatherers, and oceania fishing groups. *UC Los Angeles: Hum. Complex Systems* http://www.escholarship.org/uc/item/61n4303q (2012).
8. Collard, M., Kemery, M. & Banks, S. Causes of toolkit variation among hunter-gatherers: a test of four competing hypotheses. *Can. J. Archaeol.* **29,** 1–19 (2005).
9. Read, D. An interaction model for resource implement complexity based on risk and number of annual moves. *Am. Antiq.* **73,** 599–625 (2008).
10. Caldwell, C. A. & Millen, A. E. Human cumulative culture in the laboratory: effects of (micro) population size. *Learn. Behav.* **38,** 310–318 (2010).
11. Henrich, J. & Broesch, J. On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Phil. Trans. R. Soc. Lond. B* **366,** 1139–1148 (2011).
12. Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: why social learning is essential for human adaptation. *Proc. Natl Acad. Sci. USA* **108,** 10918–10925 (2011).
13. Diamond, J. The Tasmanians — the longest isolation, the simplest technology. *Nature* **273,** 185–186 (1978).
14. Boyd, R. & Richerson, P. J. Why does culture increase human adaptability. *Ethol. Sociobiol.* **16,** 125–143 (1995).
15. Boyd, R. & Richerson, P. J. Why culture is common, but cultural evolution is rare? *Proc. Br. Acad.* **88,** 77–93 (1996).
16. Tomasello, M. In *Social Learning in Animals: the Roots of Culture* (eds Heyes, C. M. & Galef, B. G.) 319–346 (Academic, 1996).
17. Tennie, C., Call, J. & Tomasello, M. Ratcheting up the ratchet: on the evolution of cumulative culture. *Phil. Trans. R. Soc. Lond. B* **364,** 2405–2415 (2009).
18. Lewis, H. M. & Laland, K. N. Transmission fidelity is the key to the build-up of cumulative culture. *Phil. Trans. R. Soc. Lond. B* **367,** 2171–2180 (2012).
19. Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B. & Laland, K. N. Identification of the social and cognitive processes underlying human cumulative culture. *Science* **335,** 1114–1118 (2012).
20. Derex, M., Godelle, B. & Raymond, M. Social learners require process information to outperform individual learners. *Evolution* **67,** 688–697 (2013).
21. Claidière, N. & Sperber, D. Imitation explains the propagation, not the stability of animal culture. *Proc. R. Soc. B* **277,** 651–659 (2010).
22. Diamond, J. *Guns, Germs, and Steel: the Fates of Human Societies* (W. W. Norton, 1999).
23. Marquet, P. A. *et al.* Emergence of social complexity among coastal hunter-gatherers in the Atacama Desert of northern Chile. *Proc. Natl Acad. Sci. USA* **109,** 14754–14760 (2012).
24. Krause, J., Ruxton, G. D. & Krause, S. Swarm intelligence in animals and humans. *Trends Ecol. Evol.* **25,** 28–34 (2010).
25. Basalla, G. *The Evolution of Technology* (Cambridge Univ. Press, 1988).
26. Henrich, J. & Boyd, R. On modeling cognition and culture: Why cultural evolution does not require replication of representations. *J. Cogn. Cult.* **2,** 87–112 (2002).
27. Smaldino, P. E. The cultural evolution of emergent group-level traits. *Behav. Brain Sci.* (in the press).
28. Flynn, E. & Whiten, A. Studying children's social learning experimentally "in the wild". *Learn. Behav.* **38,** 284–296 (2010).

Author Contributions M.D., B.G. and M.R. designed the study. M.-P.B. and M.D. collected data. M.D., M.P.B. and M.R. analysed the data. All authors discussed the results and commented on the manuscript.

METHODS

Participants. A total of 366 male students were randomly selected from a database managed by the Laboratory of Experimental Economics of Montpellier (LEEM) and recruited by email from various universities in Montpellier (Southern France). The subjects ranged in age from 18 to 49 years (mean = 24.1 years, s.d. = 4.4 years). Each participant was randomly assigned to one condition of the experiment. Participants received fees for travel according to the LEEM operating rule (€2 for local participants, €6 for others).

Procedure. The experiment took place in a computer room at the LEEM. For each session, a maximum of 20 players sat at physically separated and networked computers and were randomly assigned to a group (the number of players per group varied according to the treatment, see below). They could not see each other, and they were blind with regard to the purpose of the experiment and who belonged to each group. The players were instructed that communication was not allowed. The participants could read instructions on their screens about the rewards and the goal of the game, and they were requested to enter their date of birth before the start of the game. At the end of the game, each subject received a reward according to his performance (€10 on average, see rewards calculation).

Principle. The participants played a computer game (programmed in Object Pascal with Delphi 6) during which they had to maximize their 'health' using two virtual tasks, making an arrowhead or a fishing net. Before the beginning of the game, players were advised that the fishing-net task was potentially more effective than the arrowhead task but that the fishing-net construction was more difficult. The participants were also informed that the performance of an arrowhead depended only on its shape, whereas the performance of the fishing net depended on its shape and the procedure used to build it. Each player began the game by observing a video demonstration of each task from a cultural package and was instructed that the arrowhead and fishing-net demonstrations could be improved. The arrowhead demonstration involved 15 steps and was associated with a score of 1,638. The fishing-net demonstration involved 39 steps (the sequence of which mattered) and was associated with a score of 2,665. The participants were not aware of the highest achievable score for any task.

The players then had 15 trials to collect resources and improve their health score. At each trial, they had the opportunity to build either an arrowhead or fishing net. Players began the game with a health score of 3,400 units. At each trial, their health level was reduced by 1,000 units, corresponding to their daily needs. Between trials, players could benefit from social information (see below).

Construction period. During the construction period (limited to 90 s), the players had to choose between the arrowhead task and the fishing-net task to collect resources.

The arrowhead task. The performance of an arrowhead depended only on its shape. The arrowhead score ranged from 0 to 2,400 units. A simple symmetric, triangular arrowhead constituted an acceptable performance equal to the player's daily needs. As a consequence, the probability of a non-experienced player scoring below his daily needs was low.

Construction details for the arrowhead task. First, the players had to choose the rectangular grid dimension on which to draw the arrowhead (30 possible values, Extended Data Fig. 1.a). Once the grid was chosen, the players had to draw their arrowhead. By clicking on the grid, the players could draw lines between points (Extended Data Fig. 1.b). The players had to draw the outline of their arrowhead and the virtual relief. No construction rules were implemented.

Score calculation for the arrowhead task. Once an arrowhead was drawn, it was evaluated by the program. The arrowhead was scanned pixel by pixel to evaluate five parameters: the size (α) and the symmetry (β) of the arrowhead, the number of notches (γ) and their regularities (δ), and the triangular shape (λ). All the parameters were compared to a theoretical optimal value and normalized from 0 to 1. The score S was then obtained according to this formula:

$$S = \alpha.400 + \beta.400 + \gamma.800 + \delta.400 + \lambda.400 \qquad (1)$$

The fishing-net task. The participants had access to several virtual tools with which to build their nets. The performance of a net depended on its shape and the procedure used to build it. The net's score ranged from 0 to 5,135 units. Departure from the construction rules (which were unknown to the players) resulted in increased penalties during use of the fishing net. As a consequence, the probability of a non-experienced player scoring below his daily needs (1,000 units) was high.

Construction details for the fishing-net task. First, the players had to choose the squared grid dimension on which to build the net (30 possible values, Extended Data Fig. 1c). Once the grid was chosen, the players had access to different types of ropes and knots, as in a previous experiment[20]. A rope could be set between any pair of attaching points, and a knot could be tied to any attaching point, in any order (Extended Data Fig. 1d). There were limited ropes and knots available. Each

additional rope placed on the frame decreased the length of the remaining rope according to the length used. This remaining quantity was visible on the screen. There were three different types of rope available (thick (red), medium (blue) and thin (green)). Each additional knot placed on the net decreased the length of the remaining knot quantity according to the type of knot used (three sizes available). This remaining quantity of knots was visible on the screen. Modification of one parameter produced complex interactions with others to generate a complex fitness landscape. For example, the use of the thickest ropes prevented the net from breaking but increased the net visibility so that the number of potentially caught fish was reduced. In addition, the order of construction (the process), was important. For example, two ropes that intersect at an attaching point should be tied together with a knot before another rope is put on the frame. If this step is omitted, the expected score is reduced.

Score calculation for the fishing-net task. Once a fishing net was constructed, it was evaluated by the program. A global resistance score (GR) was calculated according to the number of knots and compared to the required number. A local resistance score (LR_i) was determined for each mesh i according to the length and thickness of the ropes involved. During each virtual fishing exercise, 79 fish were launched, with a unique size of 65 (arbitrary units). The probability of each fish encountering the net increased according to the net overall size (set by the grid-spacing parameter) and decreased according to its visibility. The visibility of a net was computed as the sum of the lengths of all ropes used, weighted by their thicknesses. Once a fish was set to interact with the net, random coordinates were generated to identify at which mesh the interaction took place. If the fish was smaller than the mesh, it escaped. If it was larger, the probability of the net breaking was calculated as $1 - (GR*LR_i)$. In such a case, the whole fishing process stopped. If the net did not break, the fish could escape with a probability P_{esc}, which depended on the shape of the mesh and construction-rule penalty. If the fish did not escape, its size was added to the player's score. This process was repeated until the last fish was encountered or until the net broke.

Information period. After each trial, the resulting score, along with the player's health level, was displayed. The players could also see score lists for the arrowheads and fishing nets generated by the player's group members at the previous trial, ordered by performance. During the first three trials, the cultural package (arrowhead or fishing net) was included in the corresponding list.

By clicking on a score, the players could see the step-by-step procedure needed to build the selected item. Any demonstration lasted 40 s, regardless of the number of building steps. At each information period, a player could see only one demonstration. From the fourth information period, cultural-package demonstrations were removed from the lists. The players then had access only to their group member's demonstrations. The duration of the social-information period was 70 s.

Rewards calculation. The individual rewards were €10 on average. Players who died during the game (health level dropped below 0) earned €2. The other players earned an amount €A calculated according to this formula:

$$A = H_p/H_g.[5.N + 3.N_d] + 5 \qquad (2)$$

where H_p is the player's health level, H_g is the sum of the group's health levels, N is the size of the group, and N_d is the number of dead players within the group.

Treatments. Four group sizes were considered: 2 players, 4 players, 8 players and 16 players. All treatments were replicated 12 times, except for the 2-player treatment, which was replicated 15 times.

Cultural evolution. The aim of the study was to investigate the evolution of the cultural packages that were introduced in the experimental groups. Two types of analyses were carried out; one examined the maintenance of cultural tasks (whether some individuals exploited the cultural task at the end of the game), and the other examined the performance associated with the tasks. For each of the two tasks, we focused on the best within-group information because this information drives subsequent cultural evolution (due to prestige bias).

Maintenance of cultural tasks. Two models were used. One model investigated how the simple task was maintained in comparison with the complex task. A cultural task was considered to be maintained within a group if, among the last three trials, at least one individual of the group exploited the task. The response variable was the presence or absence of each task in each group. The independent variables were the type of task (arrowhead or fishing net), group size, mean age within the group, and type of task × group size interaction, with 'group identity' as a random factor. Generalized linear mixed models (binomial) were used.

The other model investigated how cultural diversity was maintained according to group size. Cultural diversity was considered to be observed within a group if, in the last three trials, at least one individual performed the arrowhead task, whereas at least one individual performed the fishing-net task. The response variable was the presence or absence of the diversity. The independent variables were the group size and mean age within the group. A generalized linear model (binomial) was used.

Best within-group information. The performances of the best within-group arrowheads at the fifteenth trial were compared to the score of the arrowhead from the cultural package, using a one-sample Student's t-test (if the distribution significantly departed from normality, a Mann–Whitney-Wilcoxon test was also performed; results were qualitatively similar, data not shown). A further linear model was used to investigate the effect of group size. In this case, the response variable was the score of the best within-group arrowhead at the fifteenth trial, and the independent variables were group size and mean age within the group. These two analyses were carried out again for the fishing-net performances.
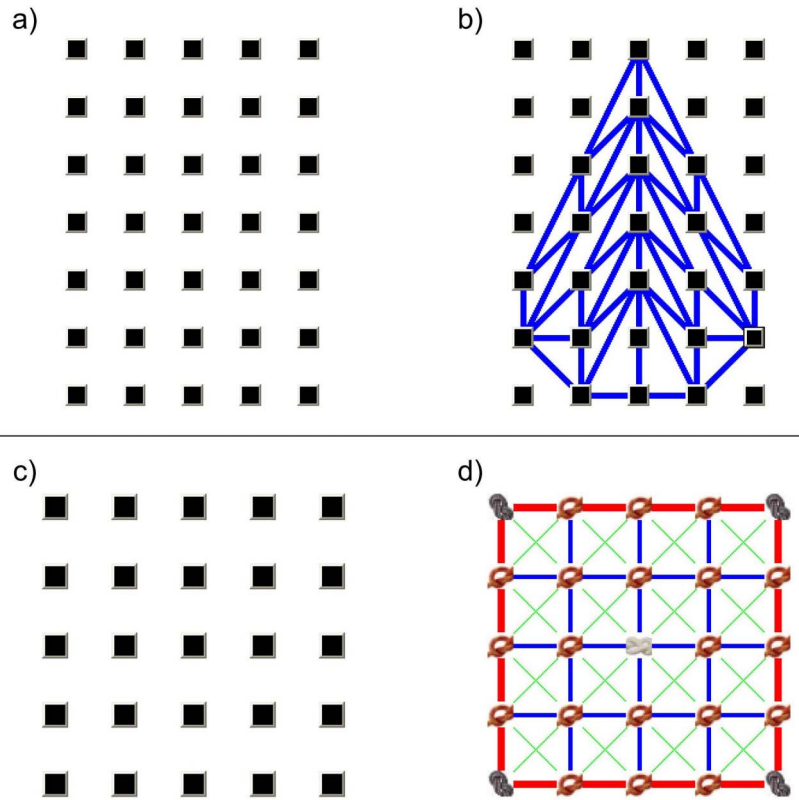
As groups could lose one of the two tasks, all analyses were carried out twice. In one case, we considered all groups, and performance score of zero was assigned when a task was lost from a group, that is, the degradation of the performance was considered complete (results shown in the main text). In the other case, we considered only the performance of the groups that conserved the task (results shown in Extended Data Figs 2 and 3).

Normality of residuals was significantly rejected (using Shapiro's test) in three models. This was owing to the presence of zero values (associated with task loss) generating a gap in the distribution between zero and the minimal score. When the presence or absence of the task was explicitly controlled for in order to estimate this gap, normality of residuals were not rejected (sometimes requiring the exclusion of only one outlier). All results described here were unchanged, whether or not these changes were made.

Fidelity of copying. Henrich's model assumes that information transmission is generally imperfect (particularly with complex tasks). Indeed, if copying is faithful, no cultural losses are expected. For each task, analyses were carried out to evaluate copying fidelity. During the observation period, players could choose a single demonstration to observe before building a new artefact. The aim was to study whether or not artefacts built by the players performed worse than the artefacts they observed.
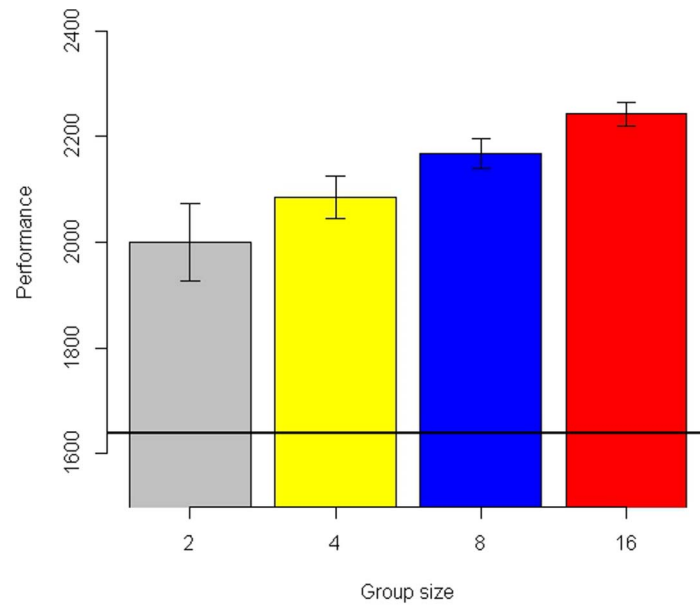
An observed artefact was considered as a model and was associated with n copies, depending on how many players observed the same model. For example, if three players observed the same model, three copies (copy 1, copy 2 and copy 3) were created. All possible pairs of artefacts were formed from the model and the copies: with one model and three copies, this corresponded to 6 pairs (model–copy 1; model–copy 2; model–copy 3; copy 1–copy 2; copy 1–copy 3; copy 2–copy 3). Comparisons of 'model–copy' represent our treatment of interest: if copying deteriorates information, the expected score difference (model score minus copy score) should be positive (null or negative otherwise). Comparisons of 'copy–copy' represent a control treatment: the expected score difference should be null. The focal artefact (first artefact from the pair) was either a copy or the model and was always compared to a copy (second artefact from the pair). The skill was considered to have deteriorated when the focal artefact outperformed the copy (score difference strictly positive). The binary response variable was the presence or absence of skill degradation. The independent variables were the type of the focal artefact ('copy' or 'model'). The identity of the focal artefact and the identity of the producer of the second artefact from the pair were included as random effects. A generalized linear mixed model (binomial) was used. All analyses were carried out separately for each task (arrowhead and fishing net).

Correlation between best within-group information and individual performances. This study was culture-centred, focusing on the state of the information available within groups (how the best within-group information performed). Considering that the best-within-group information influences the subsequent performance of the entire group, it is important to test the correlation between best within-group information and individual performances: owing to prestige bias, the best within-group information should affect the performance of the entire group. We examined the correlation between the best within-group information and the performance of the other players at the fifteenth trial using the Pearson correlation test.
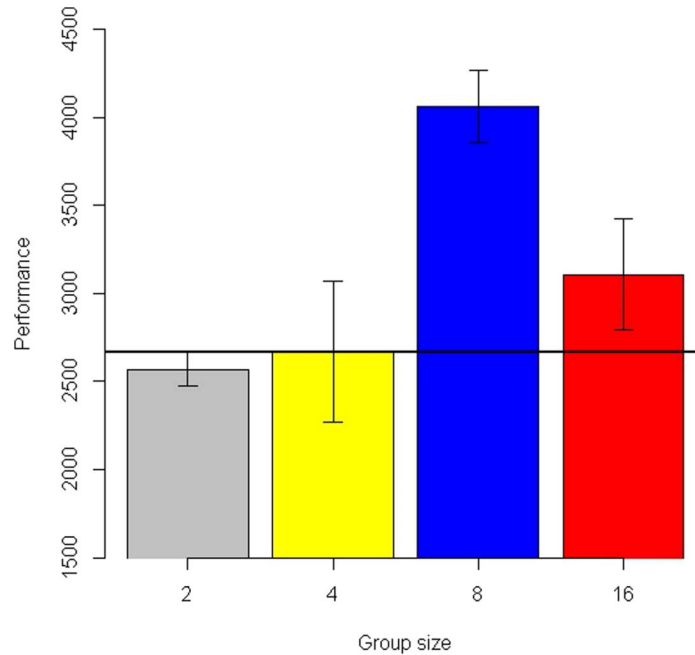
Extended Data Figure 1 | Cultural tasks. a, Rectangular grid composed of 35 attaching points in which to draw an arrowhead. The spacing between the attaching points was modifiable. **b**, An example of an arrowhead. **c**, Square grid composed of 25 attaching points in which to build a fishing net. The spacing between the attaching points was modifiable. **d**, An example of a fishing net.
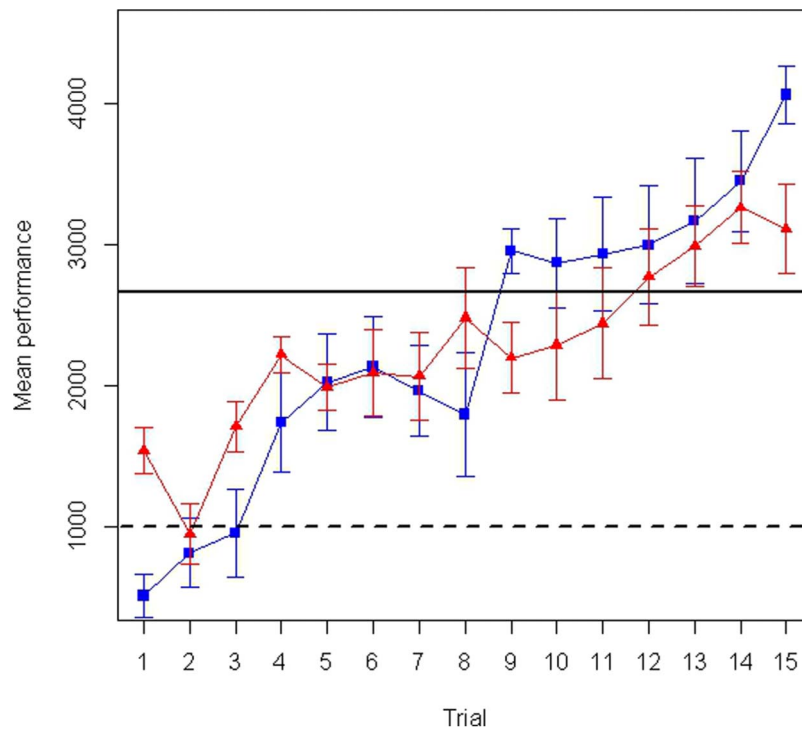
Extended Data Figure 2 | Best within-group information associated with the simple task, when conserved within the group. Performance is measured using arbitrary life units. Plotted are the mean values ± s.e.m. Considering only the performance of the groups that conserved the task (see Methods), the simple task of the cultural package was improved in all group sizes (mean performance: 2-player groups = 2,000, $t = 4.90$, d.f. = 10, $P = 0.0006$; 4-player groups = 2,085, $t = 11.12$, d.f. = 8, $P < 0.0001$; 8-player groups = 2,166, $t = 18.84$, d.f. = 11, $P < 0.0001$; 16-player groups = 2,242, $t = 27.57$, d.f. = 11, $P < 0.0001$). Group size had a linear effect on the performance of the best within-group arrowhead ($F_{1,41} = 15.3$, $P = 0.0003$). The horizontal line shows the performance of the arrowhead from the cultural package.

Extended Data Figure 3 | Best within-group information associated with the complex task, when conserved within the group. Performance is measured using arbitrary life units. Plotted are the mean values ± s.e.m. Only 4 2-player groups (26.7%) conserved the complex task and were therefore excluded from the analysis. The complex task was stable in the 4-player groups (mean performance = 2,669, $t = 0.01$, d.f. = 5, $P = 0.99$) and improved in the larger groups. The difference between 8-player groups and the demonstration of the cultural package was significant (mean = 4,059, $t = 6.79$, d.f. = 7, $P = 0.0001$, one-sided) but marginally significant concerning 16-player groups (mean = 3,108, $t = 1.40$, d.f. = 9, $P = 0.09$, one-sided). Group size had a linear and an unexpected quadratic effect on the performance of the best within-group fishing net ($F_{1,24} = 10.6$, $P = 0.003$ and $F_{1,24} = 9.88$, $P = 0.004$, respectively). This quadratic effect could indicate that participants had trouble making use of the information in a large group, but our experimental design allows us to rule out this possibility (see Supplementary Information). Instead, early performances of 16-player groups affected the probability of observing the cultural-package demonstration, hindering players from acquiring pivotal information (see Extended Data Fig. 4 and Supplementary Information). The horizontal line shows the performance of the fishing net from the cultural package.

Extended Data Figure 4 | Best within-group information associated with a fishing net (when conserved within the group) across time. The red line shows 16-player groups and the blue line shows 8-player groups. Performance is measured using arbitrary life units. Plotted are the mean values ± s.e.m. At the beginning of the game, the 16-player groups performed better than the 8-player groups ($F_{1,22} = 21.7$, $P = 0.0001$), as expected. However, the opposite was observed at the end of the game ($F_{1,16} = 5.68$, $P = 0.03$). During the first three trials, the performance associated with the best within-group fishing net affected the probability of observing the cultural-package demonstration. Thus, the probability of observing the cultural-package demonstration was lower in 16-player groups compared with 8-player groups. A lower rate of observation of the cultural-package reduced the group performance suggesting that the observation of demonstrations from other sources hindered the acquisition of pivotal information (see Supplementary Information for details). It suggests that, under specific conditions, the increasing number of valuable sources of information associated with larger group size could lead to a suboptimal cultural evolution rate. The horizontal solid line shows the performance of the fishing net from the cultural package. The horizontal dashed line shows the players' daily needs.