Group Problem Solving Behavior in a Networked Puzzle Game

Thomas N. Wisdom (tnwisdom@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th Street
Bloomington, IN 47404 USA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th Street
Bloomington, IN 47404 USA

Abstract

In this study we examined the influence on individuals of choices made by fellow participants within a group. Using a simple networked puzzle game, we studied the effects of participant group size, availability of neighbor guess information, and problem difficulty. Participants attempted to guess a hidden computer-generated picture on a small square grid game board over a series of rounds with score feedback. The guesses of other players were visible to each participant and could be copied at will. We found a significant relationship between the number of neighbor guesses participants could view and the tendency for participants to copy the solutions of others, and a decreasing tendency to imitate over rounds of a game, along with differences in score and other performance measures between isolated and grouped participants.

Keywords: Social learning; group behavior; information diffusion; innovation diffusion.

Background

Human decision-making research is increasingly concerned with the fact that many real-life decisions are made in the presence of information about similar decisions made by others, and that "social learning" using this information can strongly affect the outcomes of such decisions. While ignoring socially-mediated information in a laboratory setting may yield a clear study of important cognitive processes of individual problem solving and choice, in many situations it represents an unrealistic constraint, and may limit the relevance or applicability of the conclusions. There are important behavioral differences between people with access to social information and those who are isolated; understanding these differences is vital to the study of human decision making and problem solving.

One review of the group performance literature concludes that groups generally do not outperform their best members, although they may perform better than individuals on average (Kerr et al., 1996). More recently, modeling work by Hong and Page (2004) concludes that groups of high-ability problem solvers can be outperformed by more diverse groups of problem solvers. And though "conformity" often carries negative connotations in lay usage, other research has shown that for uncertain environments, "conformity bias" (the tendency to imitate the behavior of others instead of exploring solutions individually) can be adaptive (Boyd & Richerson, 1985; Kameda & Nakanishi, 2002).

Two processes work together to produce this benefit: innovation (generating locally novel solutions to problems) produces a diversity of solutions, and imitation (copying of solutions by others) diffuses these innovations. By imitating, a decision maker can employ solutions discovered and passed on by others without having to develop them independently using costly trial-and-error learning. When a decision maker acquires information from others about possible solutions to a problem, the resources not expended in information gathering can be used for other aspects of problem solving, thus improving performance overall. Domination of a system by either one of these processes can be problematic. An excess of innovation may be inefficient because good ideas produced by individuals are not built upon by others, and an excess of imitation may be maladaptive because suboptimal solutions that are nonetheless better than other solutions by group members are spread to the exclusion of better alternatives that are left unexplored.

Research points to several factors that can modulate the choice between innovation and imitation. Some animals are able to use the observed success (or lack thereof) of conspecifics to guide imitative foraging choices (Templeton & Giraldeau, 1996). When humans process social information, evidence of success is often delayed or distorted; to alleviate this uncertainty, other cues are often used to determine the utility of others' choices. However, this can introduce complexities of trust and observation, and thus influence behavior (Bøg, 2006).

A common characteristic of much of the human research in this area (e.g. Asch, 1951) is that some of the participants are confederates or simulated agents. This is a somewhat useful approach in that it addresses in a simple manner the influence of outside information on individual decision-making; however, it does not explore the group-level processes that emerge when individuals are all influencing each other. For instance, the topology of the social network that connects a group can substantially affect the diffusion of information within it (Noble et al., 2004), which can in turn affect how adaptive social learning will be for the members of the group. Bavelas (1950) and Leavitt (1951) examined the effects of various network structures on a group coordination task and found that the topology of the network and the complexity of the task strongly influenced individual participants' behaviors, including the types of information sent by participants and the tendency for those

in certain positions in a network to spontaneously act as "leaders" in the group. Using computer simulations, Lazer and Friedman (2005) showed that more locally connected networks (those which disseminate information relatively slowly) are better than more highly connected networks at exploring a complex problem space with many local maxima in the long run, though not in the short term. Mason, Jones, and Goldstone (2005) examined these phenomena in a study of humans exploring a one-dimensional solution space and found that the network topology and the nature of the problem space interacted in influencing how well the group performed as a whole. Consistent with Lazer and Friedman's simulations, they found that for more complex problems with multiple local maxima, networks that had more restrictive local connectivity performed better than fully connected networks.

We wished to extend the study of these phenomena to a more complex problem space and explore further contextual manipulations. Toward this end, we designed an experiment in which networked groups of people explored a multidimensional problem space and passively shared information. Each of the participants in the experimental game was given the task of drawing a picture that matched as well as possible a goal picture randomly generated by the computer. They were instructed to do this by making changes in their game boards over a series of time-limited rounds and taking into account periodic score feedback to direct subsequent board modifications. Participants could observe and copy the most recent candidate solutions of their neighbors as they explored. The form of interaction was intentionally impoverished because we wished to study both individual exploration and solution imitation behaviors, and an important precursor to the latter is determining whether information provided by others is considered useful.

To explore the effects of network efficiency, we manipulated the number of neighbors available to each participant in a uniform network structure. The apparent benefit of imitation was modulated by a manipulation of the availability of evaluative (score) information about the solutions of others. We also manipulated problem difficulty (by manipulating the size of the problem space) in order to observe the effect of differences in uncertainty on imitation behavior (conformity bias). We limited the time available to solve the problem, thus forcing participants to allocate their time and effort between innovation and imitation. There was no allowance for explicit communication, network adaptation, organized cooperation, or defection. These restrictions limited the possibilities of interaction but also provided a usefully idealized situation for probing simple innovation dissemination in social networks.

As this study is fairly novel, it is primarily exploratory; nevertheless, we made the following tentative predictions. The average scores of participants in large groups (greater than 4 participants) were expected to be greater than those in smaller groups or individuals, because the larger groups would produce a greater diversity of solutions, and thus be able to search the problem space more efficiently. Imitation was expected to be more common when scores accompanied neighbor guess information, because the utility of each neighbor's guess would be explicit and thus imitators could be more confident in their actions. Average turnover was expected to be higher when scores were visible for the same reason, and thus positively correlated with imitation behaviors. Score was expected to increase over rounds within each game due to score feedback and simple learning effects. For the same reason, the increase was expected to be greater for later conditions (order greater than 4) than for earlier conditions. Conversely, solution turnover (the amount of game board modification between rounds) would decrease within each condition as participants found more parts of the correct solution and narrowed their search to smaller areas of the board.

Method

Participants were recruited from the Indiana University Psychology Department undergraduate subject pool, and were given course credit for taking part in the study. Participants populated each session by signing up at will for scheduled experiments with a maximum capacity of 10 persons. 30 sessions (containing a total of 98 participants) were run. 6 sessions (containing a total of 18 participants) were discarded due to network or software problems, and one single-participant session was discarded due to inattention. The remaining 79 participants were distributed as shown in Table 1.

Table 1: Distribution of Participants Across Group Sizes

Group Size	1	2	3	4	5	6	7	8
Participants	7	6	12	8	5	12	21	8

We implemented the experiment using custom software run in a web browser. The task was a round-based picture-matching puzzle game with score feedback given after each round. The goal picture that participants attempted to match was a randomly-generated spline quantized to a grid of square pixels. The squares making up the spline were colored black, and the remaining squares colored white (see Figure 1 for examples).
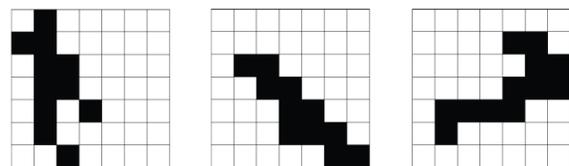

Figure 1: Examples of randomly-generated goal pictures in experimental task

The participant game board was a grid of the same dimensions as the goal picture, with each square initially colored white. The color of each square on the game board

could be toggled between black and white by clicking it with the mouse. Each participant's display included their own game board and most recent score (given as the number of squares (both black and white) marked correctly out of the total number of squares on the board), their neighbors' game boards and (in the "Score Visible" condition) scores, and indications of the current round in the game and the amount of time remaining in the current round (see Figure 2). Each game consisted of 12 rounds of 20 seconds each. After the last round in each game, participants were shown their guesses and scores for each round, along with the goal picture, and a button to click when ready to begin the next condition. When all participants had clicked this button, the next condition began.

Participants were informed that the picture they were trying to match was randomly generated and not representative of any particular object, shape, or symbol, and was generally not symmetrical; that the black squares were all connected to each other vertically, horizontally, or diagonally; and that the black part of the picture was small relative to the size of the grid.
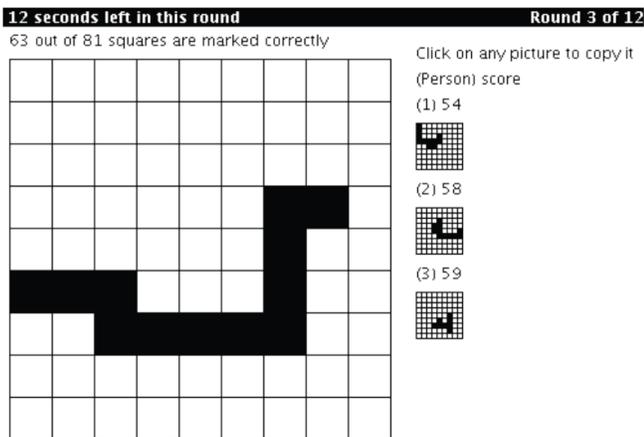
Figure 2: Screenshot of participant display

After instruction, the participants ran the first condition, after which the experimenter confirmed that each of the participants understood the mechanics of the game, and answered any questions that arose. The participants then played the remainder of the conditions without further interruption. Condition order was randomized within each session.

We defined a participant's score for a particular round as a cell-by-cell comparison (overlap) between the participant's guess for that round and the hidden goal picture (i.e. the number of cells which the two pictures have in common), divided by the total number of squares in the goal picture, to give a percentage which can be compared between conditions of varying grid size (see Figure 3).

Turnover for a particular round (greater than 1) was a measure of the amount of change in a participant's guesses between rounds. It was defined similarly to score, except

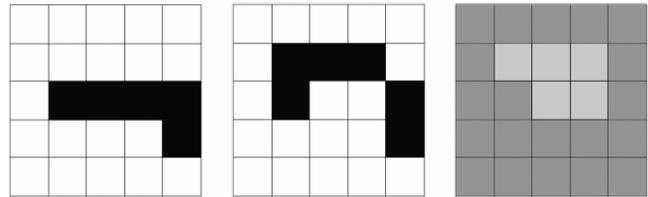that the two pictures compared were the participant's guesses from the current round and the previous round.

Figure 3. Score / overlap calculation: the first two pictures have 20 out of 25 squares in common (shown in dark grey on the right), so they have an overlap of 80%.

Imitation (a measure of whether a participant copied a neighbor's guess in a particular round) was defined as follows:

$$I_{pr} = \begin{cases} 0 : overlap(G_{pr}, G_{p(r-1)}) > \max_i(overlap(G_{pr}, G_{n_i(r-1)})) \\ 1 : overlap(G_{pr}, G_{p(r-1)}) \le \max_i(overlap(G_{pr}, G_{n_i(r-1)})) \end{cases}$$

Where G_{pr} is the guess of participant p for round r, $G_{ni(r-1)}$ is the guess of neighbor i for the round prior to round r, and *overlap* is the comparison described above for the score calculation. In other words, imitation has occurred (I_{pr}=1) if a participant's guess is closer to the most similar neighbor's previous guess than to the participant's own previous guess. Three factors with two levels each were manipulated, giving a total of 8 within-subject conditions per session. Score visibility – the scores of each participant's neighbors were either shown next to their game boards in the participant's display, or not. Network type – the set of neighbor connections between participants was either a simple lattice (each participant was connected to only two immediate neighbors) or a fully-connected network (each participant was connected to all other participants). Board dimension – the lengths of the sides of the square game board were either 7 squares or 9 squares. (with the larger board assumed to be more difficult). The size of the group participating in each session was treated as a covariate. Another factor considered was the (randomized) position of each condition in an experiment session; this was called the order.

Results

In the aggregate, participants achieved final scores of 78%, and average scores (over all rounds) of 74.6%. The average guess turnover rate per round was 13.7%, and participants engaged in imitation on 28.8% of their guesses. For the initial analysis, dependent variables were averaged across all participants and all rounds to give measures for the group's aggregate activity in each condition. These data were analyzed using a repeated-measures ANOVA, treating each group as a single subject.

Since the two network type conditions are indistinguishable for group sizes less than 4, these sessions were discarded for analyses of the effect of network type.

Similarly, sessions with only 1 participant were discarded for analyses of imitation, since isolated participants have no neighbors to imitate.

The manipulated factors appear to have had little effect on the behavior of the participants. The only significant effect observed in the ANOVA was a complicated interaction between network type and group size in which imitation was less common for the fully-connected network for groups of 4 participants, approximately equal for the two network types for groups of 5 participants, and increased for the fully-connected network with larger group sizes up to 8. This interaction suggested a simpler regression analysis using a combination of the above two factors: the number of neighbors (NN) visible to each participant in a given condition. This combination is shown in Table 20. (Note that all groups of 3 participants, and groups of 4 participants or more in the Lattice Network condition are included in the NN=2 group.)

Table 2: Combination of group size and network type to derive number of neighbors (NN)

		Group Size							
		1	2	3	4	5	6	7	8
Network Type	Lattice	0	1	2	2	2	2	2	2
	Full				3	4	5	6	7

Curve fit estimation for a linear regression of imitation against the number of visible neighbors yielded a significant positive relationship ($\beta=0.794$, F=8.501, df=6, p<.05) (see Figure 4). Interestingly, a similar analysis of final scores in each condition did not yield a significant linear relationship, but a significant quadratic relationship was apparent which peaked at NN=4 ($\beta_1=3.195$, $\beta_2=-3.259$, F=10.346, df=7, p<.05) (see Figure 5).

The same data were averaged across all participants in each game to give dependent variable measures for each round and all values of game order and group size. Linear regressions were run on the round data for each dependent variable. Analysis of score vs. round showed a slightly positive but nonsignificant relationship for sessions with one participant, and a strongly significant result for all other group sizes (see Figure 6): in the aggregate, scores in groups with between 2 and 8 participants showed a positive linear relationship ($\beta=0.966$, F=137.525, df=11, p<.001). Similarly, a slightly negative but nonsignificant relationship was found for turnover vs. round for sessions with a single participant, and a strongly significant trend for all other group sizes (see Figure 7): turnover rates in groups of between 2 and 8 participants showed a strongly positive linear relationship to round number ($\beta=-0.979$, F=210.612, df=10, p<.001). These results seem to indicate a general difference in guessing behavior between isolated individuals and groups. A significant negative trend was found for imitation rate vs. round (see Figure 8): participants tended to imitate less as each game progressed ($\beta=-0.884$, F=32.805, df=10, p<.001).

In addition, regression analysis of score vs. round showed significant positive relationships (all with p<.001; mean $\beta=0.897$) for all game order positions except the first. (It is worthwhile to note, however, that no significant trend developed over the remainder of the conditions – that is, after the first condition, participants did not get better or worse, imitate more or less, etc. over the course of the session.)
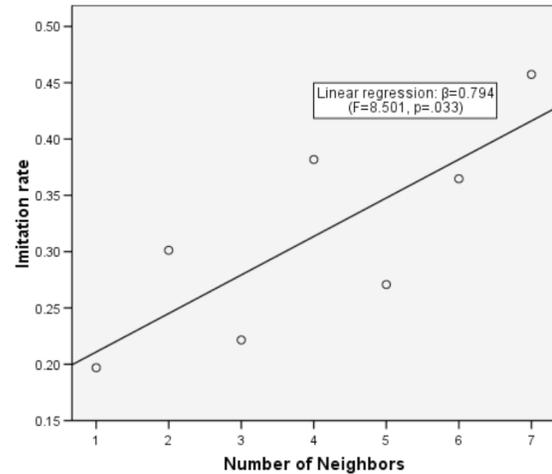

Figure 4. Linear regression of imitation vs. the number of visible neighbors (NN)
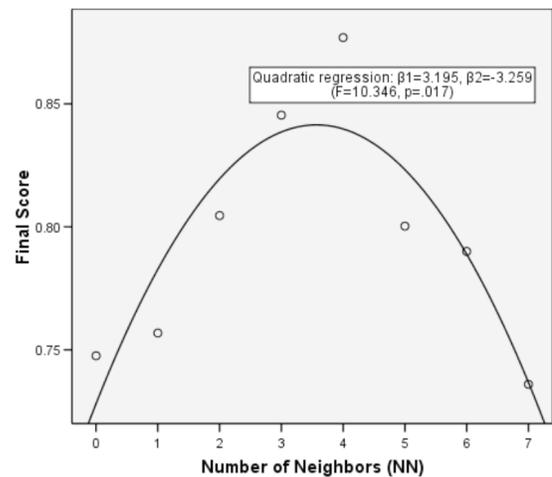

Figure 5. Quadratic regression of final score vs. the number of visible neighbors
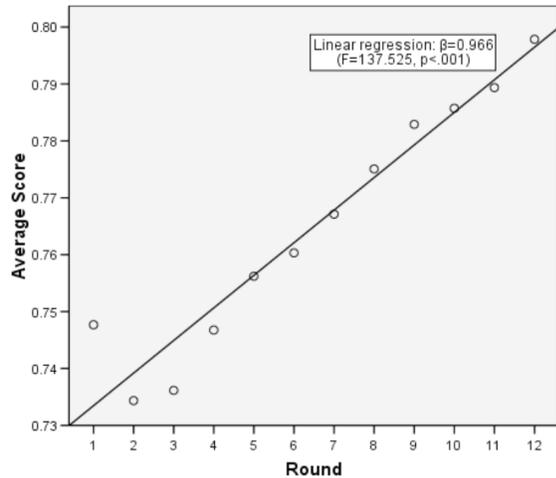
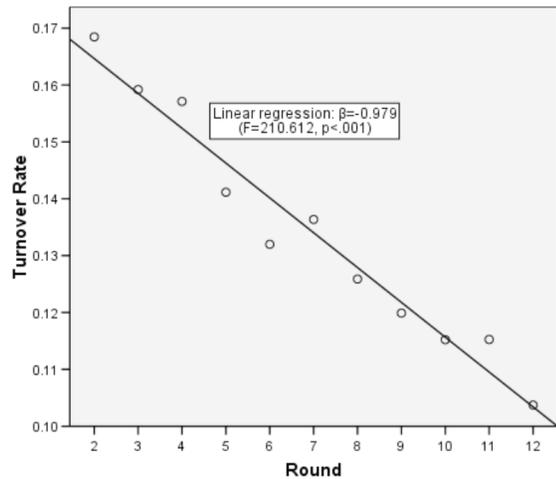Figure 6. Linear regression of score vs. round for group sizes greater than 1

Figure 7. Linear regression of turnover rate vs. round for group sizes greater than 1
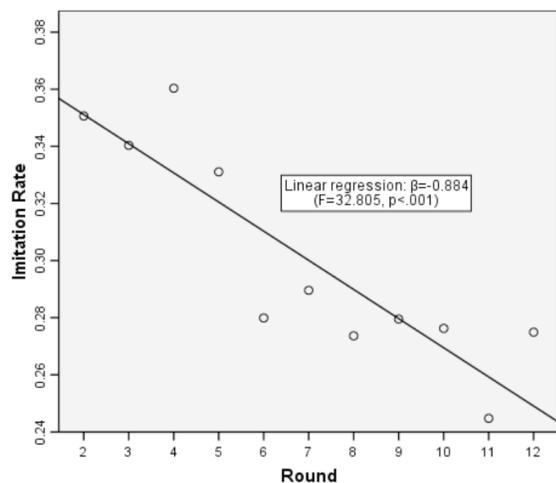

Figure 8. Linear regression of imitation rate vs. round

Discussion

The specific predictions made for the effects of the experimental manipulations were generally unsupported by the results. Most surprisingly, the rate of imitation was unaffected by the availability of score information, which seems to indicate that at least in this paradigm, this information was not perceived as useful. Likewise, score visibility did not affect turnover rate, though turnover was correlated with imitation. We observed linear increases in score and decreases in turnover over rounds, which were predicted as consequences of the progressive discovery of partial solutions and a corresponding narrowing of solution space exploration, respectively. Rates of score increase did not change significantly over the course of a session as was predicted.

The linear decrease in imitation over the course of the game (i.e. a tendency for participants to copy others' guesses less often in later rounds) has several possible interpretations. It may be that confidence in one's own personally acquired information grows with experience at working toward a particular solution, or that the effort expended in developing a solution becomes a sunk cost that participants are unwilling to write off by copying and switching to a different solution. Though competition was not specifically encouraged, it may be that participants realized that copying another solution toward the end of the game would result only in a tie with its source, and that they had to personally find a better solution to beat their fellow participants. If present, this effect is probably exacerbated by a decrease in the diversity of the solutions available for copying as participants approach the solution. These questions merit further study in future experiments, as they relate directly to questions of conformity bias and the tradeoff between imitation and innovation.

The regression results for score and turnover in relation to group size admit a simple explanation: isolated participants show less consistent changes in their score and less consistent variability in their guesses due to the fact that there are no other participants to observe or imitate. The lack of other participants limits the diversity of approaches to solving the problem to those generated internally by the participant. The regression results for score vs. round in relation to game order indicate that the first game in each session may have functioned as a practice game, during which participants were still learning the mechanics of the experiment. This suggests that in future versions, an explicit practice game should be included in order to obtain more consistent data.

The positive relationship between imitation rate and NN may indicate a tendency for imitation behavior to scale according to the number or diversity of examples available. However, there may be a simpler explanation: the more participants there are, the greater the chance that two of them will randomly resemble each other at any given time. That is, this may be an artifact of the choice of problem space and the imitation calculation.

Finally, the quadratic relationship between score and NN appears anomalous and possibly spurious: it seems reasonable to expect that scores would show a steady improvement as groups grew in size, rather than peak at an intermediate level. At the risk of over-interpretation of a limited data set, a possible answer is suggested by the concept of "span of managerial responsibility" (Drucker, 1954), which proposes that the optimal size for a typical subordinate unit to operate with minimal management control is between 3 and 6 people. This numerical rule of thumb survives in modern industrial and military practices (Department of the Army, 2003). In this experiment, scores peaked at an effective group size of 5; it may be that groups of this size allow particularly efficient information sharing and decision-making for a problem of the complexity encountered here. This is highly speculative, but it will be interesting to see if similar effects arise in future studies.

In this study we found that the behavior of isolated individuals attempting to solve a multidimensional problem differs markedly from that of people connected in groups, and that differences in the size of a group can have significant and sometimes unexpected effects on behavior. These results present several intriguing areas for further study. Specifically, having larger groups of participants would allow exploration of the possible scaling effects for group size seen here, as well as the implementation of a wider variety of network topologies. The use of alternate methods for indicating to participants the potential utility of neighbor guesses, as well as determining when imitation has occurred, would allow a closer examination of the circumstances that encourage conformity bias and similar phenomena. Overall, there are strong implications for testing the predictions of past work in the area of group problem-solving behavior, and potential applications to many real-world problems.

Acknowledgments

References

Asch, S. E. (1951). Effects of Group Pressure Upon the Modification and Distortion of Judgements. In H. Guetzkow (ed.), *Groups, Leadership, and Men*. Pittsburgh, PA: Carnegie Press.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, *22*, 725-730.

Bøg, M. (2006). Whom Should I Observe? – Learning through Observation and Experimentation in Simple Bandit Problems. Manuscript submitted for publication (copy on file with author).

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Drucker, P.F. (1954). *The Practice of Management*. New York: Harper & Brothers.

Department of the Army (2003). *Field Manual 6-0: Command and Control of Army Forces*. Washington, DC: U.S. Government Printing Office.

Hong, L., & Page, S.E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *PNAS*, *101*, 16385-16389.

Kameda, T., & Nakanishi, D. (2002). Cost–benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evolution and Human Behavior*, *23*, 373–393.

Kerr, N.L., MacCoun, R.J., & Kramer, G.P. (1996). Bias in Judgment: Comparing Individuals and Groups. *Psychological Review*, *103*, 4, 687-719

Lazer, D. & Friedman, A. 2005. The hare and the tortoise: the network structure of exploration and exploitation. *Proceedings of the 2005 National Conference on Digital Government Research*. ACM International Conference Proceeding Series, *89*. Digital Government Research Center.

Leavitt, H. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, *46*, 38-50.

Mason, W.A., Jones, A., & Goldstone, R.L. (2005). Propagation of innovations in networked groups. *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Noble, J., Davy, S., & Franks, D.W. (2004). Effects of the topology of social networks on information transmission. *From Animals to Animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior* (pp. 395-404). Cambridge, MA: MIT Press / Bradford Books.

Templeton, J.J., & Giraldeau, L.-A. (1996). Vicarious sampling: the use of personal and public information by starlings foraging in a simple patchy environment. *Behavioral Ecology and Sociobiology*, *38*, 105-113.