Collaborative learning in networks

Winter Mason[a,1] and Duncan J. Watts[b,1]

[a]Stevens Institute of Technology, Hoboken, NJ 07030; and [b]Yahoo! Research, New York, NY 10018

Complex problems in science, business, and engineering typically require some tradeoff between exploitation of known solutions and exploration for novel ones, where, in many cases, information about known solutions can also disseminate among individual problem solvers through formal or informal networks. Prior research on complex problem solving by collectives has found the counterintuitive result that inefficient networks, meaning networks that disseminate information relatively slowly, can perform better than efficient networks for problems that require extended exploration. In this paper, we report on a series of 256 Web-based experiments in which groups of 16 individuals collectively solved a complex problem and shared information through different communication networks. As expected, we found that collective exploration improved average success over independent exploration because good solutions could diffuse through the network. In contrast to prior work, however, we found that efficient networks outperformed inefficient networks, even in a problem space with qualitative properties thought to favor inefficient networks. We explain this result in terms of individual-level explore-exploit decisions, which we find were influenced by the network structure as well as by strategic considerations and the relative payoff between maxima. We conclude by discussing implications for real-world problem solving and possible extensions.

collaboration | diffusion | exploration-exploitation trade off

Many problems that arise in science, business, and engineering are "complex" in the sense that they require optimization along multiple dimensions, where changes in one dimension can have different effects depending on the values of the other dimensions. A common way to represent problem complexity of this nature is with a "fitness landscape," a multidimensional mapping from some choice of solution parameters to some measure of performance, where complexity is expressed by the "ruggedness" of the landscape (1, 2). A simple problem, that is, would correspond to a relatively smooth landscape in which the optimal solution can be found by strictly local and incremental exploration around the current best solution. By contrast, a complex problem would correspond to a landscape with many potential solutions ("peaks") separated by low-performance "valleys." In the event that the peaks are of varying heights, purely local exploration on a rugged landscape can lead to solutions that are locally optimal but globally suboptimal. The result is that when solving complex problems, problem solvers must strike a balance between local exploitation of already discovered solutions and nonlocal exploration for potential new solutions (3, 4).

In many organizational contexts, the tradeoff between exploration and exploitation is affected by the presence of other problem solvers who are attempting to solve the same or similar problems, and who communicate with each other through some network of formal or informal social ties (5–10). Intuitively, it seems clear that communication networks should aid collaborative problem solving by allowing individual problem solvers to benefit from the experience of others, wherein the faster good solutions are spread, the better off every problem solver will be. However, recent work on the relation between network structure and collaborative problem solving (11–13) has concluded that when faced with complex problems, networks of agents that exhibit lower efficiency can outperform more efficient networks,

where "efficiency" refers to the speed with which information about trial solutions can spread throughout the network. Although at first surprising, this finding is consistent with earlier results (3, 8) that slow or intermediate rates of learning in organizations result in higher long-run performance than fast rates. In both cases, the explanation is that slowing down the rate at which individuals learn, either from the "organizational code" (3) or from each other (8, 11, 13), forces them to undertake more of their own exploration, which, in turn, reduces the likelihood that the collective will converge prematurely on a suboptimal solution.

Although this explanation is persuasive, the evidence is based largely on agent-based simulations (2, 3, 8, 11, 13, 14), which necessarily make certain assumptions about the agents' behavior, and therefore could be mistaken in ways that fundamentally undermine the conclusions. In addition, one recent experiment involving real human subjects (12) found an advantage for inefficient networks in problem spaces requiring exploration. However, the practical constraints associated with running the experiments in a physical laboratory limited the size and variability of the networks considered as well as the complexity of the corresponding fitness landscapes; thus, it remains unclear to what extent the experimental findings either corroborate the simulation results or generalize to other scenarios.

Given these uncertainties, it would be desirable to test hypotheses about network structure and performance in human-subjects experiments in a way that circumvents the constraints of a physical laboratory. Fortunately, it has recently become possible to run laboratory-style human subject experiments online. Although virtual laboratories suffer from certain disadvantages relative to their physical counterparts, several recent studies have shown that many of these disadvantages can be overcome (15–17). In fact, a number of classic psychological and behavioral economics experiments conducted using Amazon's Mechanical Turk, a popular crowd-sourcing (18) site that is increasingly used by behavioral science researchers to recruit and pay human subjects, have recovered results indistinguishable from those recorded in physical laboratories (19, 20).

In this paper, we report on a series of online experiments, conducted using Amazon's Mechanical Turk, that explore the relation between network structure and collaborative learning for a complex problem with a wide range of network topologies. Several of our results are consistent with prior theorizing about problem solving in organizational settings. In particular, we find that networked groups generally outperform equal-sized collections of independent problem solvers (8), that exploration can be costly for individuals but is beneficial for the collective (3), and that centrally positioned individuals experience better performance than peripheral individuals (21, 22). However, we find no evidence to support the hypothesized superiority of inefficient
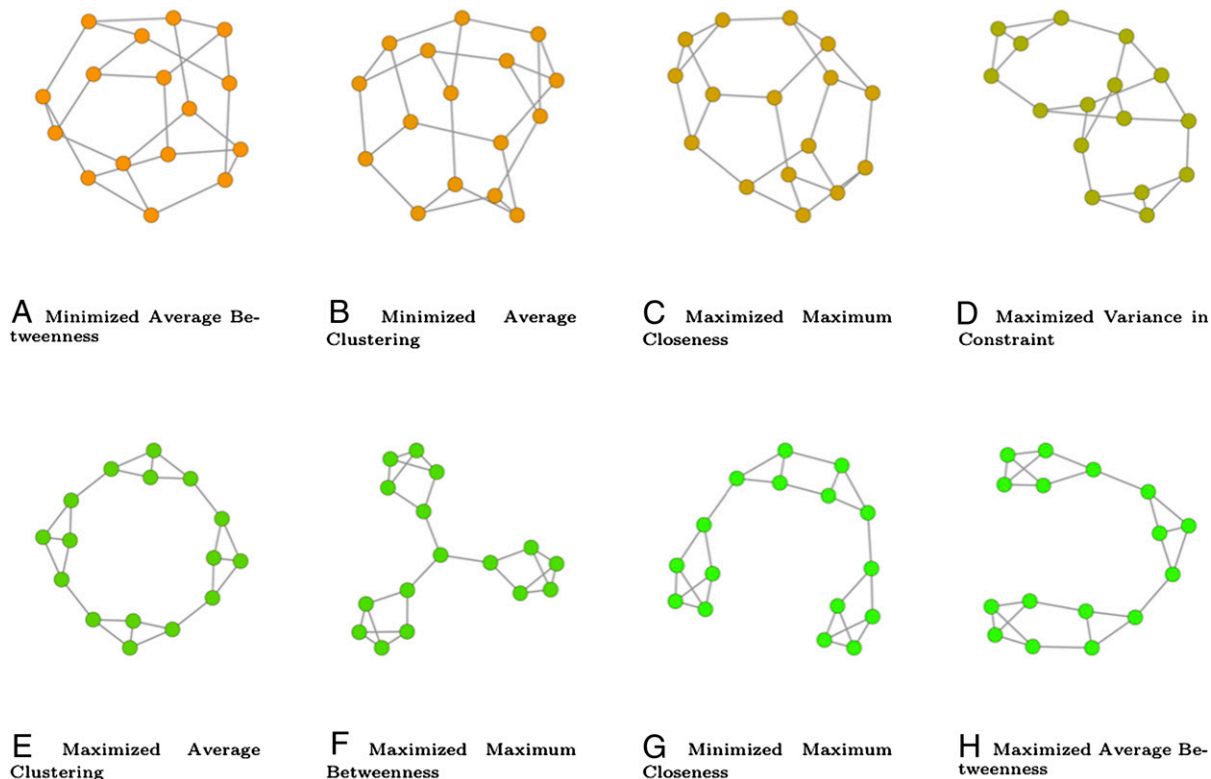
Fig. 1. Each of the 16-player, fixed-degree ($k = 3$) graphs used in the experiment, arranged in order of efficiency. The upper row constitutes networks whose decentralized nature connotes high efficiency (short path lengths), whereas the networks in the lower row are all centralized to some degree and also exhibit significant local clustering, both of which lower efficiency. The network on the upper right is an intermediate case, being decentralized but exhibiting some local structure.

networks (11–13). We note that although our problem spaces differ in some respects from prior work, they exhibit the essential features thought to advantage inefficient networks (11); indeed, the landscapes in question were designed explicitly to favor exploration. Rather, we argue the difference in our results stems from assumptions in the agent-based models that, although plausible and seemingly innocuous, turn out to misrepresent real agent behavior in an important respect.

Methods

The experiment was presented in the form of a game, called "Wildcat Wells," in which players were tasked with exploring a realistic-looking 2D desert map in search of hidden "oil fields." The players had 15 rounds to explore the map, either by entering grid coordinates by hand or by clicking directly on the map. On each round after the first, players were shown the history of their searched locations and payoffs, as well as the history of searched locations and payoffs of three "collaborators" who were assigned in a manner described in the "Networks" section. Players were paid in direct proportion to their accumulated points; hence, they were motivated to maximize their individual scores but had no incentive to deceive others (see *SI Text* and Figs. S1 and S2 for more details).

Fitness Landscape. All payoffs were determined by a hidden fitness landscape, which was constructed in three stages (more details are provided in *SI Text*):
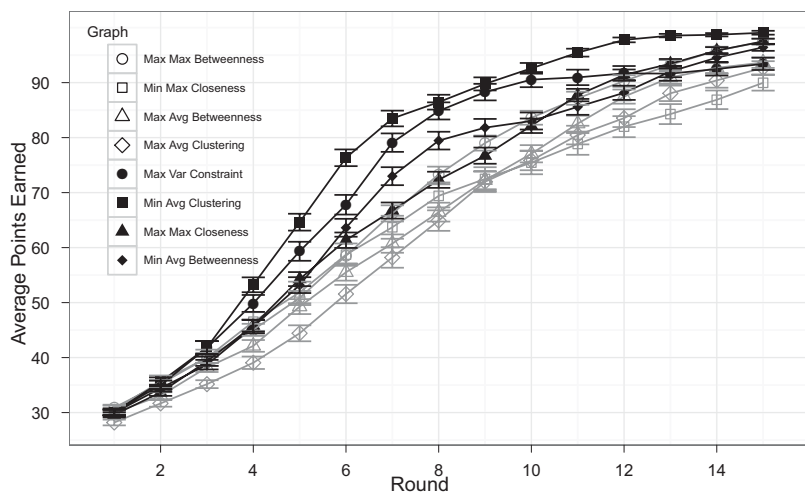
Fig. 2. Average points earned by players in the different networks over rounds (error bars are ± 1 SE) in sessions where the peak is found. Graphs with high clustering and long path lengths are shown in dark gray; those with low clustering and low path lengths are shown in light gray.

i) The "signal" in the payoff function was generated as a unimodal bivariate Gaussian distribution with the mean randomly chosen in an $L{\times}L$ grid with variance R.

ii) A pseudorandom Perlin noise (23) distribution was then created by summing a sequence of "octaves," where each octave is itself a distribution, generated in three stages:

    a) For some integer $\omega \in [\omega_{min}, \omega_{max}]$, divide the grid into $2^\omega \times 2^\omega$ cells and assign a random number drawn uniformly from the interval [0,1] to the coordinate at the center of each cell.

    b) Assign values to all other coordinates in the LxL grid by smoothing the values of cell centers using bicubic interpolation.

    c) Scale all coordinate values by ρ^ω, where ρ is the "persistence" parameter of the noise distribution.

iii) Having summed the octaves to produce the Perlin noise distribution (see also Figs. S3 and S4), the final landscape was created by superposing the signal and the noise, and then scaling the sum such that the maximum value was 100.

In its details, our method differs from other methods of generating fitness landscapes, such as March's exploration-exploitation model (3, 8, 13) and the *N-K* model (1, 2, 11), which generate *N*-dimensional problem spaces. We note, however, that all such models of fitness landscapes are highly artificial and are likely different from real-world problems in some respects. Thus, it is more important to capture the general qualitative features of complex problems than to replicate any particular fitness landscape. To satisfy this requirement, we chose $\omega \in (3, 7)$ and $\rho = 0.7$, which generates landscapes with many peaks in the range between 30 and 50 and a single dominant peak with maximum of 100 that corresponds to an unambiguously optimal solution [similar to Lazer and Friedman (11)]. Finally, we emphasize that players were not given any explicit information regarding the structure of the fitness landscape; hence, they could not be sure that the dominant peak existed in any given instance. To ensure that the peak was large enough to be found sometimes but not so large as to never be found, we conducted a series of trial experiments with different values of L and R, eventually choosing $L = 100$ and $R = 3$ (an example is shown in Fig. S5).

Networks. Before the start of each game, all players were randomly assigned to unique positions in one of eight network topologies, where each player's collaborators for that game were his or her immediate neighbors in that network. All networks comprised $n = 16$ nodes, each with $k = 3$ neighbors, but differed with respect to four commonly studied networks metrics: (*a*) betweenness centrality (24), (*b*) closeness centrality (25), (*c*) clustering coefficient (26), and (*d*) network constraint (22) (details are provided in *SI Text*). All four of these metrics have been shown to have an impact on the average path

length of a network, which, in previous work, has been equated with network efficiency (11, 13); hence, by varying them systematically (as described next), we were able to obtain networks that varied widely in terms of efficiency while maintaining the same size and connectivity.

The networks were constructed from regular random graphs ($n = 16$, $k = 3$) by making a series of degree-preserving random rewirings, where only rewirings that either increased or decreased certain properties of the four metrics above were retained. The properties of interest were the average value of the metric, the maximum value, the minimum value, and the variance. Thus, for example, the "max average clustering" network would be the network that, of all possible connected networks with fixed $n = 16$ and $k = 3$, maximized the average clustering of the network, whereas "max max betweenness" yielded the network whose most central node (in the betweenness sense) was as central as possible. In principle, this procedure could generate $2 \times 4 \times 4 = 32$ distinct equilibria (maximum/minimum of the property, 4 properties, and 4 metrics); however, many of these yielded identical or nearly identical networks. After eliminating redundancies, we then selected eight of the remaining networks shown in Fig. 1, which covered the widest possible range of network efficiency (ranging from 2.2 to 3.87), as shown in Table S1.

Experiments. Each experimental session comprised 8 games corresponding to each of the network topologies; thus, players experienced each topology exactly once in random order. We conducted 232 networked games over 29 sessions, where we note that because of the online nature of the participation, players sometimes failed to participate in some of the rounds. To check that our results were not biased by participant dropout, we also reanalyzed our data after removing 61 games in which at least 1 player failed to participate in more than half of the rounds (analysis with excluded trials is provided in *SI Text*), finding qualitatively indistinguishable results. Finally, in addition to conducting the networked experiments, we conducted a series of 24 baseline experiments, in which groups of 16 individuals searched the same landscape independently (i.e., with no network neighbors and no sharing of information), resulting in a total of 256 experiments comprising 16 players each.

Results

Collectives Performed Better than Individuals. Our first result is that networked collectives significantly outperformed equally sized groups of independent searchers. The main reason for this effect is obvious: Information about good solutions could diffuse throughout networked groups, allowing everyone to benefit from the discovery of even a single searcher in principle, whereas indepen-
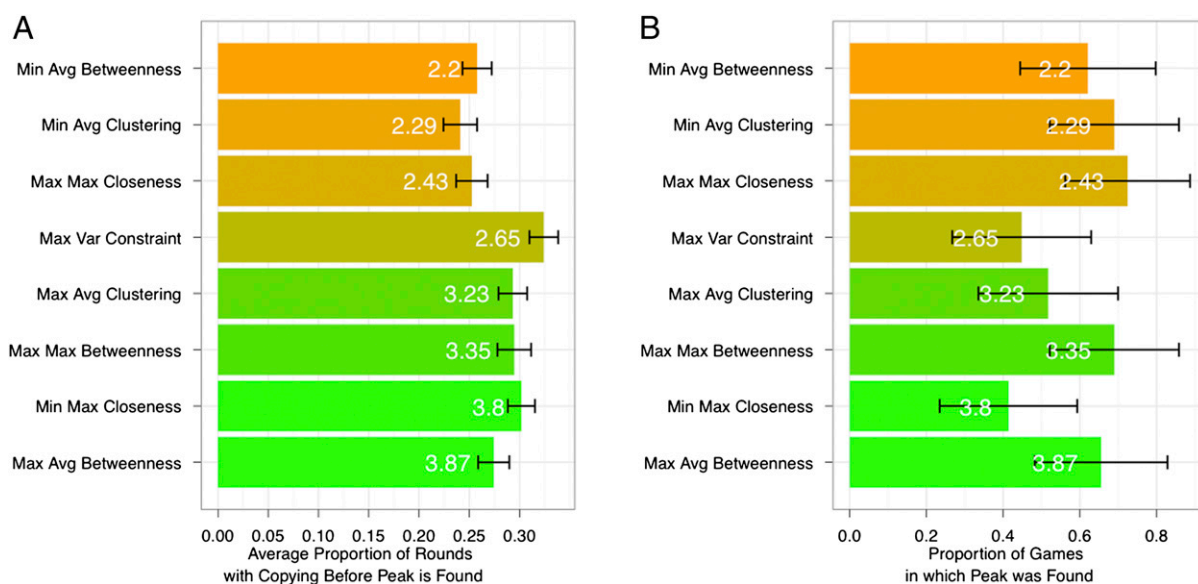
Fig. 3. (*A*) In contrast to theoretical expectations, less efficient networks displayed a higher tendency to copy; hence, they explored less than more efficient networks [numbers and colors (orange is shorter and green is longer) both indicate clustering coefficient]. (*B*) Probability of finding the peak is not reliably different between efficient and inefficient networks.
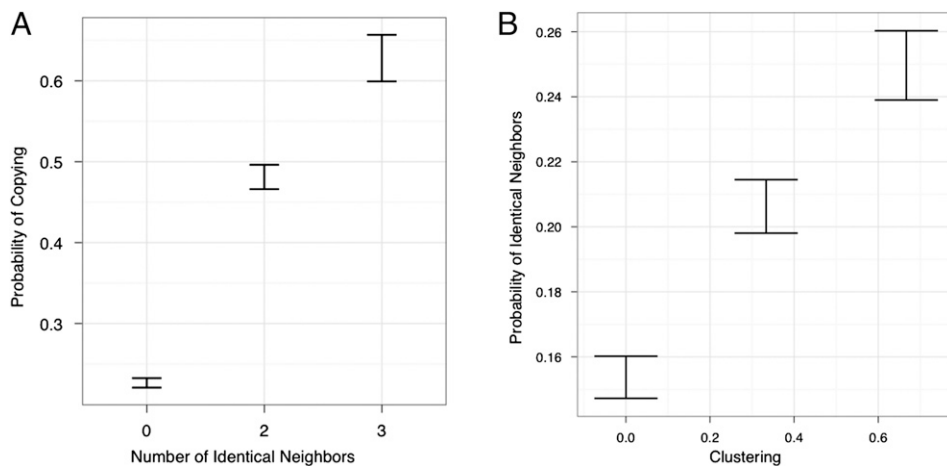
Fig. 4. (*A*) Probability of copying at least one neighbor's previous position increases with the number of neighbors who occupy identical positions. (*B*) Higher local clustering is associated with higher probability that an individual's neighbors currently occupy the same position.

dent searchers received no such benefit. Also, not surprisingly, the benefit to the diffusion of knowledge was particularly pronounced in instances in which at least one player found the main peak in the fitness function: Networked collectives earned 29 more points, on average, in such instances than individuals searching the problem spaces independently (68 vs. 39; $t = 92.5$, $P \approx 0$). A less obvious result, however, was that even when the main peak was not found, networked collectives still earned, on average, 8 points more ($M = 40.5$) than independent searchers ($M = 32$; $t = 35.6$, $P \approx 0$). Independent searchers who did not find the peak therefore earned only 2 points more than a completely random strategy, whereas individuals searching in a networked collective earned 10 points more than expected, suggesting that collectives were better able to take advantage of local maxima in the random landscape as well as the main peak.

Intuitively, one would expect that independent searchers, lacking the opportunity to copy each other, would explore more of the space than networked searchers; hence, one would also expect that the probability of at least one independent searcher finding the peak would be higher than for the equivalent number of networked searchers. Interestingly, however, although the probability was marginally higher for independent searchers, the difference was not significant (Fig. S6). Moreover, although the average distance between coordinates was indeed higher for independent searchers (Fig. S7*A*), the difference disappears when we confine the comparison to rounds in which the peak had not been discovered (Fig. S7*B*). Finally, although networked searchers did exhibit a tendency either to copy or explore close to the best-available solution in later rounds even when the peak had not been discovered, independent searchers showed a corresponding tendency to copy or explore close to their own best solutions (Fig. S8). Groups of independent searchers therefore gained no discernible benefit over networked searchers with respect to exploration, although clearly suffering with respect to exploitation.

Efficient Networks Outperformed Inefficient Networks. Next, we investigate the performance differences between our eight network topologies which, recall, varied by a factor of 50% with respect to efficiency (i.e., average path length; Table S1). We begin by noting that the result that networked collectives exploited local maxima more effectively than independent searchers demonstrates that players at times chose to exploit the local maxima of the fitness landscape rather than continuing to search for the main peak. We also note that of the 232 sessions, there were only 138 instances in which the peak was found by at least one player in a networked collective (59.5%). These circumstances, in other words, capture the qualitative features of fitness landscapes that

prior work suggests should favor inefficient networks (11, 13) over efficient networks, because individuals in inefficient networks should be less likely to get stuck on local maxima, and therefore be more likely to reap the large benefit of finding the main peak. To the contrary, however, regression analysis finds that the coefficient for average path length on score is negative and significant ($\beta = -4.63$, $P \approx 0.01$): Networks with higher efficiency (shorter path lengths) performed significantly better (Fig. S9 and Table S2).

Efficient networks outperformed inefficient networks for two reasons: first, because information about good solutions spread faster in efficient networks and, second, because, contrary to theoretical expectations, searchers in efficient networks explored more, not less, than those in inefficient networks. Fig. 2 illustrates the first finding, showing that for the 138 trials in which the peak was found, the eight networks divided into two rough groupings: those in the top row of Fig. 1, the efficient networks, and those in the bottom row, the inefficient networks. The effect of network structure on performance is therefore likely attributable to what has been called "simple contagion" (27), which depends exclusively on shortest path lengths. Supporting this conclusion, we find that the time required for the information to reach a node was exactly equal to the path length to the node that discovered the peak 71% of the time.

Fig. 3*A* illustrates the second result, showing that before finding the peak, searchers in inefficient networks had a greater tendency to copy each others' solutions than those in efficient
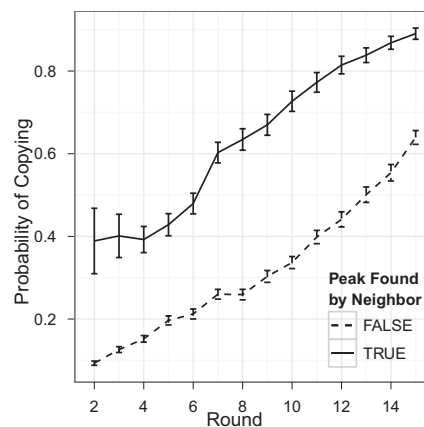
Fig. 5. Copying increases as a function of time, regardless of the discovery of the peak.
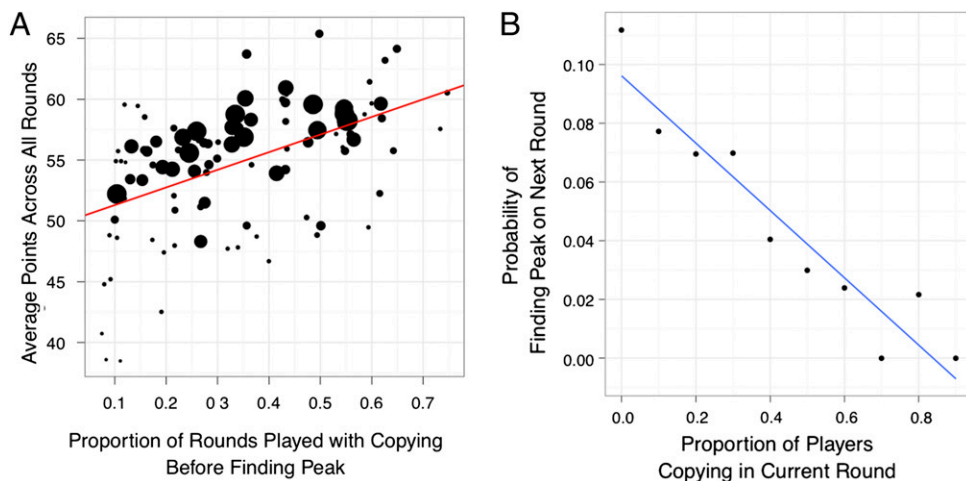
Fig. 6. (A) Average individual performance increases with individual tendency to copy his or her neighbors' positions (size of circle indicates the number of games played by a given subject, showing the same trend for frequent and infrequent participants). Larger circles indicate more games, and smaller circles indicate fewer games. (B) Proportion of players copying each other on any given round was associated with diminished probability of discovering the main peak of the fitness landscape on the subsequent round.

networks. In contrast to the theoretical expectation that lower efficiency should encourage exploration, in other words, we find it is efficient networks that explored more. Consistent with this result, it appears that efficient networks found the peak slightly more often than the inefficient networks (Fig. 3B), although this difference was not significant.

Hidden Network Structure Affected the Decision to Explore. This last finding that the network structure affected the players' tendency to copy is somewhat surprising, because, in addition to not being informed about the fitness landscape, players were not given information about the network they were in. Moreover, every player had the same number of neighbors; hence, all players' "view" of the game was identical irrespective of the details of the network or their position in it. Under similar circumstances, a recent study of public goods games (20) found that network structure had no significant impact on individual contributions; thus, one might have expected a corresponding lack of impact here. The explanation for the effect, however, can be seen in Fig. 4. First, as shown in Fig. 4A, individuals were more likely to copy their neighbors if two or more of their neighbors were exploiting the same location ($z = 38.9$, $P \approx 0$), even if that location was not near the peak of the fitness function. Second, as shown in Fig. 4B, higher clustering corresponded to a greater likelihood that an individual's neighbors would already have chosen the same location ($z = 15.96$, $P \approx 0$). What these results reveal is that local clustering allowed an individual's neighbors to see each other's locations, thereby increasing their likelihood of copying each other, which, in turn, increased the focal individual's tendency to copy his or her neighbors. Because the inefficient networks also had higher clustering, the players were more likely to copy each other than those in the efficient networks. In other words, by facilitating what amounts to "complex contagion" of information (27), in contrast to the simple contagion described above, inefficient network structure effectively reduced players' tendency to explore even though the networks themselves were invisible.

Players Faced a Social Dilemma. As noted earlier, players benefited considerably from discovering the main peak, earning 68 points on average when it was found, compared with 40 points when it was not ($t = 152.4$, $P \approx 0$). Clearly, therefore, one would expect players to copy more, and explore less, when one of their neighbors had discovered the main peak. Fig. 5 confirms this expectation; however, it also shows that the decision to exploit known solutions depended even more sensitively on time. In early rounds, that is, players mostly explored new locations whether or not the peak had been discovered (although more so when it had not), whereas in later rounds, they overwhelmingly

copied their most successful neighbor, again, regardless of whether or not their neighbors had discovered the peak. The likely explanation for these results is that players could not be certain that the peak existed, or that if it did, whether anyone would find it within the time limit. As a consequence of this uncertainty, players were increasingly unwilling to explore as the game progressed, preferring to exploit known local maxima. Less obviously, players also faced a version of a social dilemma in which individual and collective interests collide. On the one hand, Fig. 6A shows that players who copied more (before finding the peak) also tended to score higher ($t = 5.7$, $P < 0.001$); thus, players could improve their success by free-riding on others' exploration. On the other hand, Fig. 6B shows that the more players copied on any round, the lower was the likelihood of finding the peak on the next round ($z = -4.25$, $P < 0.001$); hence, the individual decision to exploit negatively impacted the performance of the collective.

Relative and Absolute Individual Success Depended on Position and Network Structure. Finally, in addition to affecting the average success of group members, network efficiency affected the distributions of success, and thereby relative individual performance, in two different and conflicting ways. First, as Fig. 7 shows, the
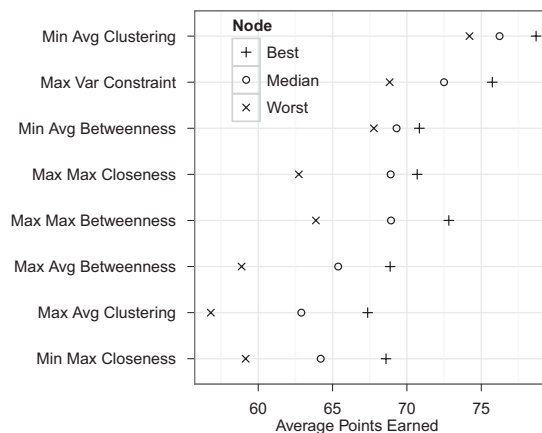
Fig. 7. Average points earned for best-positioned node, median node, and worst-positioned node in each of the graph types in sessions in which the peak was found. The decentralized graphs perform the best, and the most successful position in the centralized graphs only does as well as the median node in the decentralized graphs. Avg, average; Max, maximum; Min, minimum; Var, variance.

best-performing individuals, those who occupied central positions (i.e., with high betweenness or closeness centrality), consistently earned more points than their peers in the same network (See *SI Text* and Figs. S10 and S11 for more details). Second, however, Fig. 7 also shows that the average number of points earned for individuals in decentralized networks was comparable to the most points earned for centralized networks. Thus, depending on whether a given individual cares about relative or absolute performance, and assuming that he or she can also choose his or her network position, he or she may prefer to belong to a more or less centralized network, respectively.

Discussion

Our results reinforce certain conclusions from the literature on organizational problem solving but raise interesting questions about others. For example, our finding that exploitation is correlated with individual success but anticorrelated with collective success is consistent with previous results (3), which, interestingly, were derived from a very different model of organizational problem solving than ours. Moreover, our finding that individuals benefit from occupying central positions is reminiscent of long-standing ideas about the benefits of centrality (21) and bridging (22). However, our result that efficient networks perform unambiguously better than inefficient networks stands in contrast to previous findings that inefficient networks perform better in complex environments (11–13).

To understand the reasons for this difference better, we conducted a series of agent-based simulations in which artificial agents played the same game as our human agents, on the same fitness landscapes, but followed rules similar to those described in one previous study (11). Although we were able to replicate the result that inefficient networks outperformed efficient networks, it arose only when agents searched with an intermediate level of myopia (i.e, search that was either too local or too global yielded no difference in performance across networks). Moreover, for all parameter values, simulated agents copied more and performed worse than human agents did (see *SI Text* and Figs. S12–S17 for more details).

These results suggest that, at least for this class of problem, agent-based models have, to date, been insufficiently sophisticated and heterogeneous to reflect real human responses to changing circumstances, such as observed spatial correlations in the fitness landscape or the period of the game. For example, our comparison case (11) assumed that when at least one neighbor had a higher score, the focal agent would always copy; moreover, when agents "chose" to explore, they also adopted a fixed heuristic, independent of time and circumstances. By contrast, the search strategies that we observed in our experiments varied considerably across players, and also within players over the course of the game.

Recognizing that the results of both artificial simulations and artificial experiments should be generalized with caution, we conclude by outlining two possible implications of our findings. First, the result that networked searchers outperform independent searchers appears relevant to real-world problems, such as drug development, where competitive pressures lead firms to protect not only their successes but their failures. Given the considerable benefits to collective performance that our results indicate are associated with sharing information among problem solvers, future work should also explore mechanisms that incentivize interorganizational tie formation or otherwise increase efficiency in communication of trial solutions.

Second, we expect that network efficiency should not, on its own, lead to premature convergence on local optima. To clarify, recall that the intuition in favor of inefficiency is that slowing down the dissemination of locally optimal solutions also allows more time for globally optimal solutions to be found. However, our simulation results suggest that this extra time is useful only inasmuch as the overall rate of discovery is neither too slow nor too fast, an unlikely balance that does not appear to be displayed by actual human search strategies (more discussion is provided in *SI Text*). Moreover, we find that inefficient network structures can also decrease exploration, precisely the opposite of what the theory claims they are supposed to accomplish. We emphasize that this result does not contradict other recent work pointing out the collective advantages of diversity (28) or of other aspects of interactivity, such as turn taking (29). Rather, it suggests that manipulating network structure to impede information flow is an unreliable way to maintain organizational diversity, and indeed may have the opposite effect. Assuming that diversity can be maintained via more direct mechanisms, such as recruiting efforts, allocation of individuals to teams (28), or choice of alliances (30), our results therefore imply that efficient information flow can only be advantageous to an organization.

1. Kauffman SA (1993) *The Origins of Order: Self Organization and Selection in evolution* (Oxford Univ Press, New York).
2. Levinthal DA (1997) Adaptation on rugged landscapes. *Manage Sci* 43:934–950.
3. March JG (1991) Exploration and exploitation in organizational learning. *Organ Sci* 2: 71–87.
4. Gupta AK, Smith KG, Shalley CE (2006) The interplay between exploration and exploitation. *Acad Manage J* 49:693–706.
5. Bavelas A (1950) Communication patterns in task-oriented groups. *J Acoust Soc Am* 22:725–730.
6. Leavitt HJ (1951) Some effects of certain communication patterns on group performance. *J Abnorm Psychol* 46(1):38–50.
7. Uzzi B, Spiro J (2005) Collaboration and creativity: The small world problem. *Am J Sociol* 111:447–504.
8. Miller K, Zhao M, Calantone R (2006) Adding interpersonal learning and tacit knowledge to March's exploration-exploitation model. *Acad Manage J* 49:709–722.
9. Guimerà R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308:697–702.
10. Rendell L, et al. (2010) Why copy others? Insights from the social learning strategies tournament. *Science* 328:208–213.
11. Lazer D, Friedman A (2007) The network structure of exploration and exploitation. *Adm Sci Q* 52:667–694.
12. Mason WA, Jones A, Goldstone RL (2008) Propagation of innovations in networked groups. *J Exp Psychol Gen* 137:422–433.
13. Fang C, Lee J, Schilling MA (2010) Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organ Sci* 21:625–642.
14. Rivkin JW (2000) Imitation of complex strategies. *Manage Sci* 46:824–844.
15. Mason WA, Watts DJ (2009) *Financial Incentives and the Performance of Crowds*, Proceedings of ACM SIGKDD Workshop on Human Computation, pp 77–85.
16. Paolacci G, Chandler J, Ipeirotis P (2010) Running experiments on Amazon Mechanical Turk. *Judgm Decis Mak* 5:411–419.
17. Mason W, Suri S (2011) Conducting Behavioral Research on Amazon's Mechanical Turk. *Behavior Research Methods*, 10.3758/s13428-011-0124-6.
18. von Ahn L, Dabbish L (2008) Designing games with a purpose. *Commun ACM* 51(8): 58–67.
19. Horton J, Rand D, Zeckhauser R (2011) The online laboratory: Conducting experiments in a real labor market. *Exp Econ* 14(3):399–425.
20. Suri S, Watts DJ (2011) A study of cooperation and contagion in networked public goods experiments. *PLoS One* 6(3):e16836.
21. Bonacich P (1987) Power and centrality: A family of measures. *Am J Sociol* 92: 1170–1182.
22. Burt RS (1992) *Structural Holes: The Social Structure of Competition* (Harvard Univ Press, Cambridge, MA).
23. Perlin K (2002) Improving noise. *ACM Trans Graph* 21:681–682.
24. Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41.
25. Beauchamp MA (1965) An improved index of centrality. *Behav Sci* 10(2):161–163.
26. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393:440–442.
27. Centola D, Macy M (2007) Complex contagions and the weakness of long ties1. *Am J Sociol* 113:702–734.
28. Page S (2008) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton Univ Press, Princeton).
29. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330: 686–688.
30. Phelps C (2010) A longitudinal study of the influence of alliance network structure and composition on firm exploratory innovation. *Acad Manage J* 53:890–913.