doi:10.1038/nature22332

Locally noisy autonomous agents improve global human coordination in network experiments

Hirokazu Shirado[1,2] & Nicholas A. Christakis[1,2,3,4]

Coordination in groups faces a sub-optimization problem[1–6] and theory suggests that some randomness may help to achieve global optima[7–9]. Here we performed experiments involving a networked colour coordination game[10] in which groups of humans interacted with autonomous software agents (known as bots). Subjects ($n = 4{,}000$) were embedded in networks ($n = 230$) of 20 nodes, to which we sometimes added 3 bots. The bots were programmed with varying levels of behavioural randomness and different geodesic locations. We show that bots acting with small levels of random noise and placed in central locations meaningfully improve the collective performance of human groups, accelerating the median solution time by 55.6%. This is especially the case when the coordination problem is hard. Behavioural randomness worked not only by making the task of humans to whom the bots were connected easier, but also by affecting the gameplay of the humans among themselves and hence creating further cascades of benefit in global coordination in these heterogeneous systems.

Collective action and large-scale cooperation are important challenges[1–3]. Most work on cooperation has focused on the social dilemma aspect, namely, on getting people to be willing to make sacrifices for the greater good[11,12]. Yet, even when this dilemma can be addressed, there remains another substantial problem: coordination[4–6]. The difficulty of achieving optimal collective action in groups may arise not only from the conflicting interests among individuals, or between individuals and their group, but also as a consequence of the inability of individuals to effectively coordinate their actions globally. Even if all individuals behave properly in their local interactions, this may not result in the optimal outcome for the whole community[1,2].

Previous theoretical work has suggested a surprising, even paradoxical, solution to the coordination problem: adding 'noise'[13–15]. Noise is usually defined as meaningless information, and it is often seen as problematic[16]. When it comes to optimization, however, noise can help a system to reach a global optimum. For example, mutation has an essential role in evolution[17]; error can facilitate search for information[18]; random fish schooling may enhance survival[19]; and cooperation may benefit from deviant behaviour[7–9,20].

Here, we evaluate the benefits of noise in addressing the coordination problem of human groups[21,22]. As human interactions are embedded within social networks, we also consider the impact of network position on the potentially beneficial effect of noise[23]. We first characterize the collective-action dynamics of networks of people interacting in a classic colour coordination game[10]. Then, we test the effect of noise on collective performance using autonomous software agents (bots), manipulating both the noisiness and geodesic placement of the bots. By adding bots into experimental social networks, we therefore explore the performance of heterogeneous systems involving both real humans and autonomous agents, while also demonstrating a possible practical solution to the problem of global coordination itself.

We recruited 4,000 unique subjects online and randomly assigned them to 1 of 11 conditions in a series of 230 sessions (see Supplementary Information). Subjects were assigned a location in a network of 20 nodes, generated by a preferential attachment model[24]; the network structure was created *de novo* for each session by attaching new nodes (each with two links) to existing nodes; and subjects were placed into the resulting networks at random. The collective goal is for every node to have a colour different than all of its neighbour nodes[10]. This colour coordination game successfully captures the problem of systematic failure by sub-optimization in coordination; that is, while each individual attempts to reach a solution that is optimal for that individual, this may not be optimal for the whole group (Fig. 1a).

In the sessions, each subject was allowed to choose a colour from three choices (green, orange and purple) at any time. The number of colours made available was the minimum necessary to colour the entire network without conflicts, which is known as the chromatic number; and all networks in our experiments are, by construction, globally solvable. However, while all the networks allowed the subjects to reach the collective goal, the networks could (by chance) vary in their number of solutions (that is, the networks ranged from 6 to 13,824 possible colourings that would work, known as the chromatic polynomial; see Supplementary Information).

Subjects could see only the colours of neighbours to whom they were directly connected, in addition to their own colour. Thus, although a subject might have solved the problem from his or her own point of view, the game might continue because the network still had conflicts in other regions of the graph. In terms of the optimization problem, the cost function of the game is expressed as the sum of the number of conflicts. As in past work[10], the subjects got paid according to how long it took for all conflicts in the network to be resolved, and they had to complete the task within 5 min (see Supplementary Information for details).

Within this basic setup, we then introduced three bots into the network in exchange for the same number of humans (no bots were placed in the control sessions; see Supplementary Table 1). Subjects were not informed that there were bots in the game. We manipulated the noisiness of the bots as follows. In the 'zero noise' condition, the bots behaved with a simple, greedy strategy: when a bot had a chance to minimize colour conflicts with its neighbours, it chose that colour; otherwise, it maintained its current colour. In the other two conditions, the bots behaved with the same greedy strategy most of the time, but they also randomly picked a colour from the three permissible options regardless of their local situation with either a probability of 10% ('small noise') or 30% ('large noise'). In all of the conditions, the bots made decisions every 1.5 s, which was the typical human reaction time (Extended Data Fig. 1).

Independent of bot noise, we also manipulated their network location as follows. In the 'central' condition, the bots were assigned to the three positions that had the largest number of neighbours (the highest network degree). Likewise, in the 'peripheral' condition, the bots were assigned to the three positions with the lowest degree.

[1]Yale Institute for Network Science, Yale University, New Haven, Connecticut 06520, USA. [2]Department of Sociology, Yale University, New Haven, Connecticut 06520, USA. [3]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520, USA. [4]Department of Biomedical Engineering, Yale University, New Haven, Connecticut 06520, USA.

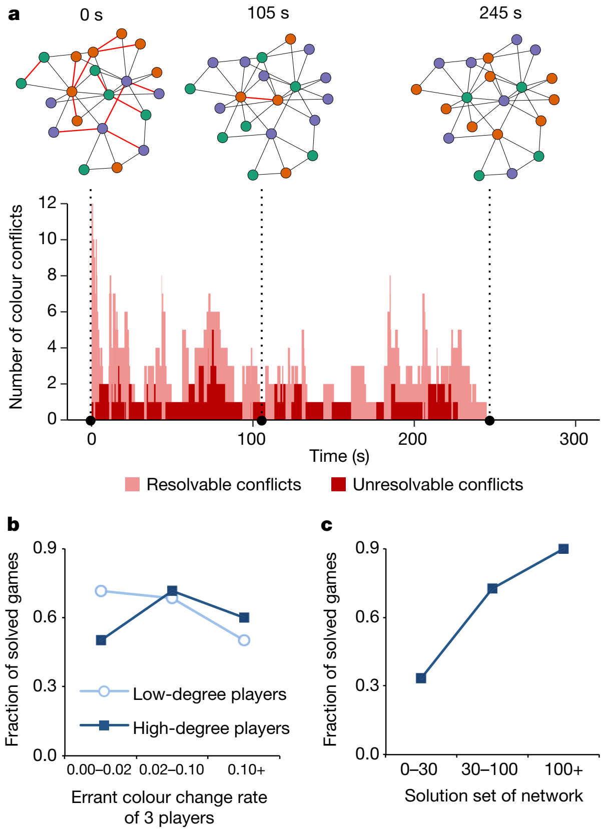


Figure 1 | Results of sessions involving only human players. a, An example of the colour coordination game. The figures are snapshots with players' node colour at 0, 105 and 245 s (see Supplementary Video 1 for full version). Red edges show that the connected players are the same colour (colour conflicts). Some conflicts can be resolved when either player selects the rarest colour among his/her neighbours (resolvable conflicts); but others cannot (unresolvable conflicts). **b**, The actual fraction of solved games depending on the behaviour of the most central or peripheral three players is shown ($n = 30$). The errant colour change rate is the ratio of colour selections (by the subjects) producing more colour conflicts divided by the opportunities to make such selections (see Supplementary Information for details). An intermediate level of errant colour choice among high-degree human players resulted in the greatest solvability (which comports with the programming strategy for helpful bots). **c**, The actual fraction of solved games in relation to the number of possible colour combinations (the 'chromatic polynomial') is shown ($n = 30$); having more possible solutions is associated with a higher solution rate.

In the 'random' condition, the bots were randomly assigned to their network locations. It was permissible for the bots to be connected to each other, by chance, in all conditions.

As noted, the bots acted using only their local information. To assess the effect of such bot behaviour compared to the much more demanding case requiring global knowledge of the entire network structure and its solution space in advance, we also carried out experiments with a 'fixed colour' condition. In this extra condition, we evaluated all colour combinations of each network that resulted in no conflicts, and then assigned the initial colours of three of the nodes based on one of those combinations (chosen at random). That is, during the game, the three nodes were not controlled by bots that coordinated with their neighbours, but rather, these nodes simply stayed at their initial colours, which were known to be consistent with a global solution to the problem. We examined this treatment only in the case in which the fixed nodes were in the central location.

In summary, we evaluated 11 conditions: 1 control condition not involving any bots; 9 treatment combinations of noise and location of bots (3 levels of behavioural randomness (0%, 10% and 30%) crossed with 3 types of location (random, central and peripheral)), and 1 final condition with 3 fixed-colour nodes. We conducted 30 sessions for the control condition and 20 sessions for each of the treatment conditions for a total of 230 sessions and 4,000 subjects.

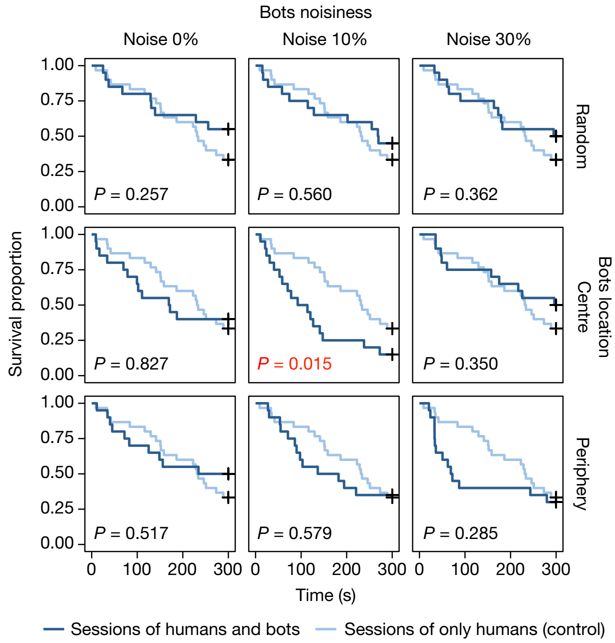


Figure 2 | Survival curves of sessions, by noisiness and location of bots. The curves show the percentage of sessions unsolved at a given time. Dark blue lines show results for the sessions including bots ($n = 20$), by their noise level (horizontal dimension) and geodesic location (vertical dimension). Light blue curves show results for the control sessions involving solely human players ($n = 30$). Total $n = 210$. Sessions are censored at 300 s; P values given by the log-rank test. Bots having 10% behavioural noise and located at the centre of the network cause a significant improvement in the solvability of the game ($P = 0.015$), and induce 55.6% acceleration in the median time to solution, from 232.4 s to 103.1 s.

For the games involving only human subjects, 20 out of 30 resulted in an optimal colouring of the network in less than the allotted 5 min (median time = 232.4 s; interquartile range (IQR) 143.7–300.0). Although the subjects aimed to eliminate all the conflicts, they often found themselves unable to reach the collective goal only by reducing their local conflicts on an individual basis. For example, as of 105 s in Fig. 1a (or Supplementary Video 1), each of the subjects had chosen one of the least common colours among their neighbours; that is, no one person could change their colour for the better. A conflict between neighbours, however, still remained. Such states in which players get caught in locally unresolvable conflicts are regarded as local minima of the cost function of the game (in contrast to resolvable conflicts that can be addressed by local action). Players would need a moderate level of deviancy from the norm of conflict minimization to overcome the local minimum and reach a global solution (for example, Fig. 1a, at 245 s).

By analysing the sessions involving only human subjects, it is possible to discern that games were more likely to be solved when some players occasionally chose a locally inappropriate colour, temporarily increasing conflicts; moreover, the effect of such behavioural deviance varied according to the geodesic location of the players, as captured by their network degree (Fig. 1b). In addition, and distinctly, some networks could be intrinsically easier to solve (that is, the chromatic polynomial could be higher) (Fig. 1c).

To demonstrate how bots could improve the performance of human groups, Fig. 2 shows survival curves of the sessions involving the nine

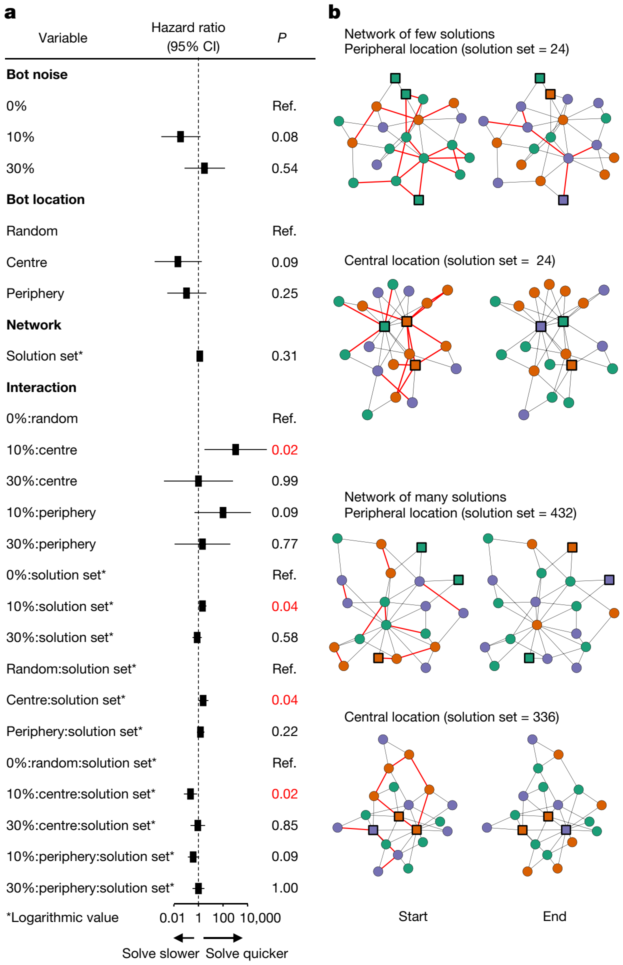


Figure 3 | Results of the survival analysis by bot and network characteristics. a, Hazard ratios for game solution time according to bot noise, bot location, number of solutions of the network (chromatic polynomial), and all interactions among these variables ($n = 180$; see Supplementary Table 3 for details). The results show that the benefit of bots varies with the solution space; when a network has few possible colour combinations, placing slightly noisy bots in a central location (high-degree nodes) facilitates resolution. 'Ref.' denotes the reference (or baseline) category for each variable. **b**, These network snapshots show initial and final states of illustrative sessions involving bots with 10% noise. Square nodes show the bots, and round nodes show human players; red edges show colour conflicts.

bot treatments. Before implementing pairwise comparisons of each treated group with the control group, we performed a log-rank test of the null hypothesis that all the survival curves are identical; that hypothesis was rejected ($P = 0.024$), indicating that at least two of the survival curves differed. The sessions having bots with 10% noise in central locations were the most likely to be solved within the allotted 5 min (17 out of 20 sessions, or 85%, compared with 20 out of the 30 control sessions, or 67%, with humans alone); moreover, the solution was achieved more than 129.3 s faster (that is, 55.6% faster) than sessions involving just humans (median time = 103.1 s (IQR 49.5–170.1) versus 232.4 s (IQR 143.7–300.0)), which was significantly better ($P = 0.015$, log-rank test).

We then examined the difference in effectiveness of the various bot treatments, while furthermore controlling for the intrinsic solvability of the network, using Cox proportional hazard models. Bot behavioural randomness of 10%, central location, and the logarithm of the chromatic polynomial all have a significantly positive effect on the completion time ($P < 0.05$; $n = 180$ bot-treated sessions; see Supplementary Information). We also evaluated another metric of the complexity of the solution space (that is, mean convergence steps with linear probabilities) and got similar results (Extended Data Fig. 2 and Supplementary Table 4). The statistical model with full interactions shows that the bots affect the solution time only when they behave with 10% randomness and are placed in the central location in the network (Fig. 3a); moreover, when the network affords many solutions, the beneficial impact of bots decreases, as shown by the three-way interaction (Fig. 3b). In short, the bots are especially helpful when the network is globally hard to solve.

We found that the impact of 10%-noise bots was comparable to the impact of assigning three nodes with fixed (constant) colours in a configuration known *ex ante* to be compatible with a global solution. There was no significant difference between the sessions with 10%-noise bots and the sessions with fixed colours ($P = 0.675$, log-rank test). Thus, the intervention of the bots, based on local decision-making alone, is equally as effective as a pre-calculated solution that (in typical circumstances) impractically would require previous global knowledge of the entire network structure and its solution space.

The bots appear to have improved collective performance in part by changing the colour-conflict behaviours of human players in the whole system (Extended Data Fig. 3). When placed at high-degree nodes, the bots with 0% behavioural randomness reduced the number of conflicts but they increased the duration of unresolvable conflicts; the bots with 30% randomness decreased the duration of unresolvable conflicts but increased overall conflicts; and only the bots with 10% randomness decreased both the number of conflicts and the duration of unresolvable conflicts, compared with the control sessions. By contrast, when placed at low-degree nodes, the bots were less likely to influence the entire network of humans, regardless of their noisiness.

When the bots were placed in high-degree positions, their behavioural randomness was able not only to facilitate the solution of their own conflicts, but also to nudge neighbouring humans to change their behaviour in ways that appear to have further facilitated a global solution. The bots with 0% behavioural randomness reduced the randomness of other human players (Fig. 4a), which made the human players, particularly the middle-degree players, come to be stuck in unresolvable conflicts (Fig. 4d). The bots with 30% behavioural randomness destabilized the entire network, including the low-degree players, who displayed more noise in their own actions (Fig. 4c); as a result, the sessions with 30%-noise bots showed the same level of unresolvable conflicts as those without bots (Fig. 4f). The bots with 10% behavioural randomness increased the randomness of the central players but reduced that of the peripheral players (Fig. 4b); hence, through the influence of their behavioural randomness, the 10%-noise bots reduced the unresolvable conflicts not only of themselves but also of the entire network, including links between human subjects unconnected to the bots (Fig. 4e). These results were obtained even though the subjects were, in fact, less and less satisfied with their counterparts the more noisy the bots became (Extended Data Fig. 4).

In a separate, further experiment involving an additional 340 subjects and a matched set of $n = 20$ graphs, we found that these beneficial effects on group coordination and learning were obtained even when players knew they were interacting with bots (see Supplementary Information). The solution time was statistically indistinguishable (Extended Data Fig. 5) and the effect on players throughout the system was also similar (Extended Data Fig. 6).

Adding autonomous agents with simple strategies into social systems may make it easier for groups of humans to achieve global optima for complex group-wide tasks. Here, the setting was a global coordination game, but other settings might include cooperation, sharing or navigation[5,12,25]. Any such bots, however, might only be helpful if they have certain properties, including noisiness or particular geodesic locations. Indeed, like other situations[13,14,17,18,20], some noise may be

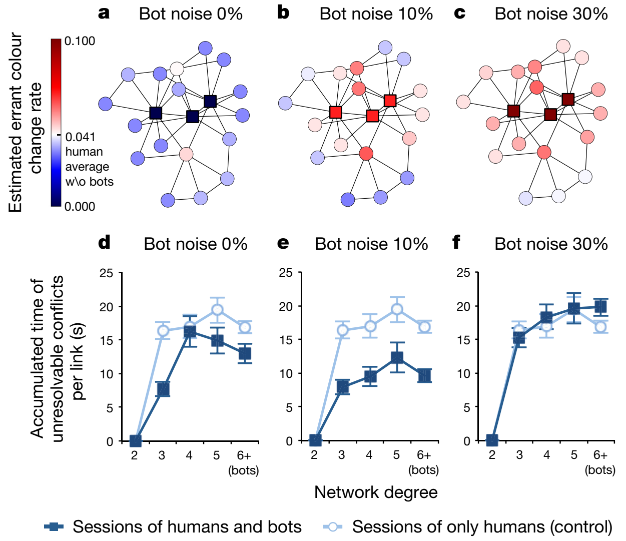


Figure 4 | Effect of bots on the behaviour of human players. **a**–**c**, Snapshots show estimates of the errant colour change rate (that is, humans choosing 'wrong' colours) in the same network with central bots, depending on bot noise. Square nodes show bots and round nodes show humans (see Supplementary Information and Supplementary Table 5 for regression modelling details). Note that the intermediate white colour shows the estimated errant rate of average human players in sessions without bots (0.041); thus, the red colour shows that human players behave in a more noisy way as a result of the influence of the bots; the blue colour shows the opposite. **d**–**f**, These graphs show the average accumulated time of unresolvable conflicts per link for each geodesic location of the players. Dark blue lines show results for sessions with central bots (whose degree was typically ≥ 6), by their noise level, and light blue lines show results for the control sessions with only humans. Bots with 10% noise (**e**) change the behaviours of the human players in the whole system for the better. Error bars denote s.e.m.

good from the point of view of the group. Moreover, bots with some noise, with solely local information, improved global outcomes here just as much as bots using global information acquired in advance.

We find that these slightly noisy bots work not only by making the task of humans to whom they are connected easier, but also by affecting the game play of the humans themselves when they interact with still other humans in the group, thus creating cascades of benefit. And this happens even when people know they are interacting with bots. In this sense, even simple artificial intelligence (AI) agents can serve a teaching function, changing the strategy of their human counterparts and modifying human–human interactions, and not just affecting human–bot interactions. More generally, our work illustrates the performance of combined, heterogeneous groups composed neither solely of humans nor solely of robots attempting to coordinate their actions. Future work can explore even more realistic or complex interactions, such as military or commercial robots working within human groups, or autonomous vehicles moving in a world of human-driven cars.

Although laboratory experiments afford robust causal inference, they must sacrifice some realism and breadth. Guided by prior theory, we chose to focus on only two aspects of bot contributions (noise and placement) and their effect on one primary outcome (success of global coordination in a standard game[10]). We also necessarily made other design choices, including using a scale-free network limited to 20 people (which was required if the games were to be tractable). But there are other features of social interactions that might affect the ability of groups to coordinate to solve a problem, such as group size, network topology[10], and bot fraction; whether the networks are dynamic or static[26,27]; or whether social institutions (for example, policing, sanctions or norms) are present. These elements are important directions for future work.

Adding bots of moderate noisiness to strategic positions within human networks might help to address diverse problems, especially when the particular coordination problem is hard. For example, narrowly focused workers might each labour to enhance their own productivity, but this might actually decrease overall company performance. Crowd-sourcing applications in science (such as solving quantum problems[28] or other types of 'citizen science' ranging from protein folding[29] to the assessment of archaeological or astronomical images) might be facilitated by adding some bots or noise to groups working collaboratively. Moreover, our work reinforces the idea that both simple and sophisticated AI might be useful. For instance, simple bots might help to reduce racist remarks online[30]. The simplicity and transparency of decision-making in simple AI of the kind we explore here might also make it intelligible to humans, thereby eliciting an effective, long-term relationship[11]. Simple autonomous agents, when mixed into complex social systems, might offer substantial advantages, and they could help groups of humans to help themselves.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 December 2016; accepted 5 April 2017.

1. Hardin, G. The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality. *Science* **162,** 1243–1248 (1968).
2. Dawes, R. M. Social deilemmas. *Annu. Rev. Psychol.* **31,** 169–193 (1980).
3. Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl Acad. Sci. USA* **108,** 11825–11829 (2011).
4. Calvert, R. Leadership and its basis in problems of social coordination. *Int. Polit. Sci. Rev.* **13,** 7–24 (1992).
5. Dyer, J. R. G., Johansson, A., Helbing, D., Couzin, I. D. & Krause, J. Leadership, consensus decision making and collective behaviour in humans. *Phil. Trans. R. Soc. Lond. B* **364,** 781–789 (2009).
6. Van Huyck, J. B., Battalio, R. C. & Beil, R. O. Tacit coordination games, strategic uncertainty, and coordination failure. *Am. Econ. Rev.* **80,** 234–248 (1990).
7. Nowak, M. Stochastic strategies in the prisoner's dilemma. *Theor. Popul. Biol.* **38,** 93–112 (1990).
8. Kandori, M., Mailath, G. J. & Rob, R. Learning, mutation, and long run equilibria in games. *Econometrica* **61,** 29–56 (1993).
9. Young, H. P. Learning by trial and error. *Games Econ. Behav.* **65,** 626–643 (2009).
10. Kearns, M., Suri, S. & Montfort, N. An experimental study of the coloring problem on human subject networks. *Science* **313,** 824–827 (2006).
11. Axelrod, R. *The Evolution of Cooperation* (Basic Books, 1984).
12. Rand, D. G. & Nowak, M. A. Human cooperation. *Trends Cogn. Sci.* **17,** 413–425 (2013).
13. Jiang, J.-J., Huang, Z.-G., Huang, L., Liu, H. & Lai, Y.-C. Directed dynamical influence is more detectable with noise. *Sci. Rep.* **6,** 24088 (2016).
14. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467,** 167–173 (2010).
15. Couzin, I. D. *et al.* Uninformed individuals promote democratic consensus in animal groups. *Science* **334,** 1578–1580 (2011).
16. Gao, J., Barzel, B. & Barabási, A.-L. Universal resilience patterns in complex networks. *Nature* **530,** 307–312 (2016).
17. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in experimental populations of E. coli. *Nature* **387,** 703–705 (1997).
18. Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. Optimization by simulated annealing. *Science* **220,** 671–680 (1983).
19. Kadri, U., Brümmer, F. & Kadri, A. Random patterns in fish schooling enhance alertness: a hydrodynamic perspective. *EPL* **116,** 1–6 (2016).
20. Traulsen, A., Semmann, D., Sommerfeld, R. D., Krambeck, H. J. & Milinski, M. Human strategy updating in evolutionary games. *Proc. Natl Acad. Sci. USA* **107,** 2962–2966 (2010).
21. Helbing, D. & Yu, W. The outbreak of cooperation among success-driven individuals under noisy conditions. *Proc. Natl Acad. Sci. USA* **106,** 3680–3685 (2009).
22. Couzin, I. D., Krause, J., Franks, N. R. & Levin, S. A. Effective leadership and decision-making in animal groups on the move. *Nature* **433,** 513–516 (2005).
23. Liu, Y.-Y., Slotine, J.-J. & Barabási, A. L. Controllability of complex networks. *Nature* **473,** 167–173 (2011).
24. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).

25. Kesting, A., Treiber, M., Schönhof, M. & Helbing, D. Adaptive cruise control design for active congestion avoidance. *Transp. Res., Part C Emerg. Technol*. **16,** 668–683 (2008).
26. Rand, D. G., Arbesman, S. & Christakis, N. A. Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl Acad. Sci. USA* **108,** 19193–19198 (2011).
27. Shirado, H., Fu, F., Fowler, J. H. & Christakis, N. A. Quality versus quantity of social ties in experimental cooperative networks. *Nat. Commun.* **4,** 2814 (2013).
28. Sørensen, J. J. W. H. *et al.* Exploring the quantum speed limit with computer games. *Nature* **532,** 210–213 (2016).
29. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466,** 756–760 (2010).
30. Munger, K. Tweetment effects on the tweeted: experimentally reducing racist harassment. *Polit. Behav.* http://dx.doi.org/10.1007/s11109-016-9373-5 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Allison, F. Fu, M. Kearns, G. Kraft-Todd, A. Oswald, D. Rand and D. Spielman for comments. M. McKnight provided technical support and programming for our Breadboard platform. Support for this research was provided by grants from the Robert Wood Johnson Foundation, the National Institute of Social Sciences, and the National Institutes of Health (P30-AG034420).

Author Contributions H.S. and N.A.C. designed the project. H.S. collected the data and performed the statistical calculations. H.S. and N.A.C. analysed the results. H.S. and N.A.C. wrote the manuscript. N.A.C. obtained funding.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to N.A.C. (nicholas.christakis@yale.edu).

Reviewer Information *Nature* thanks C. A. Hidalgo, I. D. Couzin, C. F. Camerer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

A total of 4,000 unique subjects (plus a further 340 for the secondary experiment regarding bot visibility; see Supplementary Information) participated in our incentivized economic game experiments. They were recruited using Amazon Mechanical Turk (AMT; see Supplementary Information), and they interacted anonymously over the Internet using customized software playable in a browser window (available at http://breadboard.yale.edu). While keeping other initial conditions the same, we completed 30 sessions for the only-human condition (control) and 20 sessions for each bot-treated condition (treatment). In each session (after passing various tutorials), the subjects were paid a US$2 show-up fee and a declining bonus of up to US$3 depending on the speed of reaching a global solution to the coordination problem (in which every player in a group had chosen a different colour from their connected neighbours). When they did not reach a global solution within 5 min, the game was stopped and the subjects earned no bonus.

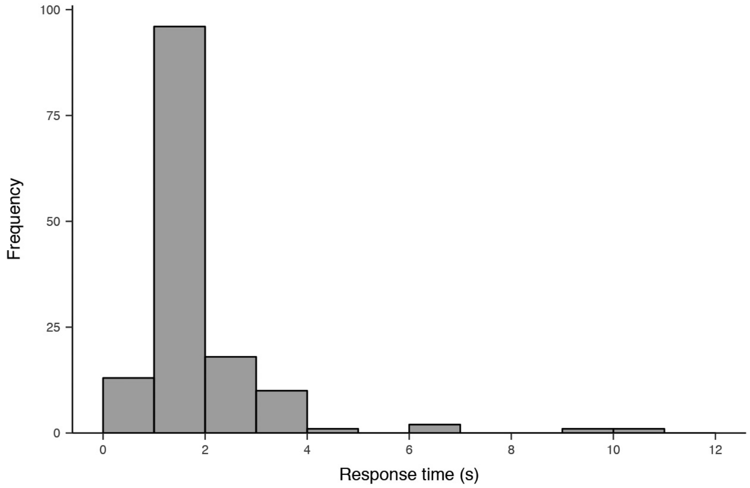
Except for the control group sessions, the networks had 3 bots in addition to 17 human subjects. These bots were assigned to three geodesic locations (peripheral, central, or random locations). The bots were controlled programmatically with a simple, greedy algorithm incorporating a random element; we drew a random number from a uniform distribution between 0.0 and 1.0; if the random number was less than a preset threshold ('behavioural noise'), the bot picked a colour among the three colour options at random; otherwise, it behaved based on the colours of its neighbours—that is, if the bot's current colour was not the least common among its neighbours, it changed to the least common colour; otherwise, it maintained the current colour.

To evaluate the difference in effectiveness between the various bot treatments, we analysed the solution time of the $n = 180$ sessions using Cox proportional hazard models. The sessions that were not solved within 300 s were regarded as censored. Each network session had a distinct level of complexity with respect to finding a colouring solution because it was generated *de novo*; thus, we controlled for the number of possible colour combinations of the network (the chromatic polynomial). We also performed various statistical robustness checks (see Supplementary Information).

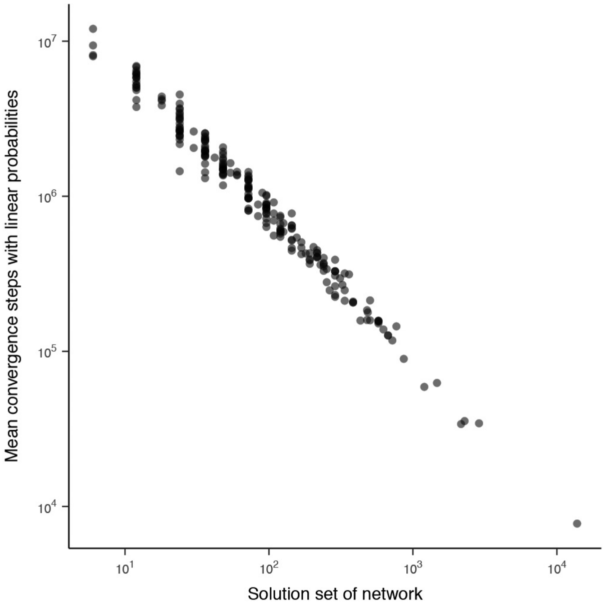
We examined the impact of bots' behavioural noise on the humans' behaviour using a generalized linear mixed model (GLMM) involving logistic regression (see Supplementary Information). The dependent variable is the errant colour-change rate evinced by the human players (that is, choices that deviated from the simple, greedy strategy to minimize local conflicts). The model incorporated fixed effects for the behavioural noise of bots, the number of neighbours, the number of neighbouring bots, the session length, and random effects for session. We predetermined sample size so as to be able to evaluate at least a 30% difference in game solvability, based on two-sample tests for equality of proportions. The investigators were not blinded to allocation during analysis.

This research was approved by the Yale University Committee of the Use of Human Subjects. All the subjects were informed about the use of their behavioural data for research purposes upon enrolment in the experiment and consented. We verified colour perception and understanding of the game rules in all the subjects using five multiple-choice questions; we excluded applicants who failed to select the correct answer in any of these questions.

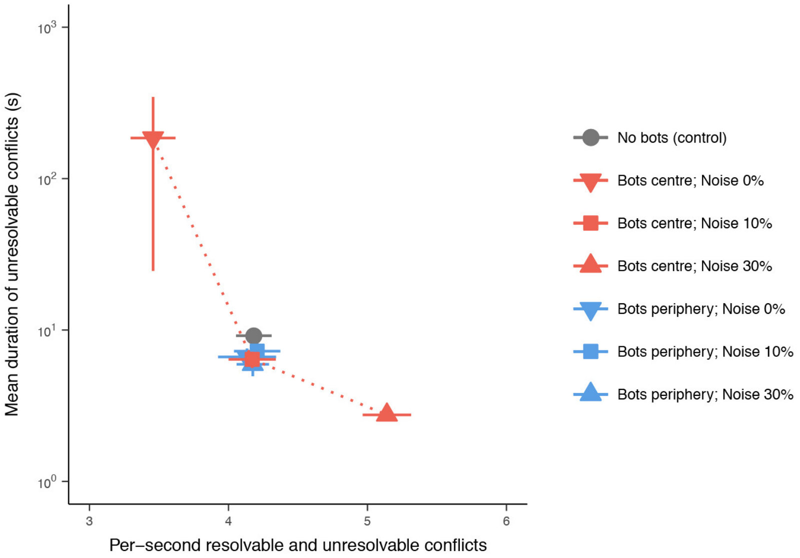
Data availability. The data reported in this paper are archived at Yale Institute for Network Science and are available upon request.



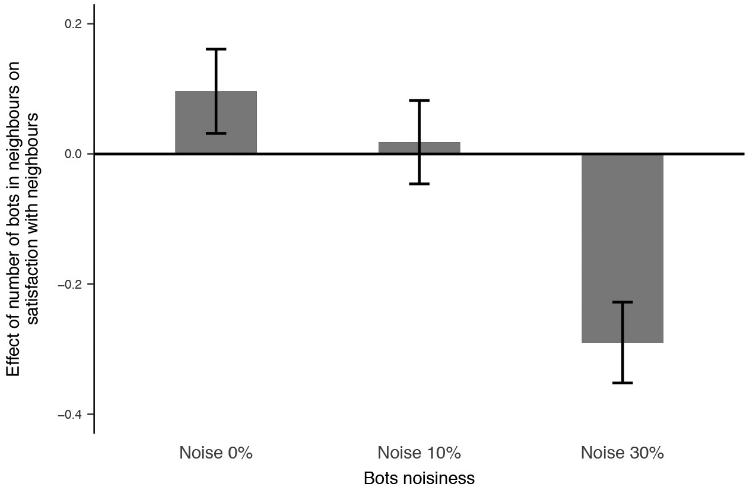
Extended Data Figure 1 | Histogram of the response time of humans in the colour-matching test ($n = 142$). In the colour-matching test in our preliminary experiments, subjects were asked three times to click the same colour button as a picture on the screen with five options: green, orange, purple, pink and yellow. This histogram shows the response time (from when a colour in question showed up on screen to when a subject clicked a button) for 142 pilot subjects. Most subjects clicked the correct button in 1.0–2.0 s (median time = 1.59 s).



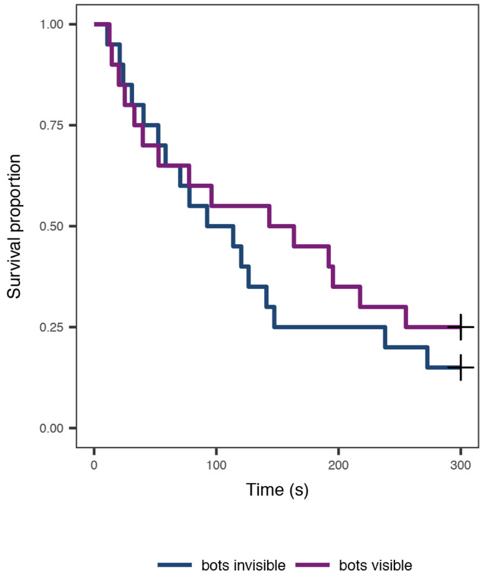
Extended Data Figure 2 | Relationship between different measures of the structure-based complexity of the graph colouring sessions. The correlation coefficient after logarithmic transformation is −0.990 ($P < 0.001$; $n = 230$). The solution set (x axis), known as the chromatic polynomial, is the number of possible colour combinations that satisfy the task of colouring the network. The average number of steps to reach a solution (y axis) involves computing the following statistic: a node is randomly selected and changes its colour to one that is different from its random neighbour and this is repeated until a solution is reached; the number of steps is then measured. This linear probability algorithm offers the advantage of allowing us to evaluate the landscape of the solution space starting from an arbitrary initial value. The mean convergence steps statistic was calculated for 100 iterations of each experimental network given the same initial colouring.



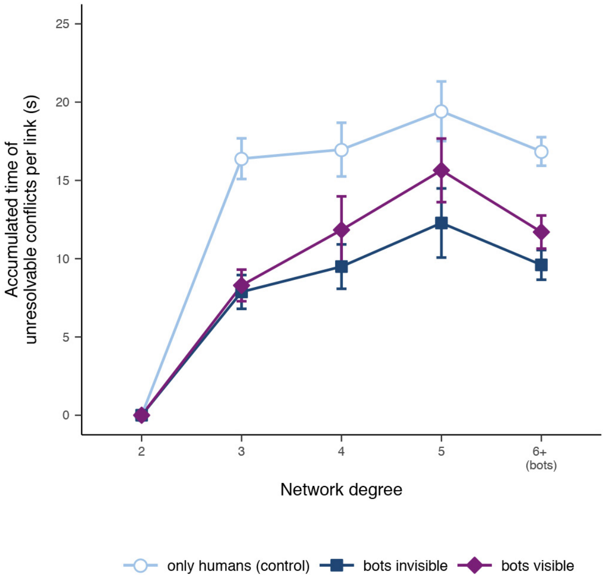
Extended Data Figure 3 | Impact of bots on colour conflicts over the entire network. The error bars are s.e.m. ($n = 30$ for the no-bots sessions; $n = 20$ for all the bot-treated sessions). When placed in the centre, bots with 0% behavioural noise reduce the number of conflicts but increase the duration of unresolvable conflicts; bots with 30% noise decrease the duration of unresolvable conflicts but increase the overall conflicts; and bots with 10% noise decrease both the number of conflicts and the duration of unresolvable conflicts, compared with results of only human players. In contrast to central placement, when bots are placed in the periphery, conflict status does not vary with behavioural noise (data points are overlapping).



Extended Data Figure 4 | Effect of bots' behavioural noise on players' satisfaction with their neighbours. After each session was completed, subjects rated their satisfaction with the actions of their neighbours on a five-point scale: very satisfied, satisfied, neither, dissatisfied, and very dissatisfied (the specific question asked was: "How satisfied were you with the actions of your neighbours you were connected with?"). These coefficients show the effect of number of bots among neighbours on subjects' satisfaction with their neighbours, estimated by a proportional odds logistic regression, incorporating number of neighbours and whether the session was solved. The error bars are s.e.m. ($n = 3,035$).



Extended Data Figure 5 | Survival curves for sessions by bot visibility. The curves show the percentage of sessions unsolved at a given time. Dark blue lines show the $n = 20$ sessions (involving $n = 340$ additional subjects) where human players were informed of which nodes were played by bots (visible-bots condition; $n = 20$), and light blue lines show the sessions where humans were not informed (invisible-bots condition; $n = 20$). The difference of the survival curves is not statistically significant ($P = 0.435$, log-rank test).



Extended Data Figure 6 | Effect of bot visibility on players' unresolvable conflicts for each geodesic location. The dark purple line shows results for the sessions where human players were informed of which nodes were played by the bots (visible-bots condition; $n = 20$), the dark blue line shows results from the sessions where humans were not informed (invisible-bots condition; $n = 20$). In both conditions, the bots were located at high-degree nodes with 10% noise. The light blue line shows results for the sessions with all human players as a control ($n = 30$). The error bars are s.e.m. by session. Except for the addition of the dark purple line (the results of the visible-bots condition), this figure is the same as Fig. 4e. Pertinently, the dark purple and dark blue lines are not statistically distinguishable, suggesting that making the bots visible has a similar effect throughout the network on players' behaviour compared to keeping them invisible.