Behavioral experiments on biased voting in networks

Michael Kearns[1], Stephen Judd, Jinsong Tan, and Jennifer Wortman

University of Pennsylvania, Department of Computer and Information Science, 3330 Walnut Street, Philadelphia, PA 19104

Edited by Ronald L. Graham, University of California at San Diego, La Jolla, CA, and approved December 5, 2008 (received for review August 19, 2008)

Many distributed collective decision-making processes must balance diverse individual preferences with a desire for collective unity. We report here on an extensive session of behavioral experiments on biased voting in networks of individuals. In each of 81 experiments, 36 human subjects arranged in a virtual network were financially motivated to reach global consensus to one of two opposing choices. No payments were made unless the entire population reached a unanimous decision within 1 min, but different subjects were paid more for consensus to one choice or the other, and subjects could view only the current choices of their network neighbors, thus creating tensions between private incentives and preferences, global unity, and network structure. Along with analyses of how collective and individual performance vary with network structure and incentives generally, we find that there are well-studied network topologies in which the minority preference consistently wins globally; that the presence of "extremist" individuals, or the awareness of opposing incentives, reliably improve collective performance; and that certain behavioral characteristics of individual subjects, such as "stubbornness," are strongly correlated with earnings.

behavioral game theory | collective decision making | network science

The tension between the expression of individual preferences and the desire for collective unity appears in decision-making and voting processes in politics, business, and many other arenas. Furthermore, such processes often take place in social or organizational networks, in which individuals are most influenced by, or aware of, the current views of their network neighbors.

The 2008 Democratic National Primary race offers a recent, if approximate, example of this phenomenon. On the one hand, individual voters held opposing and sometimes strong preferences that were apparently very nearly balanced across the population; however, there was a strong and explicit desire that once the winning candidate was identified, the entire party should unify behind that candidate (1). Obviously primary voters could be influenced by many global factors (such as polls and mainstream media) outside the scope of their individual social and organizational networks, but presumably for many voters these local influences still played an important and perhaps even dominant role.

Although there is now a significant literature on the diffusion of opinion in social networks (2–4), the topic is typically studied in the absence of any incentives toward collective unity. In many contagion-metaphor models, individuals are simply more or less susceptible to "catching" an opinion or fad from their neighbors, and are not directly cognizant of, or concerned with, the global state. In contrast, we are specifically interested in scenarios in which individual preferences are present but are subordinate to reaching a unanimous global consensus.

We report here on an extensive session of human-subject experiments meant to provide a simple abstraction of the key properties and tensions discussed above. In each experiment, 36 subjects each simultaneously sit at workstations and control the state of a single vertex in a 36-vertex network whose connectivity structure is determined exogenously and is unknown to the subjects. The state of a subject's vertex is simply one of 2 colors (red or blue), and can be asynchronously updated as often as desired during the 1-min experiment. Subjects are able to view the current color choices of their immediate neighbors in the network at all times but otherwise have no global information on the current state of the network (aside from a crude and relatively uninformative "progress bar"; see Fig. 1). No communication between subjects outside the experimental platform is permitted.

In each experiment, each subject is given a financial incentive that varies across the network, and specifies both individual preferences and the demand for collective unity. For instance, one player might be paid \$1.25 for blue consensus and \$0.75 for red consensus, whereas another might be paid \$0.50 for blue consensus and \$1.50 for red consensus, thus creating distinct and competing preferences across individuals. However, payments for an experiment are made only if (red or blue) global unanimity is reached, so subjects must balance their preference for higher payoffs with their desire for any payoff at all. A screenshot for a particular subject in a typical experiment is shown in Fig. 1. We note that our experiments may also be viewed as a distributed, networked version of the classic "Battle of the Sexes" game, or as a networked coordination game (5). Compared with the traditional analyses of these games, we are particularly interested in the effects arising as a result of the interactions of varying network structure and varying incentive schemes.

We note that although our experimental framework deliberately omits global "broadcast" mechanisms for consensus (other than the aforementioned progress bar) that are common in many public electoral processes—such as media polls, "mainstream" media reports and analyses—many other real-world sources of both small- and large-scale influence can be modeled via network structure. For instance, individuals whose opinion reaches an inordinately large number of others (such as might be expected of some political bloggers) can be modeled by high-degree vertices. Cohesive or close-knit groups of like-minded individuals can be modeled by subsets of vertices with similar incentives and dense connectivity. Our experiments deliberately introduce such structures and others. We also remark that our demand for complete unanimity before any payoffs are made is an abstraction of most real decision-making and voting processes, where a sufficiently strong consensus is typically enough to yield the benefits of unity. Although we expect most of our findings would be robust to such weakening, we leave its investigation to future research.

The experiments described here are part of an extensive and continuing series that have been conducted at the University of Pennsylvania since 2005, in which collective problem-solving from only local interactions in a network has been studied on a wide range of tasks, including graph coloring (6), trading of virtual goods (7), and several other problems. An overarching goal of this line of research is to establish the ways in which network structure and task type and difficulty interact to influence individual and collective behavior and performance.

Author contributions: M.K., S.J., J.T., and J.W. designed research; M.K., S.J., J.T., and J.W. performed research; M.K., S.J., J.T., and J.W. contributed new reagents/analytic tools; M.K., S.J., J.T., and J.W. analyzed data; and M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: mkearns@cis.upenn.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808147106/DCSupplemental.

ECONOMIC SCIENCES

COMPUTER SCIENCES

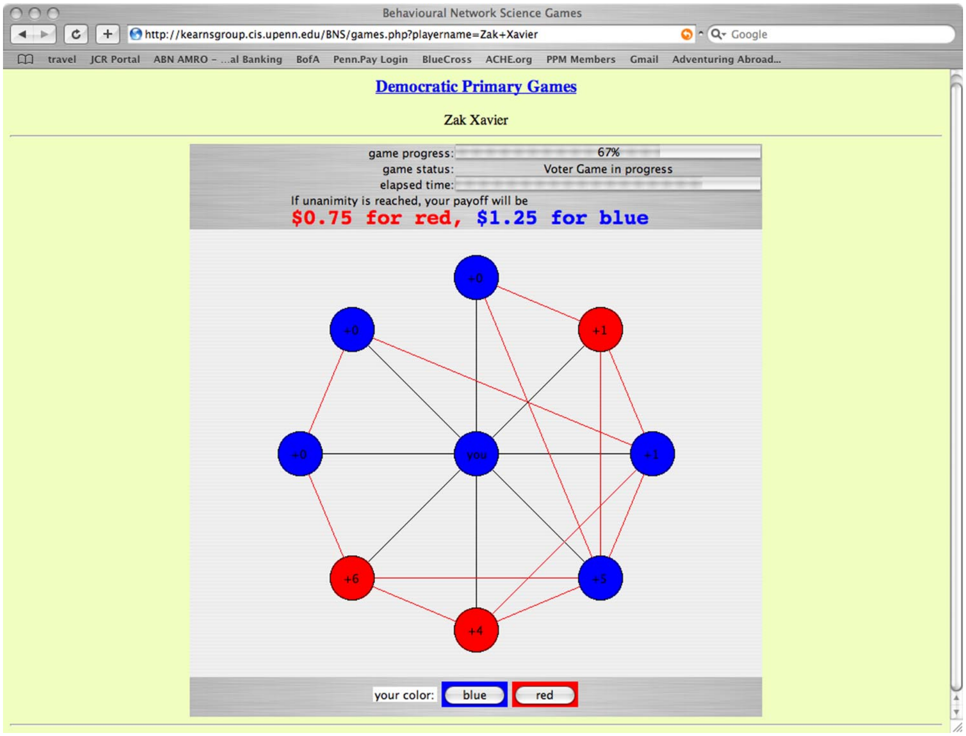

Fig. 1. Screenshot of the user interface for a typical experiment. Each subject sees only a local ("ego network") view of the global 36-vertex network, showing their own vertex at the center and their immediate neighbors surrounding. Edges between connected neighbors are also shown, as are integers denoting how many unseen neighbors each neighbor has. Vertex colors are the current color choices of the corresponding subjects, which can be changed at any time using the buttons at the bottom. The subject's payoffs for the experiment are shown (in this case $0.75 for global red consensus, $1.25 for blue), and simple bars show the elapsed time in the experiment and the "game progress," a simple global quantity measuring the fraction of edges in the network with the same color on each end. This progress bar is primarily intended to make subjects aware that there is activity elsewhere in the network to promote attention, and is uninformative regarding the current majority choice.

Experimental Design

There are two main design variables underlying our experiments: the connectivity structure of the underlying network and the financial incentives and their placement in the network. In each experiment, the network structure and the incentives were chosen in a coordinated fashion to examine specific scenarios or hypotheses. We now describe these choices and hypotheses in greater detail.

The 81 experiments fell into 2 broad categories that we call the Cohesion experiments (54 experiments) and the Minority Power experiments (27 experiments), named for the phenomena they were designed to investigate. All of the networks used had 36 vertices and nearly identical edge counts (101 ± 1), thus fixing edge density; only the arrangement of connectivity varied, and not the amount.

In the Cohesion experiments [named in part for a particular measure of inter- and intra-group connectivity (8)], vertices were divided into 2 groups of 18. Vertices in one group (the "red" group) were given incentives paying more for a red global consensus, whereas vertices in the other group (the "blue" group) were given incentives paying more for a blue global consensus. The relative strengths of these incentives were varied, as were the amount and nature of the connectivity within and between the two groups. In particular, we varied whether the typical vertex had more or fewer inter-group than intra-group edges, thus controlling whether local neighborhoods were comprised primarily of individuals with aligned incentives (high cohesion), competing incentives (low cohesion), or approximately balanced incentives. We also varied the nature of this connectivity; half of the Cohesion experiments used networks whose edges were generated (subject to the inter/intra group constraints) by a random or Erdos–Renyi process (9) (in which all edges are chosen randomly and independently with some fixed probability), the other half by the preferential attachment process (10) (which is known to generate the oft-observed power law distribution of connectivity). These two network formation models are well-studied and together provide significant variation over a number of common structural properties, including network diameter, degree distribution, and clustering. See Fig. 2 and the SI for further details.

The overarching goal of the Cohesion experiments was to systematically investigate how collective and individual performance and behavior varied with neighborhood diversity and the strength of preferences. Although it is perhaps most natural to hypothesize that increased inter-group connectivity should improve collective performance—this would be consistent with several mathematical network theories and metrics, including the aforementioned cohesion, and notions of expansion from the graph theory literature (9)—the degree of improvement, and how it might be influenced by the detailed structure (Erdos–Renyi vs. preferential attachment), the variability of individual human behavior, and so on, are difficult to predict.

In the Minority Power experiments, all networks were generated via preferential attachment (10). A minority of the vertices with the highest degrees (number of neighbors) were then assigned incentives preferring red global consensus to blue, whereas the remaining majority were assigned incentives preferring blue global consensus. The size of the chosen minority was varied (6, 9, or 14), as were the relative strengths of preferences. See Fig. 2 and the SI for further details.

The overarching goal of the Minority Power experiments was to systematically investigate the influence that a small but well-connected set of individuals could have on collective decision-making—in particular, to investigate whether such a group could reliably cause their preferred outcome to hold globally and unanimously.

For each of the different network structures in the Cohesion and Minority Power families, we ran experiments in which there were "strong symmetric," "weak symmetric," and "asymmetric" incentive structures. By "symmetric" we mean that the incentives of those players preferring blue and those preferring red were symmetrically opposed (such as $0.75/$1.25 for consensus to red/blue vs. $1.25/$0.75); by "weak" and "strong" we refer to the relative magnitudes of the preferred and non-preferred payments ($1.25 to $0.75 for weak, $1.50 to $0.50 for strong). In the asymmetric incentives experiments, the group preferring one color would be given strong incentives, whereas groups preferring the other color would be given weak incentives. We thus imposed scenarios in which 2

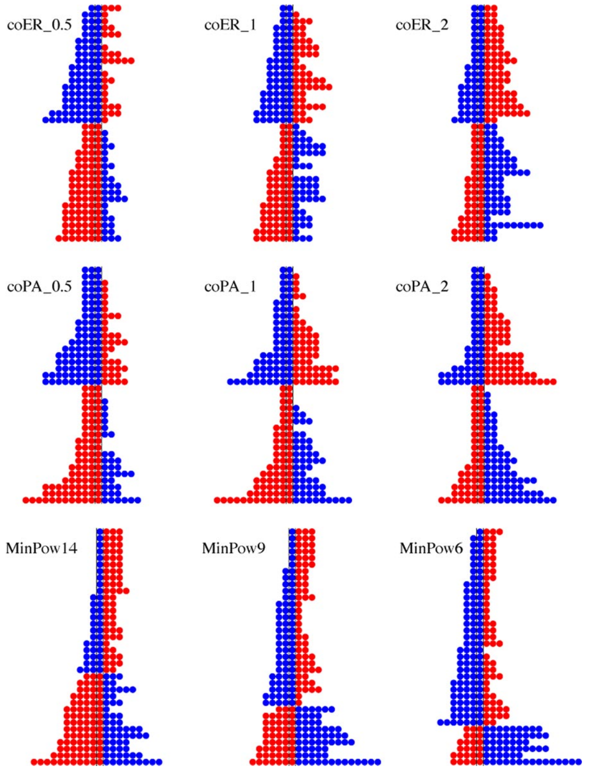

Fig. 2. Visualization of network and incentive structures. For each of the 9 network and incentive structures there is a diagram consisting of 36 rows of colored dots. Each row corresponds to a single subject or vertex in the network, and the dots in that row represent that subject and his or her network neighbors. The color of the central dot indicates the preferred (higher payoff) color for the corresponding subject, according to the incentives. The dots to the left of center indicate the number of neighboring subjects with the same preference; the dots to the right indicate the number with the opposite preference. Vertices are ordered within groups by their overall degree. (*Top*) Cohesion experiments with Erdos–Renyi connectivity in which there is more intra- than inter-group connectivity between the two groups (specifically a 1:2 inter:intra ratio) (*Left*); balanced connectivity (1:1 ratio) (*Center*); and more inter- than intra-group connectivity (2:1 inter:intra ratio) (*Right*). This is demonstrated by the migration of dots from left of center to right of center as we move from column 1 to 2 to 3. (*Middle*) Cohesion experiments with preferential attachment connectivity in the same inter:intra ratios as the coER row above. Comparison with the first row reveals clear differences in the overall degree distributions, because the variance in the total number of neighbors of subjects is much higher for preferential attachment and those diagrams reveal the presence of subjects with very large numbers of neighbors. (*Bottom*) Minority Power experiments, where again we see the heavy-tailed degree distributions typical of preferential attachment but in which now the blue-preferring vertices are selected to be a minority of varying sizes (14, 9, and 6) with the highest degrees. As discussed in the text, each of these 9 network structures was combined with payoff amounts that were weak symmetric, strong symmetric and asymmetric, yielding 27 distinct scenarios that were each executed in 3 trials, for a total of 81 experiments.

opposing groups "cared" equally but mildly about the global outcome, equally and strongly, or in which one group cared more than the other.

Thus, each of the 9 network structures was combined with weak symmetric, strong symmetric and asymmetric incentive schemes, yielding 27 distinct scenarios that were each executed in 3 trials, for a total of 81 experiments.

Human Subject Methodology

We now briefly remark on some further details of the experimental methodology and system. All experiments were held in a single session lasting several hours, and the participants were 36 University of Pennsylvania students enrolled in an undergraduate survey course on network science. Each of the 81 experiments had a fixed network and incentive structure, and the system assigned each of the 36 subjects randomly to one of the 36 network positions at the start of each experiment, thus assuring there was no systematic bias in the position of subjects in the networks. To prevent the establishment of social conventions that could trivialize the experiments (such as all subjects playing red for the remainder of the session following a successful global consensus to red), the system used a local randomization scheme on the colors, which might make what appeared red to one player appear blue to another. Each experiment had a 1-min limit for the population to reach a unanimous color choice; if they did so before then, the experiment ended and payments were tallied by the system. The session was closely proctored to ensure that no communication between subjects took place outside of the system, and physical partitions were erected around workstations to prevent inadvertent information leakage.

Results

Collective Behavior. Overall the subject population exhibited fairly strong collective performance. Of the 81 experiments, 55 ended in global consensus within 1 min (resulting in some payoff to all participants), with the mean completion time of the successful experiments being 43.9 s (standard deviation 9.6 s). We now proceed to describe more specific findings quantifying the impact of network structure, incentive schemes, and individual behavior.

Network structure influenced collective performance in a variety of notable ways. The Cohesion experiments were considerably harder for the subjects than the Minority Power experiments; only 31 of 54 of the former were solved compared with 24 of 27 of the latter (difference significant at $P < 0.001$). Furthermore, in all 24 of the successfully completed Minority Power experiments, the global consensus reached was in fact the preferred color of the well-connected minority. Together these results suggest that not only can an influentially positioned minority group reliably override the majority preference, but that such a group can in fact facilitate global unity.

Within the Cohesion experiments, generating connectivity according to preferential attachment (20/27 solved) yielded better collective performance than generating it via Erdos–Renyi (11/27 solved; difference significant at $P \approx 0.013$). When combined with the high success rate of the preferential attachment Minority Power experiments (the difference between the 44/54 solved instances of all preferential attachment networks and the 11/27 solved Erdos–Renyi networks is significant at $P < 0.001$), this finding indicates that, for this class of consensus problems, preferential attachment connectivity may generally be easier for subjects than Erdos–Renyi connectivity, an interesting contrast to problems of social differentiation such as graph coloring (6), where preferential attachment networks appear to create behavioral difficulties.

Independent of the method for generating connectivity, Cohesion performance improved systematically as within-group connectivity was replaced by between-group connectivity, with the strongest performance coming from Cohesion networks in which most subjects might have a preferred color different from those of a majority of their neighbors. Across all Cohesion experiments, the success rate on the networks with the highest level of inter-group connectivity (14/18 solved) and the success rate when connectivity was either mainly intra-group or balanced (17/36 solved) are significantly different ($P < 0.03$). Thus, increased awareness of the presence of opposing preferences improves social welfare. In terms of behavioral collective dynamics, it appears that this awareness leads to early "experimentation" with subjects' nonpreferred colors, resulting in more rapid mixing of the population choices.

Across all network structures, asymmetric incentives yielded the strongest collective performance (the overall asymmetric success rate of 22/27 differs from the combined weak/strong symmetric success rate of 33/54 at $P < 0.05$), and, indeed, the extremist's preferences were dominant, determining the consen-

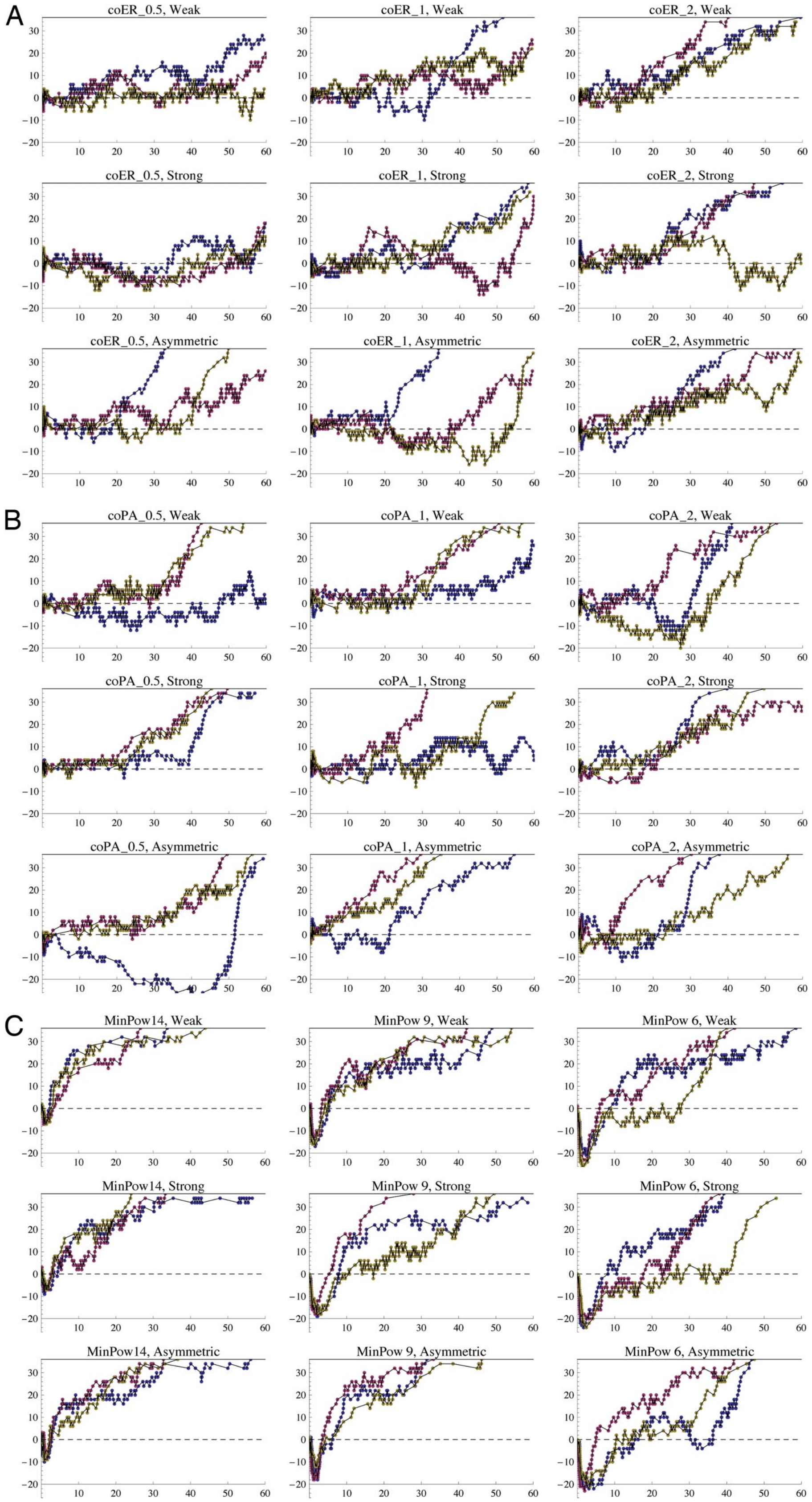

Fig. 3. Visualization of the collective dynamics for all 81 experiments. For each network and incentive structure there is a set of axes with 3 plots corresponding to the 3 trials of those structures. Each plot shows the number of players choosing the eventual collective consensus or majority color minus the number of players choosing the opposite color (*y* axis) at each moment of time in the experiment (*x* axis). All plots start at 0 before any color choices have been made; plots reaching a value of 36 within 60 s are those that succeeded in reaching unanimous consensus. Negative values indicate moments where the current majority color is the opposite of its eventual value. Plots are grouped by network structure first (Cohesion experiments with Erdos–Renyi connectivity in *A*; Cohesion experiments with preferential attachment connectivity in *B*; Minority Power experiments in *C*), and then labeled with details on the network and incentive structure. Within the Cohesion experiments, inter-group connectivity increases from left to right; within the Minority Power experiments, the minority size is decreasing from left to right. Several distinctive effects of network structure on the dynamics can be observed. Many Cohesion experiments spend a significant period ''wandering'' far from the eventual consensus solution. In contrast, Minority Power experiments invariably experience an initial rush into negative territory as the majority select their preferred color, but are then quickly influenced by the well-connected minority. Several instances of rather sudden convergence to the final color can also be seen, even after long periods of near-consensus to the opposite color (e.g., blue plot in lower left corner axes of *B* at ≈50 s). Fig. 4 below provides a visual summary of some of the qualitative effects of network structure on these dynamics.

sus outcome in 18 of the 22 successful asymmetric experiments. Strong symmetric incentives (14/27 successes) yielded worse performance than weak symmetric ones (19/27 successes). Thus, it appears most beneficial to have extremists present in a relatively indifferent population, and most harmful to have 2 opposing extremist groups.

The results on collective behavior described so far have focused on the final outcomes of experiments. The collective dynamics within individual experiments is also revealing, and shows notable effects of network structure. In Fig. 3 we provide visualizations of the collective dynamics in each of the 81 experiments, grouped by network structure and incentive scheme. As described in the caption, for each experiment there is a plot charting the progression of the number of players choosing the eventual consensus or majority color as a function of time within the experiment. Notable features include a ritual initial flurry of activity away from the minority preference in the Minority Power experiments, followed by an inevitable assertion of the minority influence over the population. There are also many instances in which a significant fraction of the experiment is spent quite far away from the eventual consensus choice, including near-total reversals of the collectively chosen color; see Fig. 3 and its caption for further details.

Although these visualizations of the dynamics are rich in detail, it is difficult to extract meaningful structural effects from them. In Fig. 4 we thus show the results of fitting simple 2-segment random walk models to the experimental dynamics within each family of experiments (fixed network and incentive structure). These models clearly show the effects of structure on collective dynamics: In terms of the rate of approach to the eventually favored color, Cohesion experiments with Erdos–Renyi connectivity tend to both begin and end slowly, whereas those with preferential attachment connectivity begin slowly but end more rapidly. Higher inter-group connectivity consistently increased late-game speed toward consensus. The Minority Power dynamics ended relatively fast, but early speed was heavily influenced by the size of their minorities.

Individual Behavior. It is natural to investigate the extent to which different human subjects exhibited distinct strategies or styles of play across the experimental session, and the degree to which such stylistic differences did or did not influence individual earnings. For any measure M of individual subject behavior within an experiment (such as the number of color changes made by the subject), we can compute the 36 average values for M obtained by taking the 81-game average for each subject, and compare these to the distribution of "random observer" averages, obtained by picking a random subject to observe in each experiment, and averaging the resulting 81 M values. Because subjects were in fact randomly assigned their network positions and incentives at the start of each experiment, if the variance of the 36 actual subject averages significantly exceeds that of the random observer distribution (according to a standard variance test), we can conclude that subjects exhibited meaningful (greater than chance) variation on measure M. See Fig. 5.

Most noteworthy is the fact that when the measure is wealth, subjects did not exhibit meaningful variation—thus the disparity in average or total wealth across the session (which ranged from \$46.50 total earnings to \$58.75, with a mean of \$52.76 and standard deviation of \$2.46) is already well-explained by the random assignment of subjects to positions. However, this finding in no way precludes the possibility that subjects still display distinct "personalities," nor that these differences might strongly correlate with final wealth. For instance, subject "stubbornness"—as measured by the amount of time a subject is playing their preferred color, but is the minority color in their neighborhood—varies meaningfully (Fig. 5) and is positively correlated with average wealth (correlation coefficient ≈ 0.43, $P < 0.01$). Being stubborn at the outset of an experiment (during the first 9 s) shows even stronger correlation with wealth (correlation coefficient ≈ 0.55, $P < 0.001$). The number of color changes made by subjects in the opening seconds of an experiment also varies significantly (Fig. 5) and is strongly negatively correlated with wealth (-0.58, $P < 0.001$). Together, these results suggest that stubborn and stable players set the tone of an experiment early.

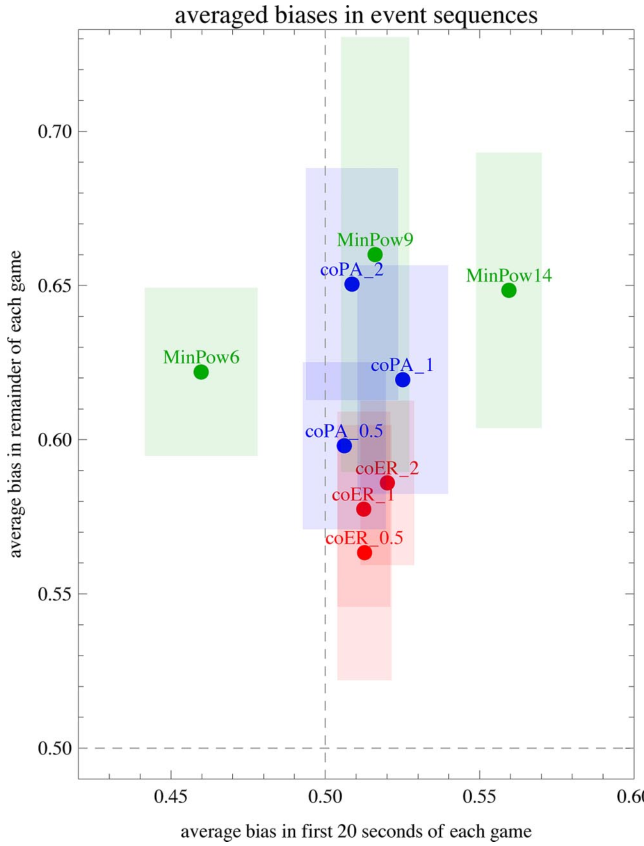

Fig. 4. Visualization of biased random-walk model fits to the dynamics of Fig. 3. For each of the 81 individual experimental plots in Fig. 3, we fit a 2-segment random walk model to the data—one segment for the first 20 s of the experiment, and one for the remainder of the experiment (similar findings result from different cut points between the two segments). Within each segment, we simply compute the fraction p of "upwards" moves (the number of moves toward the eventual majority color, divided by the total number of moves within the segment). This can be interpreted as modeling the collective dynamics by a random walk with probabilities p and $1 - p$ of upwards and downward moves, respectively; we refer to p as the bias of the model. Permitting independent bias values in the two segments allows us to separately model the dynamics in the early and later portions of each experiment. This yields a 2-parameter model for each of the 81 plots. Above we show the result of averaging over all incentive schemes and all repeated trials within the 9 families of network structures (Cohesion with Erdos–Renyi connectivity and 3 settings of inter- vs. intra-group connectivity; Cohesion with preferential attachment connectivity in 3 inter- vs. intra- settings; and Minority Power with 3 different minority group sizes). For each of these 9 families, we plot a point showing the average bias in the two segments, along with a shaded rectangle delimiting the standard deviation in both bias parameters for that family. The dashed lines show $p = 0.5$, where the model is unbiased (equal upward and downward probability). Several qualitative effects of network structure are apparent. For instance, Cohesion experiments tend to begin slowly (bias only slightly larger than 0.5), but preferential attachment connectivity leads to more rapid convergence in the later portion than does Erdos–Renyi connectivity. Increasing inter-group connectivity speeds the later dynamics regardless of the connectivity type. Minority Power experiments tend to conclude rapidly, but their early dynamics are strongly dependent on the minority size, with smaller minorities slowing the early progress toward the eventual majority choice. When the minority size is only 6, the first 20 seconds typically have a downward drift (bias $p < 0.5$).

Player stubbornness warrants further investigation, because it strikes at the heart of the tension that is a focal point of the experiments—by being stubborn, one might improve the chances of swaying the population toward one's preferred color, but one also risks preventing global consensus being reached in time (and thus forgoing any payoff). It is clear that no subject was infinitely

ECONOMIC SCIENCES

COMPUTER SCIENCES

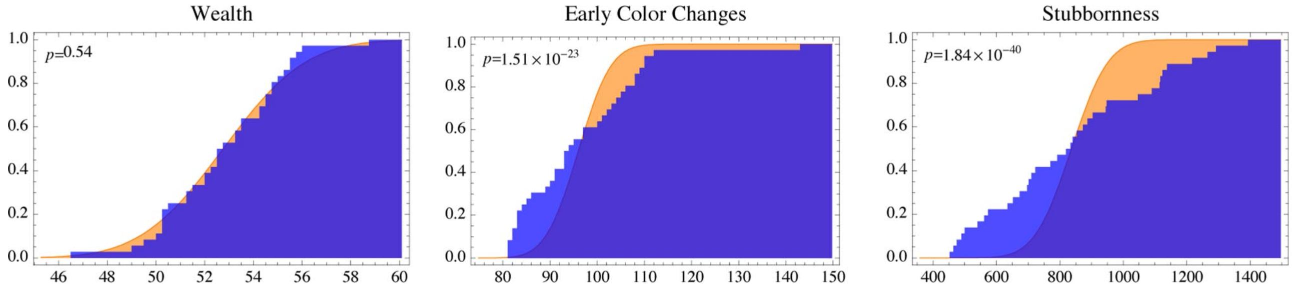

Fig. 5. Illustration of the ''random observer'' method for detecting meaningful variation in subject behavior. (*Left*) Empirical cumulative distribution function (CDF) of total player wealth (blue), in which wealth (*x* axis) is plotted against the fraction of the 36 subjects earning at least that amount (*y* axis). It is very well-modeled by the theoretical expected CDF generated by choosing a random player's wealth independently in each experiment (orange), so we may conclude that the variation in player wealth is explained by the random assignments to network position. In contrast, the CDFs of the number of color changes taken by each player in the first several seconds (*Middle*) and the total amount of ''stubborn'' time (*Right*) are poorly modeled by the random observer CDF, showing considerably greater variance in both cases. See text for details.

stubborn: The wealthiest player had their preferred color 28 times out of 55 successful games but acquiesced to group dynamics and accepted the lower payoff 27 times. All other players acquiesced more often—up to as many as 40 times out of 55. In the 26 games that failed to achieve unanimity, there were only 30 individual cases of players defying all of their neighbors as time expired, and only 5 games ended in failure due to players that defied all neighbors for more than the last 2 seconds of play. Only 3 individual players ever caused this kind of failure; one did it 3 times, but also acquiesced 38 out of 55 times and garnered relatively poor overall earnings. These facts combined with the aforementioned correlation of stubbornness with wealth suggest that successful players managed to be "tastefully" stubborn, and that overall behavior was quite acquiescent.

In addition to the raw experimental data, subjects were given an exit survey in which they were invited to comment on their own and others' strategies, and these surveys provide a rich and often consistent source of insight into individual styles of play. Twenty-four subjects explicitly mentioned starting off by choosing the color that would give them the higher payoff upon consensus. Twenty-seven subjects mentioned either trying to signal others, or noticing others trying to signal; however, many also found this behavior annoying and said that it did not help. Twenty-one subjects noticed others being irrationally stubborn, or expressed suspicion that others were being irrationally stubborn. (Here we use the term "stubborn" in the informal way it was given in the surveys, as opposed to the formal measure discussed above.) Three subjects mentioned being stubborn themselves because they did not want small payoffs. Seven subjects mentioned using different strategies depending on whether their incentives were weak (\$0.75 vs. \$1.25) or strong (\$0.50 vs. \$1.50). Three subjects mentioned changing their behavior as the night progressed, 1 subject developed a strategy, and 2 subjects simply became tired. We note that there is no evidence in the data of the collective performance improving or degrading significantly as the session progressed; for instance, plotting the accumulated collective wealth vs. the progression of experiments in the order they were conducted yields an almost perfectly linear curve.

Finally, 27 subjects mentioned following the action choices of their high degree neighbors and/or being more stubborn when they themselves had high degree. It is interesting to note that the average degree of subjects is much more weakly correlated with their wealth (0.38, $P \approx 0.09$) than the stubbornness and stability properties discussed above, despite these reports of conditioning behavior on degrees. There is no inherent contradiction here, because conditioning on degrees may appear primarily in the decision on how stubborn and stable to play.

Despite the observed and reported variations in individual subject strategies, it is interesting that one can approximately reproduce salient aspects of the collective behavior with rather simple and homogenous theoretical models of individual behavior. For example, consider a "multiplicative" model in which a player who is paid $w(c)$ for global convergence to color c, and a fraction $f(c)$ of whose neighbors are currently playing c, plays c in the next time step with probability proportional to $w(c)f(c)$ (11). Such agents combine their preferences (as given by the values $w(c)$) with the current trend in their neighborhoods (the $f(c)$) to stochastically select their next color in a natural manner. If such agents are simulated using the same networks and incentives as in the 81 human subject experiments, and the number of simulation steps is capped (as it effectively is by the 1-min time limit of the human experiments), there is rather strong correlation (0.60, $P < 0.001$) between subject and simulation times to consensus.

Discussion

A number of further investigations are suggested by the findings summarized here. In particular, the variations in individual behavior and the apparently helpful presence of "extremists" raise the question of whether certain mixtures of behaviors and attitudes are required for optimal collective problem-solving. It would also be interesting to use the data from our experiments to develop richer statistical models of individual and population behavior, whose predictions in turn could be tested on further behavioral experiments.

1. Zeleny J (June 28, 2008) Working together, Obama and Clinton try to show unity. *NY Times*. Available at www.nytimes.com/2008/06/28/us/politics/28unity.html.
2. Kleinberg J (2007) Cascading behavior in networks: Algorithmic and economic issues, *Algorithmic Game Theory*, eds Nisan N, Roughgarden T, Tardos E, Vazirani V (Cambridge Univ Press, Cambridge, UK), pp 613–632.
3. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83:1420–1443.
4. Schelling T (1978) *Micromotives and Macrobehavior* (Norton, New York).
5. Luce RD, Raiffa H (1957) *Games and Decisions*: Introduction and critical survey (Wiley, New York).
6. Kearns M, Suri S, Montfort N (2006) An experimental study of the coloring problem on human subject networks. *Science* 313:824–827.
7. Kearns M, Judd S (2008) Behavioral experiments in networked trade. *Proceedings of the 2008 ACM Conference on Electronic Commerce*, eds Sandholm T, Riedl J, Fortnow L (Assoc Computing Machinery, New York), pp 150–159.
8. Morris S (2000) Contagion. *Rev Econ Studies* 6757–78.
9. Bollabass B (2001) *Random Graphs* (Cambridge Univ Press, Cambridge, UK).
10. Barabasi A, Albert R (1999) Emergence of scaling in random networks. *Science*, 286:509–512.
11. Kearns M, Wortman J (2008) Learning from collective behavior, *Proceedings of the 21st Conference on Learning Theory*, eds Servedio R, Zhang T (Omnipress, Madison, WI), pp 99–110.