

Running head: Collective Search

Collective Search in Concrete and Abstract Spaces

Robert L. Goldstone

Michael E. Roberts

Winter Mason

Todd Gureckis

Indiana University

Corresponding Address: Psychology Building  
Indiana University  
Bloomington, IN. 47401  
[rgoldsto@indiana.edu](mailto:rgoldsto@indiana.edu)

© 2007 Robert L. Goldstone, Michael E. Roberts, Winter Mason, and Todd Gureckis

Do NOT Quote Without Permission

To appear as: Goldstone, R. L., Roberts, M. E., Mason, W., & Gureckis, T. (in press).  
Collective search in concrete and abstract spaces. In T. Kugler, C. Smith, and T.  
Connelly (Eds.) Decision modeling and behavior in uncertain and complex  
environments. Springer Press.

## Abstract

Our laboratory has been studying the emergence of collective search behavior from a complex systems perspective. We have developed an internet-based experimental platform that allows groups of people to interact with each other in real time on networked computers. The experiments implement virtual environments where participants can see the moment-to-moment actions of their peers and immediately respond to their environment. Agent-based computational models are used as accounts of the experimental results. We describe two paradigms for collective search – one in physical space and the other in an abstract problem space. The physical search situation concerns competitive foraging for resources by individuals inhabiting an environment consisting largely of other individuals foraging for the same resources. The abstract search concerns the dissemination of innovations in social networks. Across both scenarios, the group-level behavior that emerges reveals influences of exploration and exploitation, bandwagon effects, population waves, and compromises between individuals using their own information and information obtained from their peers.

### Collective Search in Concrete and Abstract Spaces

The purpose of much of human cognition can be characterized as solving search problems. Concrete examples of important search problems facing us are finding food, friends, a mate, shelter, and parking places. More abstract examples of search are scouring the web for valuable data, finding a scientific research area that is novel yet impactful, and finding sequences of moves that will allow us to play an effective game of chess. In the concrete cases, the sought-after resources are distributed in an actual physical space, and people have to move through this space to sample locations for the desired resource. In the abstract cases, if spaces exist at all, they are formal constructions, they will often times have more than three dimensions, and they may not operate under standard Euclidean metric assumptions. Despite these differences between abstract and concrete search problems, there is a growing sentiment that they may share fundamental cognitive operations (Hills, 2006; Todd & Gigerenzer, 2000). Pursuing this premise, we will juxtapose concrete and abstract search paradigms to reveal some of this shared cognitive underpinning.

The notion that cognition centrally involves search is hardly new. It was the motivation for Newell and Simon's (1972) classic work on solving problems by searching for sequences of operators that would take an agent from an initial to a goal state. However, in this work, and the considerable body of subsequent work in cognitive science that it inspired, the focus has been on individuals solving problems on their own or in teams. Our work focuses on individuals searching for solutions in an environment that consists of other individuals also searching. The motivation for this focus is that individuals rarely solve important problems in isolation from one another, in controlled laboratory cubicles. For example, one can think of the continued advancement of science and technology as a massive, real-world collective search problem. While we might view the individual scientist as a single solution-searching unit, in fact each scientist's work is influenced by their own discoveries and the successes, failures, and current efforts of others. Indeed, the presence of peers with similar motivations fundamentally changes the search process for any individual. Depending upon the circumstances, individuals will be either attracted to or repelled by the solutions of others. Attraction occurs when the cost of exploring a problem space on one's own is high (Boyd & Richerson, 1985) and when others can act as informative scouts for assessing the quality of solutions that would be costly to personally gauge. Repulsion occurs when competition for resources is keen and early consumers of a resource can preempt the effective consumption of the resource by

subsequent individuals. In a world that consists of many individuals competitively searching their environment for resources, it is not enough to employ efficient, independent search strategies. Individuals also need to know when to modify their own search to follow others' leads or, conversely, to avoid the crowd. One's peers are a big part of one's environment.

The importance of one's peers does not stop there, because peers change the rest of the environment as well. When limited resources -- like bananas, but not like web sites -- are used, they are no longer available for others. In these cases, one individual's consumption negatively affects the resources available for the rest of the group. However, the influences that individuals have on the environment need not only be competitive. For web sites, use may facilitate subsequent use, as when early users' preferences are used to make helpful suggestions for subsequent users (Clark, 2003). A generalization of this notion of affecting peers by affecting their environment is the notion of "Stigmergy." Stigmergy is a form of indirect communication between agents that is achieved by agents modifying their environment and also responding to these modifications (Dorigo, Bonabeau, & Theraulaz, 2000). This effect has been well documented in ant swarms, in which ants lay down pheromones as they walk that attract subsequent ants (Theraulaz & Bonabeau, 1995). An analogous stigmergic effect is achieved by "swarms" of humans that make a terrain more attractive to others by wearing down the vegetation with their own steps (Goldstone, Jones, & Roberts, 2006; Goldstone & Roberts, 2006; Helbing, Keltsch, & Molnár, 1997). Stigmergy has recently been proposed as an important mechanism for achieving multi-robot cooperation (Howsman, O'Neil, & Craft, 2004) and robustly interacting software systems (Parunak & Brueckner, 2004).

For these reasons, searching in a group changes the essential nature of the search process. The experiments and computational models that we will describe concern this process of collective search. Our approach stresses the macroscopic, group-level behavior that emerges when the groups' members pursue their self-oriented strategies. This work is an effort to complement cognitive scientists' tendency to focus on the behavior of single individuals thinking and perceiving on their own. Social phenomena such as rumors, the emergence of a standard currency, transportation systems, the World Wide Web, resource harvesting, crowding, and scientific establishments arise because of individuals' beliefs and goals, but the eventual form that these phenomena take is rarely dictated by any individual (Goldstone & Janssen, 2005).

### **Foraging for Concrete Resources**

A problem faced by all mobile organisms is how to search their environment for resources. Animals forage their environment for food, web-users surf the Internet for desired data, and businesses mine the land for valuable minerals. When an organism forages in an environment that consists, in part, of other organisms that are also foraging, then unique complexities arise. The resources available to an organism are affected not just by the foraging behavior of the organism itself, but also by the simultaneous foraging behavior of all of the other organisms. The optimal resource foraging strategy for an organism is no longer a simple function of the distribution of resources and movement costs, but it is also a function of the strategies adopted by other organisms.

### **Group Experiments on Foraging**

One model in biology for studying the foraging behavior of populations is the Ideal Free Distribution (IDF) model (Fretwell & Lucas, 1972; Seth, 2001). This model assumes that animals distribute themselves among patches so as to maximize the gained resources. The specific assumptions of the model are that animals 1) are free to move between resource patches without cost, 2) have correct (“ideal”) knowledge of the rate of food occurrence at each patch, and 3) are equal in their abilities to compete for resources. The model predicts an equilibrium distribution of foragers such that no forager can profit by moving elsewhere. This condition is met if the distribution of foragers matches the distribution of resources across patches.

Consistent with this model, groups of animals often distribute themselves in a nearly optimal manner, with their distribution matching the distribution of resources. For example, Godin and Keenleyside (1984) distributed edible larvae to two ends of a tank filled with cichlid fish. The food was distributed in ratios of 1:1, 2:1, or 5:1. The cichlids quickly distributed themselves in rough accord with the relative rates of the food distribution before many of the fish had even acquired a single larva and before most fish had acquired larvae from both ends.

Although animals frequently distribute themselves in approximate accord with an ideal free distribution, systematic deviations are also observed. One common result is undermatching, defined as a distribution of animals that is less extreme than the distribution of resources (Kennedy & Gray, 1993). An

example would be a 75/25 distribution of foragers when the resources have an 80/20 distribution. When undermatching occurs, there are fewer animals at the richer patch, and more animals at the leaner patch, than is optimal. The few experiments that have examined group foraging behavior with humans have also found undermatching (Kraft & Baum, 2001; Madden, Peden, & Yamaguchi, 2002).

Our experiments extend the previous studies of group foraging in humans in a few directions. First, we have developed a computer-based platform for the foraging experiment that allows us to manipulate experimental variables that would be difficult to manipulate in a more naturalistic environment. Second, we collect second-by-second data on the amount of resources and number of participants at different pools, which allows us to explore variation in resource use with high temporal resolution. Third, although our environment is virtual, it is naturalistic in one important respect – resources are distributed in a continuous spatial environment rather than at two discrete locations. Fourth, we do not designate or identify the resource alternatives to participants. As in many natural situations (Kennedy & Gray, 1993), the participants must discover the number and locations of resource patches themselves.

Using our virtual environment with interacting participants, we manipulated the relative outputs of the different resource pools and the knowledge possessed by the agents. In Godin and Keenleyside's (1984) experiment with cichlids, every cichlid could see the other cichlids as well as the larvae resources at both ends of the tank. Gallistel (1990) argued that this kind of information is important for the cichlids to distribute themselves rapidly in accord with the resource distribution. They are learning about the resource distributions by observing events that do not directly involve themselves. However, in individual reinforcement learning situations, an agent only has access to the outcomes of its own actions. It does not have access to the values of options not selected. Both situations occur naturally, and it is possible that the ability of a group to efficiently distribute itself to resources depends on the information at each agent's disposal (Templeton & Giraldeau, 1996).

In Experiment 1 (Goldstone & Ashpole, 2004), two resource pools were created with different rates of replenishment. The participants' task was to obtain as many resource tokens as possible during an experiment. A participant obtained a token by being the first to move on top of it. Resources were split evenly 50/50, or had a 65/35, or 80/20 split. In our visible condition, each participant could see each other

and the entire food distribution. In our invisible condition, they could not see other participants, and they gradually acquired knowledge of the resource distributions by virtue of their reinforcement histories.

166 undergraduate students from Indiana University served as participants and were run in 8 sessions with about 21 participants in each session. Each participant worked at their own computer and was told that they were being asked to participate in an experiment on group behavior. They were instructed to try to pick up as many “food” resources as possible by moving their icons’ position on top of food locations. Participants within a group co-existed in a virtual environment consisting of an 80 X 80 grid of squares populated by replenishing resource pools and other human-controlled agents. Participants controlled their position within this world by moving up, down, left, and right using the four arrow keys on their computers’ keyboard. Each participant was represented by a yellow dot. In the “Visible” condition, all of the other participants’ locations were represented by blue dots, and available food resources were represented by green dots. In the “Invisible” condition, participants only saw their own position on the screen and any food gathered by that participant in the last two seconds. After this time interval, these consumed food pieces disappeared.

The rate of distribution of food was based on the number of participants, with one piece of food delivered every  $4/N$  seconds, where  $N$  is the number of participants. This yields an average of one food piece per participant per four seconds. When a piece of food was delivered, it was assigned to a pool probabilistically based upon the distribution rate. For example, for the 80/20 condition, the food would be delivered to the more plentiful pool 80% of the time, and in the less plentiful pool 20% of the time. The location of the food within the pool followed a Gaussian distribution. Every experiment was divided into six 5-minute sessions. These six games consisted of all combinations of the two levels of knowledge (Visible versus Invisible) and the three levels of resource distribution (50/50, 65/35, 80/20). The locations of the resource pools were different for each of the six sessions so that participants would not have knowledge that carried over sessions. However, the distances between the two resource pools were kept constant across the sessions.

As a preliminary analysis of the distribution of agents across resource pools, Figure 1 shows the frequency with which each of the 80 X 80 grid cells was visited by participants, broken down by the six experimental conditions. The brightness of a cell increases proportionally with the number of times the cell was visited. The few isolated bright specks can be attributed to participants who decided not to move

for extended periods of time. The thick and thin circles show one standard deviation of the food distribution for the more and less plentiful resources, respectively. An inspection of this figure indicates that agents spend the majority of their time within relatively small regions centered on the two resource pools. The concentration of agents in pools' centers is greater for visible than invisible conditions, and is greater for the more plentiful pool. For the invisible conditions, there is substantial scatter of travel outside of one standard deviation of the pools' centers.

The dynamics of the distribution of agents to resources is shown in Figure 2, broken down by the six conditions. In this figure, the proportion of agents in the two pools is plotted over time within a session. We plot the proportion of agents in the pools, only counting agents that are within 3 standard deviations of one pool or the other. For the Visible and Invisible Resources conditions, an average of .7% and 16.7% of participants, respectively, were excluded because they were not in either resource pool. This large difference in exclusion rates is most likely due to the need for exploratory foraging in the Invisible Resources condition.

The 50/50 graph is not particularly diagnostic. For the 65/35 graph, horizontal lines indicate the proportions that would match the distribution of food. Although fast adaptation takes place, the asymptotic distribution of agents systematically undermatches the optimal probabilities. For example, for the 65/35 distribution the 65% pool attracts an average of 60.6% of the agents in the 50-300 second interval, a value that is significantly different from 65%. Undermatching is similarly found for the 80/20 distribution. So, if we were efficiency consultants, we would recommend that foragers in the less productive pool should move to the more productive pool -- the resources there are being relatively underutilized.

A final analysis of interest explores the possibility of periodic fluctuations in resource use. Informal experimental observations suggested the occurrence of waves of overuse and underuse of pools. Participants seemed to heavily congregate at a pool for a period of time, and then became frustrated with the difficulty of collecting food in the pool (due to the large population in the pool), precipitating a migration from this pool to the other pool. If a relatively large subpopulation within a pool decides at roughly the same time to migrate from one pool to another, then cyclic waves of population change may emerge. A Fourier transformation of the time series data was applied to test this. Fourier transformations translate a time-varying signal into a set of sinusoidal components. Each sinusoidal component is characterized by a phase (where it crosses the Y-intercept), amplitude, and frequency. For our purposes,

the desired output is a frequency plot of the amount of power at different frequencies. Large power at a particular frequency indicates a strong periodic response. The frequency plots for Experiment 1 show significantly greater power in the low frequency spectra for invisible than visible conditions. The power in lower frequencies is particularly high for the invisible condition with an 80/20 distribution. For all three invisible conditions, the peak power is at approximately .02 cycles/second. This means that the agents tend to have waves of relative dense crowding at one pool that repeat about once every 50 seconds. This 50-second period includes both the time to migrate from the first pool to the second pool and to return to the first pool. A pronounced power peak at lower frequencies is absent for the visible condition. One account for the difference in the two visibility conditions is that in the visible condition, each agent can see whether other agents are closer than themselves to underexploited resource pools. The temptation to leave a dissatisfying pool for a potentially more lucrative pool would be tempered by the awareness that other agents are already heading away from the dissatisfying pool and toward the lucrative pool. However, in the invisible condition, agents may become dissatisfied with a pool populated with many other agents, but as they leave the pool they would not be aware that other agents are also leaving. Thus, the ironic consequence of people's shared desire to avoid crowds is the emergence of migratory crowds! In a related irony, Rapaport, Kugler, Dugar, and Gisches (under review; see also Chapter X in this volume) report that when individuals within a group each choose paths so as to have a low-cost journey, the collective result can be average journey costs that are high. The problem in both of their and our group experiments is that people end up adopting similar courses of action despite their best intentions, and inefficient congestion arises.

In a second experiment (Goldstone, Ashpole, & Roberts, 2005), we wished to decouple the visibility of agents and food resources. Accordingly, we ran groups of participants in conditions where food resources, but not fellow foragers, were visible, and vice versa. This decoupling allows us to determine whether people use other people as information sources about where food may be located, and if so, how this information is used. On the one hand, an organism may be attracted toward patches occupied by its conspecifics. An animal can use the prevalence of conspecifics in a patch as information that the patch is highly productive. Consistent with this hypothesis, field experiments on migratory birds have shown that the presence of birds attracts other birds to a region (Pöysä, Elmberg, & Sjöberg, 1998). Adding birds to a site makes it more likely for still more birds to choose the site for nesting, which is why

duck hunters will put out decoys. Another familiar example is the tendency of buzzards to use the presence of other buzzards as an indicator of possible food sources, and therefore to fly to where there is a large group of buzzards.

On the other hand, an animal may avoid sites that already have a crowd of conspecifics. Pulliam and Danielson's (1991) ideal preemptive distribution hypothesis is that the first animals to arrive in an area will take the best territory, with subsequent arrivals taking the best remaining territories. This pattern has been observed with aphids. One of the central questions examined by Experiment 2 is: Are people more like buzzards or aphids with respect to the influence of conspecifics on foraging strategies?

The results, shown in Figure 3, indicate both systematic undermatching and overmatching in the distribution of human participants to resources over time. Consistent overmatching was found when resources were visible but other agents were invisible. When agents but not resources were visible, undermatching was found. The results support the hypothesis of ideal preemptive distribution rather than conspecific attraction. If participants were attracted to a resource pool because of the presence of other foragers at the pool, then overmatching would have been predicted with invisible resources and visible agents. That is, in a situation where direct knowledge of resources was lacking but the popularity of a pool could be used to estimate the pool's productivity, the presence of a relatively large number of participants at the richer pool would be expected to draw still more participants to the pool. In fact, a modest level of undermatching was observed in this condition. By contrast, according to the ideal preemptive distribution hypothesis, individuals at a site preempt other individuals from occupying that site. This is consistent with the undermatching observed when agents but not resources are visible, and it is also consistent with the release from undermatching (i.e., overmatching) observed when resources but not agents are visible. By this account, overmatching is found because participants are attracted to the rich, productive pools, and are not dissuaded from approaching the pools by the presence of other participants (who are invisible). As with Experiment 1, cyclic waves of population migration were suggested by a Fourier analysis. Together, the results suggest that our participants did not naturally make second-order inferences that other participants would be influenced by the same factors (i.e. dearth of resources in a pool, or sudden onset of food resources) that influenced themselves.

### **An Agent-based Model of Collective Foraging**

In developing EPICURE (Roberts & Goldstone, 2006, in press), an agent-based computational model, our intention was to build as simple a model as possible, with only those strategies included that the empirical results directly implicated. An interactive version of the resulting model is available at <http://cognitn.psych.indiana.edu/Epicure.html>. We populated a world with agents that probabilistically decided from moment to moment what spatial grid location they would move toward. The likelihood that a particular location is selected as the target destination is based on the location's value relative to all other locations. Value is influenced by several factors. The first is the distance of location. The closer a location is, the more likely it is to be selected as a target destination. Second, once a location has been selected as a target, we increase its value so that it will tend to be selected as a target at the next moment too. This is a way of incorporating inertia to target locations, and we call this goal bias. Our empirical results showed that people tend to stick around a resource for a while and then switch to the other pool if they know where it is. We do not see people start to move toward a pool, get midway into “no-man's land,” and then head back to the pool they just left. However, computational foragers did exactly this until we incorporated inertia for their destinations.

There are different rule variants for the foragers in the visible and invisible conditions. For the agents who can see all of the other agents and food, a third factor is that the value of a location increases as the density of food in its vicinity increases, and fourth, the location's value decreases as the density of other agents increases. The motivation for this is that if other agents are present in a region, then there will be more competition to get the food and hence the region is less attractive. Neither of these sources of information is available in the invisible condition. In the invisible condition, agents must gradually accumulate a personal history of where they have found food. Every time food is found in a location, the location's value increases, and this increase diffuses to the nearby locations. Conversely, if a cell is visited but contained no food, then its value, and its neighbors' values decrease.

EPICURE is able to account for the empirically observed pattern of overmatching and undermatching for the four visibility conditions in Experiments 1 and 2. For the visible condition of Experiment 1, EPICURE's agents rapidly converge on a distribution that approximates the food distributions but consistently undermatches the food distributions (as shown in Figure 3). For the invisible condition, we also find that the agents can learn where the food clusters are given a bit more time, but that there is also asymptotic, stable undermatching over the entire experiment. Finding undermatching was

surprising to us because we had anticipated that we would only get undermatching if we included an explicit bias in our agents to assume that all discovered resource patches had approximately equal frequencies of outputs. In fact, our agent-based system spontaneously produced undermatching like our human subjects did, even though it makes none of the typical assumptions posited by the biology literature to explain undermatching (Kennedy & Gray, 1993). It does not need a bias to spend equal time at different resource pools, unequal competitive abilities among foragers, or foraging interference. Why does EPICURE predict undermatching? The critical notion is spatial turfs. A single agent can efficiently patrol a compact region of about 10 squares despite large differences in food productivity. Although the 80% pool has 4 times the productivity of the 20% pool, they both have the same spatial extent and variance, and so can support agents in numbers that are more similar than predicted by the pools' productivities.

A corollary of this account of overmatching is that the matching of agents to resources should depend heavily on the variances of the pools. EPICURE's predictions are shown in Figure 4. When a resource pool occupies a larger area, then all else being equal it will attract more EPICUREans. When the pool variances are identical (8 vs. 8), the agents slightly undermatch the resources. When the 80% pool has half the variance of the 20% pool (8 vs. 16), dramatic undermatching occurs because it takes more agents in the 20% pool to cover the much larger area. When the variances are reversed (16 vs. 8), the rarely observed phenomenon of overmatching occurs, and the explanation lies in the fact that the *densities* of the pools are equal, but the *coverage times* are unequal because the food rate is low. Agents in the 20% pool have less area to cover and fewer pieces of available food, so the average pick-up time remains relatively low and food does not last long enough to attract agents, as it does in the 80% pool. Finally, in the 16 vs. 16 condition, nearly perfect matching is observed. EPICURE's prediction, comparing the 8 vs. 8 to the 16 vs. 16 conditions, is that as the spatial extent of two pools increase, better matching should be found. Baum and Kraft (1998) found similar results with pigeons competitively foraging for food from two bowls (small resource pools) or troughs (large).

EPICURE also correctly predicts our results from the mixed visibility conditions of Experiment 2. When agents and resources are both visible, there is a tendency for participants not to go to the more prolific pools because of the restraining effect of high agent density. However, when only the resources are visible, agents cannot see the high agent density of the prolific resource, and hence the empirically

observed overmatching is also found in the model. EPICURE also spontaneously produces population waves revealed by Fourier analysis, and like our empirical data, the strongest frequency response was at about .02 cycles/second. The model correctly predicts an average of about 2-3 pool switches across a 5-minute experiment. There are several results in the literature that are also predicted by the model. Greater undermatching is predicted and empirically found as the number of agents increase (Gillis & Kramer, 1987). As the number of agents increase, or the resource area gets more constricted (Baum & Kraft, 1998), then the number of unoccupied turfs decreases and the resources are well covered by agents staking out their small turfs. The spatial coverage of the agents becomes more important than the resource pool productivity in determining the distribution of agents.

A final, somewhat counterintuitive prediction of EPICURE is that increasing the distance between two resource pools, and hence travel cost, should decrease, not increase, undermatching. For example, in one set of simulations, we expanded the gridworld to 120 x 120 cells, and compared simulations with pool centers at (20, 20) and (100, 100) to simulations with pool centers at (40, 40) and (80, 80). With visible agents, the increased distance between pools led to nearly perfect matching. These results agree with the empirical pigeon foraging results found by Baum and Kraft (1998). As the distance between pools increased, the pigeons more closely matched the Ideal Free Distribution model, and they switched pools significantly fewer times. Milinski (1987) also found that stickleback fish switch between pools significantly less often as the distance between pools increases. Likewise, in our simulation, the far-apart pools led to significantly fewer average switches than the closer pools. The dynamics of the constrained visible model offer a simple explanation for both the switching and matching results. As the pools become more separated, it is much less likely that an agent will probabilistically choose to switch pools, because the other pool's resources are so far away. Furthermore, if the agent does decide to switch, the longer distance means there are more opportunities for the agent to change its decision and choose a food pellet in the previous pool, although the goal bias tempers this change in decision. The decreased switching, in turn, promotes better matching because the new pool must appear to be consistently better in order for the agent to complete the journey. The burden of switching is higher, so agents are more likely to switch only when there is a true advantage. An additional consideration is that as travel costs increase, by further separating pools, undermatching increases because it becomes less likely that an agent will

sample from both pools and if the pools are equally large then they have approximately equal chances to originally capture an agent.

It might be argued that EPICURE can be fit to experimental results, but does not genuinely predict outcomes from foraging experiments. In EPICURE's defense, many of the model's behaviors were observed before we knew of their empirical support. These predictions are invariant with changes to parameters within reasonable ranges. This includes its surprising prediction of decreased undermatching with increased travel costs, and increased undermatching with smaller resource patches. However, for the future record, we also offer further genuine predictions of EPICURE for which we do not currently know of supporting evidence. First, EPICURE predicts greater disparity of wealth (resources retrieved) among agents in the invisible than visible resources condition. If resources are invisible, then some fortunate agents can effectively monopolize a resource pool without other agents getting a chance to sample from it. Second, EPICURE predicts greater disparity of wealth among agents in the more prolific, compared to less prolific, pool. Third, EPICURE predicts strong reliance on social information with an environment containing a few large resource pools that stochastically appear and deplete in different regions of the environment, but predicts foragers to rely on privately obtained information if those same resource pools are small and quickly depleted. We are currently conducting experiments to test these predictions.

In summary, we believe that EPICURE provides an elegant synthesis of several results from the literature on human and animal foraging. Our empirical results highlight the importance of knowledge on group-level behavior. We find three empirical inefficiencies in our groups' behavior: 1) undermatching in the sense that there were too many participants in the less plentiful resource and too few participants in the more plentiful resource, 2) participants were more scattered than were the food resources, and 3) systematic cycles of population change are apparent whereby the migration of people from one pool to another is roughly synchronized. All three of these inefficiencies were more pronounced for invisible than visible conditions. Knowledge of food distributions allows an agent to more effectively match those distributions, whereas knowledge of other agents allows an agent to more effectively decouple their responses from others. The importance of agent and food information led us to feature these in our computational model of foraging. One of the best ways to evaluate a complex adaptive system model is to see whether behaviors arise that are not explicitly forced by the rules – are we getting out more than we knew we were putting in the model. By this measure, the model does a good job of explaining collective

foraging behavior. EPICURE shows high-level behaviors like undermatching and population cycles even though it was not built with the intention of creating them, and the model also predicts the specific dependencies of these high-level behaviors on population size, and the location, variance, and the productivity of resources.

### **Propagation of Innovations in a Group**

The previous experiments and computational model described a situation with competitive foraging for spatial resources. We would now like to turn our attention to a situation with collective foraging for abstract resources. In the concrete spatial foraging task, resources consumed by one agent could not be consumed by others. However in our abstract search scenario, resources are less tangible and thus those used by one agent can still be used by another. This is akin to searching for the solution to a math problem which, once found, can be imitated with no loss to the discoverer. Nevertheless, like the foraging task, agents can still benefit from knowing where other agents are searching. Furthermore, similar to the foraging task, the abstract search situation is not genuinely cooperative because each agent is still trying to maximize its own, not the group's performance.

### **Group Experiments on Innovation Propagation**

Humans are uniquely adept at adopting each others' innovations. While imitation is commonly thought to be the last resort for dull and dim-witted individuals, cases of true imitation are rare among non-human animals (Blackmore, 1999), requiring complex cognitive processes of perception, analogical reasoning, and action preparation. This capacity for imitation has been termed "no-trial learning" by Bandura (1965), who stressed that, by imitating one another, people perform behaviors that they would not have otherwise considered. When combined with variation and adaptation based on reinforcement, imitation is one of the most powerful methods for quick and effective learning. Cultural identity is largely due to the dissemination of concepts, beliefs, and artifacts across people. The tendency for humans to imitate is so ubiquitous that Meltzoff (1988) has even suggested that humans be called "Homo imitans."

In social psychology, there has been a long and robust literature on conformity in groups (Cialdini & Goldstein, 2004; Sherif, 1935). The usual finding is that people conform to majorities in groups. To some degree, conformity is found because people desire to obtain social approval from others.

For example, sometimes when people give their answers privately, they are less likely to conform to the group's opinion than when responding publicly (Deutsch & Gerard, 1955). However, at other times, the conformity runs deeper than this, and people continue to conform to the group's opinion even privately (Sherif, 1935). In our experiments and modeling, we are interested in the use of information provided by others even when social approval motivations are minimized because the group members never meet one another and are anonymous.

Conformity to others' ideas has been a major field of research not only in social psychology, but also in economics, political science and sociology. It is common in models of collective action to make an individual's decision to participate based upon their expectations for how many other people will participate (Chwe, 1999). A common outcome of a collective "I'll do it if you do it" mentality, is for "tipping points" to arise in which adding a couple more participants to an action leads to a positive feedback cycle in which still more participants sign on, leading to an exponential increase in participation for a time (Gladwell, 2000). This behavior is a sensible policy both because the likelihood of success of an innovation depends upon its public adoption rate (Bullheimer et al., 1998) and because other people may have privileged information unavailable to the individual making a choice. The potential cost of this bandwagon behavior is wasted time, money, and effort in adopting new innovations (Rosenkopf & Abrahamson, 1999; Strang & Macy, 2001).

Our studies explore the diffusion of innovative ideas among a group of participants, each of whom is trying to individually find the best solution that they can to a search problem. The work fills an important gap in research. There are several excellent computational models for how agents in a population exchange information (Axelrod, 1997; Kennedy & Eberhart, 2001; Nowak et al., 1990). There is also excellent work in social psychology on how individuals conform or use information provided by others (Gigone & Hastie, 1996). Fieldwork also explores actual small groups of people engaged in cooperative problem solving (Arrow et al., 2000). However, there is very little work with laboratory-controlled conditions that explores the dynamics of a group of participants solving problems as they exchange information. One related study is Latané and L'Herrou's (1996) exploration of participants' sending E-mail messages to each other (see also Latané & Bourgeois, 1996), as they tried to predict which of two options their group would select. Over the course of message exchanges, neighboring participants in the network tended to adopt similar choices (consolidation) but there was also continued diversity of

choices across the entire network. In contrast to this work, our research predominantly focuses on situations where participants are trying to find good solutions to a problem rather than trying to conform to their neighbors. For example, farmers may discuss the benefits of various crop rotation techniques with their neighbors, and may be convinced to try a new one by a neighbor's success, but there is no reward to conforming to a neighbor's behavior in itself.

In creating a paradigm for studying information dissemination, our desiderata were: 1) a problem to solve with answers that vary continuously on a quantitative measure of quality, 2) a problem search space that is sufficiently large that no individual can cover it all in a reasonable amount of time, and 3) simple communications between participants that are amenable to computational modeling. We settled on a minimal search task in which participants guess numbers between 0-100 and the computer reveals to them how many points were obtained from the guess by consulting a hidden fitness function (Mason, Jones, & Goldstone, in press). Additionally, random noise was added to the points earned, so that repeated sampling was necessary to accurately determine the underlying function relating guesses to scores. Over 15 rounds of guesses, participants try to maximize their earned points. Importantly, participants get feedback not only on how well their own guess fared, but also on their neighbors' guesses. In this manner, participants can choose to imitate high-scoring guesses from their neighbors. We experimentally manipulated the network topology that determines who counts as neighbors, as well as the fitness function that converts guesses to earned points.

We created neighborhoods of participants according to random, regular lattice, fully connected, and small world graphs. Examples of the graph topologies for groups of 10 participants are shown in Figure 5. In the random graph, connections are randomly created under the constraint that the resulting graph is connected -- there is a path from every individual to every other individual. Random graphs have the property that individuals tend to be connected to other individuals via paths that do not require passing through many other individuals. This property has been popularized as the notion of "six degrees of separation" connecting any two people in the world, and has been experimentally supported (Milgram, 1967). More formally, the average path length connecting two randomly selected nodes in a random graph is  $\ln(N)/\ln(K)$  where  $N$  is the number of nodes and  $K$  is the average number of neighbors connected to each node. The regular lattice can be used to represent a group with an inherent spatial ordering such that people are connected to each other if and only if they are close to one other. The regular lattice also

captures the notion of social “cliques” in that if there is no short path from A to Z, then there will be no direct connection from any of A’s neighbors to any of Z’s neighbors. In regular lattices, the average path required to connect two individuals requires going through  $N/2K$  other individuals. Thus, the paths connecting people are much longer, on average, for lattice than random graphs.

Random graphs have short paths, but unfortunately (from the perspective of modeling social phenomena) do not contain cliques. Lattices show cliques, but do not have short path lengths. Recently, considerable interest has been generated in networks that have both desirable properties, so-called “small world networks.” These networks can be formed by starting with a lattice and randomly rewiring (or adding new connections, in the case of our experiments and Figure 5) a small number of connections (Watts & Strogatz, 1998). The result is a graph that still has cliques because nodes that are connected to the same node tend to be spatially close themselves, yet also have a short average path length. From an information processing perspective, these are attractive networks because the spatial structure of the networks allows information search to proceed systematically, and the short-cut paths allow the search to proceed quickly (Kleinberg, 2000). Notice, in Figure 5, that all three of the described networks have a total of 12 connections between 10 participants. Thus, if there is a difference in information dissemination in these networks, then it must be due to the topology, not density, of the connections. A fourth network, a fully connected graph, allowed every participant to see the guesses and outcomes of every other participant.

We compared two hidden functions for converting guessed numbers to points. The unimodal function has a single best solution that can always be eventually found with a hill-climbing method (see Figure 6). The trimodal function increased the difficulty of the search by introducing local maxima. A local maximum is a solution that is better than all of its immediate neighboring solutions, yet is not the best solution possible. Thus, a simple hill-climbing method might not find the best possible solution.

Twelve groups of Indiana University undergraduate students ranging in size from 7 – 18 people with a median of 14 people per group participated for partial course credit, for a total of 153 participants. Each group participated in eight experiments that consisted of every combination of the four network types (Figure 5) and two fitness functions (Figure 6). Participants were told to try to maximize their total number of points acquired over 15 rounds of number guessing, and that the same guess would be worth about the same number of points from round to round, but that a certain amount of randomness was added

to the earned points. Participants were also told that they would see the guesses and points earned by some of the other participants, and that these others would also see the participants' guesses and earnings.

The results from this experiment are shown in Figure 7, expressed in terms of the percentage of participants within one-half standard deviation of the global maximum for a fitness function. Over the 15 rounds, increasingly many participants find the global maximum. For the unimodal function, the fully connected network finds the global maximum most quickly, and the advantage of the fully connected network over the other three networks is particularly striking for Rounds 2-4. Around Round 5, the small world network catches up to the performance level of the fully connected network, and for the rest of the rounds, these two network types continue to outperform the other two networks. This pattern of results is readily explainable in terms of the propensity of a network to disseminate innovations quickly.

Innovations disseminate most quickly in the full network because every individual is informationally connected to every other individual.

For the multimodal payout function, the small-world network performs better than the fully connected network for the first six rounds. One account for its superiority over the full network is that the small world network is able to thoroughly search the problem space. The fully connected groups frequently get stuck in local maximum because the groups prematurely converge on a good, but not great, solution. The small-world structure is an effective compromise between fully exploring a search space and also quickly disseminating good solutions once they are found. Much like how our foraging simulations counter-intuitively revealed a more optimal distribution of agents to resources when the environment limited the ability of the agents to easily explore each site, the most surprising aspect of these results is that the truism of "the more information, the better" is not supported. Giving each participant all of the results from all of the agents does not lead to the best group solution for the multimodal problem – the problem with this policy is that with the fully connected network, everybody ends up knowing the same information. Participants thereby become too like-minded, acting like a single explorer, rather than a federation of independent explorers.

The general point from this first experiment is that before one decides how to connect a group, one should know about the nature of the problem the group needs to solve. A candidate generalization is that the more exploration a group needs to do, the more clustered and locally-connected the network should be. Conversely, the more quickly a group needs to exploit emerging solutions, the more globally

connected individuals should be. Problem spaces that require considerable exploration to find the global maximum should benefit from networks that have relatively well isolated neighborhoods that can explore different regions of a problem space. To test this hypothesis, in a separate experiment we also tested the more difficult fitness function shown in Figure 8 that we call the needle function. This function features one very broad local maximum, and one hard-to-find global maximum. We tested 12 groups of participants in needle functions like this, with each group connected in the same four network topologies we used before. For this function, Figure 9 shows that the lattice network performed better than the other three network types, starting by Round 7 if not earlier. The lattice network fosters the most exploration because of its spatially segregated network neighborhoods. Exploration of the problem space is exactly what is needed for the needle function because of its hard-to-find global maximum.

The three payout functions are ordered by the demands they place on broad exploration of a problem space. The benefit for exploration increases going from the unimodal to the multimodal to the needle function. In parallel, the network structures are ordered by their preservation of local cliques of nodes. Cliquishness increases going from full to small world to lattice networks. These two progressions are coordinated, with the full network performing best with the unimodal function, the small world network performing best with the multimodal function, and the lattice performing best with the needle function. In contrast to arguments for a general informational advantage of small world networks (Watts & Strogatz, 1998), we find that what network is best depends on the kind of problem a group must solve (see also Lazer & Friedman, 2005). As broader exploration is needed to discover good solutions, increasingly cliquish networks are desirable.

### **A Computational Model of Innovation Propagation**

We have developed an agent-based computational model of our experiments based on the premise that members of a group can choose to explore a problem space on their own or take advantage of the solutions found by others. In the model, called SSEC (for Self-, Social-, and Exploration-based Choices), every agent on every round probabilistically chooses between three strategies: using their own guess on the last round, using their neighbors' best guess on the last round, and randomly exploring. Each agent randomly chooses between these strategies, with the likelihood of each strategy based on its intrinsic

bias and also its observed success. Guesses also include normally distributed randomness to avoid perfect imitation. The model, thus, can be expressed as

$$p(C_x) = \frac{B_x S_x}{\sum_n B_n S_n}$$

where  $p(C_x)$  is the probability of using Strategy  $x$ ,  $B_x$  is the bias associated with the strategy and  $S_x$  is the score obtained from the strategy. The participant's guess is then  $G_x + N(\mu = 1, \sigma = 1)$ , including normally distributed randomness, with  $G_x$  being the guess associated with Strategy  $x$ . When the random exploration strategy is selected, a uniform distribution is used to select the next guess. This model is motivated by the Particle Swarm Algorithm (Kennedy, Eberhart, & Shi, 2001). However, unlike the swarm algorithm, the SSEC model allows sudden jumps in guesses rather than smoothly changing patterns of oscillations around promising solutions. The experimental results showed that participants frequently jumped from one guess to a completely different guess, a behavior that the original Particle Swarm Algorithm does not accommodate.

The simplest version of this model with mostly default parameter values for the biases was able to accommodate some, if not all, of the trends in the results. In particular, we tested a version of the model in which  $B_1$  (the bias for using one's own previous guess) is 1,  $B_2$  (the bias for using one's neighbor's best scoring guess) is 1, and  $B_3$  (the bias for randomly exploring) is 0.1. This is essentially a one-parameter control of biases because  $B_1$  and  $B_2$  were constrained to be equal, and only the relative, not absolute value of  $B_3$  matters given the choice model used to determine strategy choice. In addition, the value of  $\sigma$  that determines the mutation/drift rate for guesses was set to 3, and noise with a variance of 30 and a mean of 0 was added to the fitness function's output, just as it was to experiment scores. Each of the four network types was run 1000 times with each of the three fitness functions for 15 rounds of guessing and 15 agents per group. This model showed the best performance with the full network with the unimodal function, the best performance with the small world network with the multimodal function, and a two-way tie for the best final performance between the lattice and small world networks for the needle function. This pattern is the same as empirically observed, except for the needle function, where we found the lattice network performing better than all other network types.

Given the promising results of this original set of simulations, we parametrically manipulated the network connectivity to continuously shift from a regular lattice with only local connectivity to a fully

connected network in which every agent is directly connected to every other agent. This was achieved by connecting 15 agents via a lattice, and then adding a number of additional random connections between agents. As the number of random connections increases, the network initially transforms from a random network to a small world network. Then, as the connectivity further increases, the network transforms from a small world network to a fully connected network. If more information communicated in a network always increases group performance, then we expect better performance (shown by brightness in Figures 10-12) as connectivity increases.

Independently, we manipulated the relative weight given to information obtained from oneself compared to others. Keeping  $B_3$  constant at 0.1, we varied  $B_1$  from 0 to 1 and set  $B_2$  equal to  $(1 - B_1)$ . Thus, we varied the degree to which each agents' guesses are based on their own previous guess compared to others' guesses. In Figures 10-12, as we go from left to right, we go from "sheepish" agents that base their guesses completely on others' guesses (and an occasional random guess) to "mavericks" that always go their own way without any influence of others.

Figures 10-12 show that the influences of connectivity and agent independence are not constant, but rather depend on the shape of the problem space. For the easy-to-solve unimodal problem, Figure 10 shows that group performance increases monotonically with both increased reliance on others' information and increased connectivity. Both trends can be explained by the fast propagation of innovations obtained when agents follow their best peers, and have many peers to follow. For single-peaked problems, there are no local maxima and so no concern with hasty collective convergence on sub-optimal solutions.

For the multimodal function (Figure 11), optimal group performance involves intermediate levels of both connectivity and self-reliance. These two factors trade off with one another such that increases in connectivity can be offset by decreases in conformity. Networks that have only local connectivity and self-reliant individuals perform relatively poorly because good solutions are inefficiently spread. Conversely, networks that have global connectivity and conformist individuals also perform poorly because the group frequently converges on local rather than global maxima. Good group performance is found when a group can both search a problem space for good solutions, and yet spread those solutions quickly once they are found. This is achieved when conformist individuals are combined with a network that limits connectivity, or when self-reliant individuals are combined with more broadly connected

networks. If one is able to engineer a social network, then one's target network should depend both on the problem, and "personalities" (mavericks vs. sheeps) of the nodes in the network.

For the trickier needle function (Figure 12), the best performing networks are pushed even further in the direction of increasing self-reliance and decreasing connectivity. Consistent with our empirical results, the needle function requires more exploration, and both limiting connectivity and increasing self-reliance promote independent exploration of group members. As with the multimodal function, there is a tradeoff between network connectivity and individual self-reliance.

A major conclusion from both the experiments and modeling is that propagating more information is not always good for the group. Lazer and Friedman's (2005) computational model converges on the same maxim: full access to what everybody else in a group is doing can lead human and computational agents to prematurely converge on suboptimal local maxima. Networks that preserve spatial neighborhoods promote exploration, and this can explain why the full network is the best network for the unimodal function, the small world network and its intermediate level of connectivity does best with the trimodal function, and the lattice function with no long-range connections does best with the difficult needle function.

Although more information is not always better as far as the group goes, it is always in the best interest of individuals to use all of the information at their disposal. Accordingly, our innovation propagation paradigm provides an unexpected example of a social dilemma (Goldstone & Janssen, 2005; Ostrom, Gardner, & Walker, 1994). Individuals, looking out for their own self-interest, will seek out as much information from others as possible, but this can inhibit the group as a whole from widely exploring a search space. Thus, obtaining information from as many peers as possible is non-cooperative behavior even though it does make links between individuals. Searching a problem space on one's own is cooperative in the sense of allowing the group as a whole to collect the most points possible, by avoiding local maxima. Our simulations show that every individual agent is best off linking to as many other people as possible. Agents with relatively many links outperform those with relatively few links. However, if every agent links maximally to every other agent, then the entire group does not perform well due to premature convergence on good, but not optimal, solutions. Sensitivity to this conflict between individual and group interests may help in the design of adaptive social networks. Designing for the greater common good may sometimes entail placing limits on individuals' ability to connect with each

other. Problems with difficult, hard to find solutions often drive people to look to others for hints and clues, but these are exactly the kinds of problems for which limited, local connectivity is advantageous.

This analysis of the conflict between the good of the individual and group becomes particularly relevant when we turn to situations where people can choose their connectivity, rather than having it imposed. Pursuing experimental paradigms in which people can create their own social networks would be valuable as connecting with both the mathematical literature on the evolution of networks (Dorogovtsev & Mendes, 2003) and the social science literature on coalition formation (Kahan & Rapaport, 1984). In many naturally occurring groups, people have some choice in who they will share information with, and what information they will reveal. From our perspective on human groups as complex systems, one of the interesting issues will be to study the global efficiency of information transmission in self-organized networks, and how incentives to individuals can be structured so that globally advantageous networks emerge.

### **Towards Unifying Collective Search Paradigms**

We have presented two cases of collective search – in concrete and abstract problem spaces. Although collective resource harvesting in space seems, at first, to have little to do with how innovations are propagated in a community, the continuity and parallels between them increasingly impress us. To see these connections, it is useful to consider some salient differences between the paradigms in terms of concreteness, competition, and the importance of information sharing.

In terms of concreteness, the physical space of the foraging environment clearly presents search constraints missing in the number-guessing scenario. Foragers must pass through intermediate locations, and accrue travel costs when trekking over spaces between resource pools. In the number-guessing experiments, participants can hop from any number to any other number. However, in both scenarios, the resources themselves are “clumpy,” distributed in compact physical spaces for the foragers, or smooth fitness functions in the innovation propagation experiments. Furthermore, the physical constraints in the forager environment can easily be relaxed. The simple methodological change of allowing participants to point and click to destinations and immediately transport there eliminates the constraints of space. Alternatively, a version of the number-guessing game in which participants can only alter their guesses slowly serves to implement a one-dimensional space. These modifications make a strong case for

unifying these paradigms. Travel costs, metric spaces, and ordered dimensions can be either present or absent in both paradigms, for both participants' movements and the resource distributions.

A second difference between the paradigms concerns competitiveness. In the foraging experiments, a food token that is picked up by one participant is no longer available for others. Resource collection is competitive. By contrast, in the number guessing experiment, any number of participants can guess the same number without penalty. Resources are not consumed as they are sampled. This difference in competition yielded different results. For our foragers, the presence of people at a resource pool often dissuaded others from harvesting the pool, but for our number-guessers, there was no evidence for being deterred by crowds. However, it is easy to imagine versions of foraging without competition. For example, when "foraging" for good stock investments, stock prices do not fall, but rather typically increase, as more individuals buy a stock. Alternatively, competition can be introduced into the number-guessing experiment by dividing the points earned by a numeric guess by the number of individuals making similar guesses. In fact, exactly this tactic of "resource sharing" has been proposed by genetic algorithm researchers in order to increase the diversity of a population of evolved solutions to multi-dimensional problems (Goldberg, 1989). The competitiveness of resource harvesting is an important factor in determining the diversity of solutions and the advantage of traveling with and apart from crowds, but it is also a factor that may or may not accompany either scenario.

Finally, the paradigms seem to differ in the importance of information sharing. For the number guessing experiment, participants who do not ever copy others are at a distinct disadvantage. Sharing information about solution quality was sufficiently important that we explicitly manipulated the network configuration that determined the pathways for this sharing. Network connectivity was not manipulated in the foraging experiments, and a forager can perform well without ever following another forager's lead. However, as we imagine expanding the forager world, decreasing the visibility of resources, and increasing the visibility of fellow foragers, the importance of information sharing increases. Even in our relatively small foraging environments, people often used other people as information scouts when people but not food resources were visible.

All of these factors affect real world search tasks. The results of our empirical foraging task and subsequent model can be extended towards concrete foraging tasks in the real world, such as chimpanzees foraging for fruit (Hashimoto et. al, 2003), the migratory patterns of early hominid foragers, and guppies

foraging for mates (Dugatkin, 1992). Furthermore, many of the same principles also apply to abstract foraging tasks, such as human information foraging (Pirolli and Card, 1999), information foraging on the World Wide Web, and Internet dating sites. Meanwhile, the results from our number-guessing scenario indicate that the social networking structure can determine the dynamics of collective search tasks.

Together, these studies may lead to a greater understanding of individuals' usage of personal information versus public information, and how this usage is contingent on the social context of a task. For example, undergraduate students face a dilemma when choosing courses for the semester. They certainly want to take courses that involve their own interests and satisfy requirements, but they also want to make use of social information that indicates which classes are easy, which professors are the best, etc. The social information changes as students begin relying on websites' ratings of professors and courses, rather than simply the students' local networks of friends and acquaintances. The students' local networks are presumably a more reliable source of knowledge, but the websites offer a greater breadth and accumulation of knowledge. This shift in public information usage can have real consequences in terms of the enrollment in particular courses and the subsequent course offerings.

Similar issues of acquiring and weighting information occur in the purchasing of consumer goods. Individuals choose cars and cellphones according to features and style, but they also care about relevant information from their peers (Janssen & Jager, 2003), whether that information concerns the popularity of a product or simply the product's reliability. Sometimes the criteria of the search space are relatively well-defined, such as a car's mechanical reliability or the average number of dropped calls on a cellphone, but the criteria often include intangibles, such as the aesthetics of a cellphone's sleek style. Furthermore, we generally receive this product information from a variety of social networks, including our real-world social network, web social networks such as MySpace, and search engines that may approximate full connectivity via their abundance of links.

Clearly these search situations can become increasingly complex as individuals weigh their own personal information and the public information acquired from various sources, but our empirical paradigm provides a rich environment for both isolating and combining these factors. In our current work, we are exploring several tasks to better elucidate these real world scenarios. In one experiment, we are examining the effects of a social network on an individual's rating of musical pieces. Unlike the number-guessing scenario, there is no correct, optimal solution. In a second experiment, we are studying a task

very similar to the number-guessing task, except it involves guessing two-dimensional pictures by filling-in squares and receiving both individual feedback and feedback regarding network neighbors' guesses. This task involves a more difficult search space, and as a result, we expect an increased reliance on social information. Finally, we are also conducting an experiment that examines the relative participation in producer and consumer roles in a social network. In this experiment, we are trying to gauge collective search performance when each experiment round involves individuals' choosing to either search for their own solutions or consume from their neighbors' current guesses. We expect to find population oscillations as individuals alternate between scavenging group information and being forced to search for their own information, but we may also find that a mix of strategies is crucial for good group performance.

Ultimately, we will use the results from these studies to explore increasingly rich collective search tasks. By manipulating concreteness, competition, the importance of information sharing, and the inclusion of multiple sources of personal and public information, we can create a family of experiments unified by general principles of collective search. In the current article, we have described several key principles gleaned from our initial collective search experiments. These principles, exemplified in both of our paradigms, include: 1) a tradeoff between exploration and exploitation, 2) a compromise between individuals using self- versus other-obtained information, and 3) the emergence of group-level resource use patterns that result from individual interests but are not always favorable to these interests. Unfavorable manifestations of these principles include inefficient population waves, bandwagon effects, mismatches between agent and resource distributions, disadvantages for highly connected networks, and premature convergence of populations on local maxima.

Despite these difficulties, collective search continues to be an immensely powerful case of distributed cognition (Gureckis & Goldstone, in press) for the simple reason that individual search often fails to provide a solution in a limited time. The shopping website Amazon has a feature that tracks one's searches and uses them as information to provide recommendations to others with similar searches. By including this additional information, the website transforms the individual's search for good products into an instance of collective search, with the same benefits and foibles as described in this chapter. Similarly, the search for missing children or wanted criminals relies on a group of individuals searching and sharing information about their search. If the features of the search are known (e.g., how far the criminal could have traveled, or whether the search problem is easy or difficult), the communications between searchers

can be set up to benefit the group and avoid unfavorable outcomes. In light of the peril and promise, it behooves cognitive scientists of all stripes to work together toward solving the problem of understanding and improving collective search.

## References

- Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). Small groups as complex systems: Formation, coordination, development, and adaptation. Newbury Park, CA: Sage.
- Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 4, 203-226.
- Bandura A. (1965). Behavioral modification through modeling procedures. In L. Krasner & L. P. Ulmann (Eds.), *Research in behavior modification: new development and implications* (pp. 310-340). New York: Rinehart and Winston.
- Baum, W.M., & Kraft, J.R. (1998). Group choice: Competition, travel, and the ideal free distribution. *Journal of the Experimental Analysis of Behavior*, 69, 227-245.
- Blackmore, S. J. (1999). *The meme machine*. Oxford: Oxford University Press.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.
- Bullheimer B., Dawid H., Zeller R. (1998). Learning from own and foreign experience: technological adaptation by imitating firms. *Computational and Mathematical Organizational Theory*, 4, 267-282.
- Cialdini, R. B. & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591-621.
- Chwe, M. S-Y. (1999). Structure and strategy in collective action. *American Journal of Sociology*, 105, 128-156.
- Clark, A. (2003). Natural born cyborgs. Oxford: Oxford University Press.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2003). *Evolution of networks: From biological nets to the internet and WWW*. Oxford: Oxford University Press.
- Dugatkin, L.A. (1992). Sexual selection and imitation: Females copy the mate choice of others, American Naturalist, 139, 1384-1389.
- Fretwell, S. D. & Lucas, H. J. (1970). Ideal free distribution. *Acta Biotheory* 19, 16-21
- Gallistel, C.R. (1990). *The organization of learning*, MIT Press (Cambridge).
- Gigone, D., & Hastie, R. (1996). The impact of information on group judgment: A model and computer simulation. In J. H. Davis & E. Witte (Eds.), *Understanding group behavior: Consensual action by small groups* (vol. 1, pp. 221-251). Hillsdale, NJ: Erlbaum.
- Gillis, D.M., & Kramer, D.L. (1987). Ideal interference distributions: Population density and patch use by zebrafish. *Animal Behavior*, 35, 1875-1882.
- Gladwell, M. (2000). *The tipping point*. London: Little Brown and Company.
- Godin, M.J., & Keenleyside, M.H.A. (1984). Foraging on patchily distributed prey by a cichlid fish (Teleostii, Chichlidae): A test of the ideal free distribution theory, *Animal Behavior*, 32, 120.
- Goldberg, D. E. (1989). Genetic Algorithms. Reading, MA: Addison-Wesley. (Chapter 1. pp. 1-23).
- Goldstone, R. L., & Ashpole, B. C. (2004). Human foraging behavior in a virtual environment. Psychonomic Bulletin & Review, 11, 508-514.

- Goldstone, R. L., Ashpole, B. C., & Roberts, M. E., (2005). Knowledge of resources and competitors in human foraging. Psychonomic Bulletin & Review, 12, 81-87.
- Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. Trends in Cognitive Science, 9, 424-430.
- Goldstone, R. L., Jones, A., & Roberts, M. E. (2006). Group path formation. IEEE Transactions on System, Man, and Cybernetics, Part A, 36, 611-620.
- Goldstone, R. L., Roberts, M. E. (2006). Self-organized trail systems in groups of humans. Complexity, 15, 43-50.
- Gureckis, T. M., & Goldstone, R. L. (in press). Thinking in groups. Pragmatics and Cognition.
- Hashimoto, C., Suzuki, S., Takenoshita, Y., Yamagiwa, J., Basabose, A.K., & Furuichi, T. (2003). How fruit abundance affects the chimpanzee party size: a comparison between four study sites. Primates, 44, 77-81.
- Helbing, D., Keltsch, J., & Molnár, P. (1997). Modeling the evolution of human trail systems. Nature, 388, 47-50.
- Hills, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. Cognitive Science, 30, 3-41.
- Howsman, T. G., O'Neil, D., & Craft, M. A. (2004). A stigmeric cooperative multi-robot control architecture. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. A. Watson (Eds.). Artificial Life IX. (pp. 88-93). Cambridge, MA: MIT Press.
- Janssen, M.A. and W. Jager (2003), Simulating Market Dynamics: The Interactions of Consumer Psychology and Structure of Social Networks, Artificial Life, 9, 343-356
- Kahan, J. P., & Rapoport, A. (1984). Theories of coalition formation. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kennedy, J., Eberhart, R. C., & Shi, Y. (2001). *Swarm intelligence*. San Francisco: Morgan Kaufmann Publishers.
- Kennedy, M., & Gray, R. D. (1993). Can ecological theory predict the distribution of foraging animals? A critical analysis of experiments on the ideal free distribution. Oikos, 68, 158-166.
- Kleinberg, J. (2000). Navigation in a small world. *Nature*, 406, 845.
- Kraft, J. R., & Baum, W. M. (2001). Group choice: The ideal free distribution of human social behavior. Journal of the Experimental Analysis of Behavior, 76, 21-42.
- Latané, B. & Bourgeois, M. J. (1996). Dynamic social impact in electronic groups: Consolidation and clustering in computer-mediated communication. *Journal of Communication*, 46, 35-47.
- Latané, B. & L'Herrou, T. L. (1996). Spatial clustering in the conformity game: dynamic social impact in electronic groups. *Journal of Personality and Social Psychology*, 70, 1218-1230.
- Lazer, D. & Friedman, A. (2005). The parable of the hare and the tortoise: small worlds, diversity, and system performance. KSG Working Paper No. RWP05-058. Available at: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=832627](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=832627).
- Madden, G. J., Peden, B. F., & Tamaguchi, T. (2002). Human group choice: Discrete-trial and free-operant tests of idea free distribution. Journal of the Experimental Analysis of Behavior, 78, 1-15.

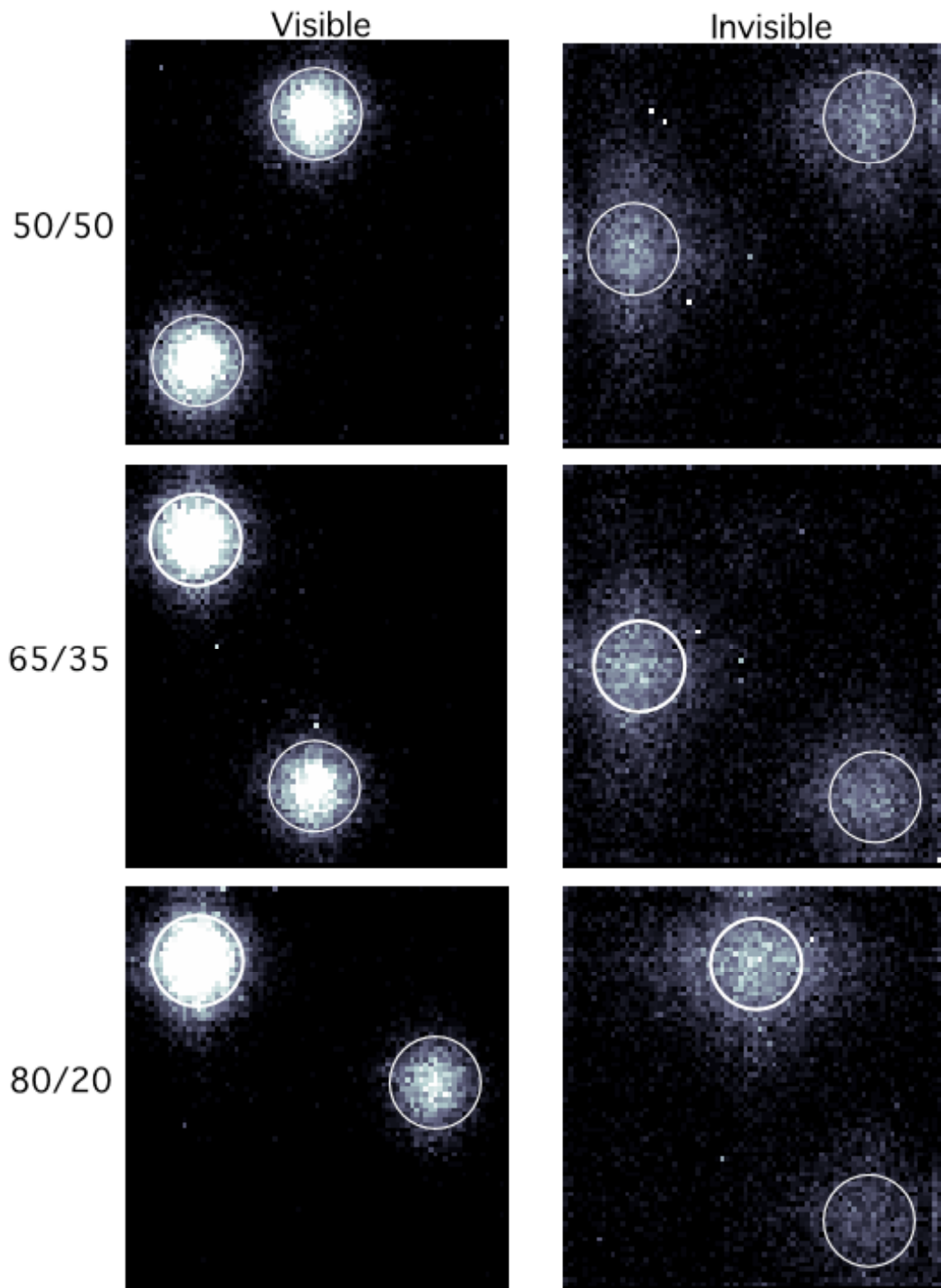
- Mason, W. A., Jones, A., & Goldstone, R. L. (in press). Propagation of innovations in networked groups. Journal of Experimental Psychology: General.
- Melzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, 24, 470-476.
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60-67.
- Milinski, M. (1987). Competition for non-depleting resources: The ideal free distribution in sticklebacks. In A.C. Kamil, J.R. Krebs, & H.R. Pulliam (Eds.), *Foraging Behavior*. NY: Plenum Press. (pp. 363-388).
- Newell, A., Simon, H. A. (1972): *Human problem solving*. Englewood Cliffs, NJ, USA, Prentice Hall.
- Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97, 362-376.
- Ostrom, E., Gardner, R., & Walker, J. (1994). Rules, games, and common-pool resources. Ann Arbor: University of Michigan Press.
- Parunak, H. V. D., & Brueckner, S. A. (2004). Engineering swarming systems. In F. Bergenti, M.-P. Gleizes, and F. Zambonelli (Eds.), Methodologies and Software Engineering for Agent Systems. Amsterdam: Kluwer (pp. 341-376).
- Pirolli, P., & Card, S. (1999). Information foraging. Psychological Review, 106, 643-675.
- Pöysä, H, Elmberg, J, & Sjöberg, K. (1998). Habitat selection rules in breeding mallards (*Anas platyrhynchos*): A test of two competing hypotheses. Oecologia, 114:283-287.
- Pulliam H. R, & Danielson, B. J. (1991). Sources, sinks, and habitat selection: a landscape perspective on population dynamics. American Naturalist, 137, S50-S66.
- Rapopoort, A., Kugler, T., Dugar, S., & Gisches, E. (under review). Choice of Routes in Congested Traffic Networks: Experimental Tests of the Braess Paradox.
- Roberts, M., & Goldstone, R. L. (2006). EPICURE: An agent-based foraging model. Artificial Life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems. Cambridge, MA: MIT Press (379-385).
- Roberts, M.E., & Goldstone, R.L. (in press). EPICURE: Spatial and knowledge limitations in group foraging. *Adaptive Behavior*.
- Rosenkopf, L., & Abrahamson, E. (1999). Modeling reputational and information influences in threshold models of bandwagon innovation diffusion. Journal of Computational and Mathematical Organizational Theory, 5, 361-384.
- Seth, A. K. (2001). Modeling group foraging: Individual suboptimality, interference, and a kind of matching. Adaptive Behavior, 9, 67-91.
- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, 187, 60.
- Strang, D., & Macy, M. (2001). In search of excellence: Fads, success stories, and adaptive emulation. American Journal of Sociology, 107, 147-182.
- 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, 105-114.

- Todd, P.M., and Gigerenzer, G. (2000). Simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23(5), 727-741.
- Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440-442.

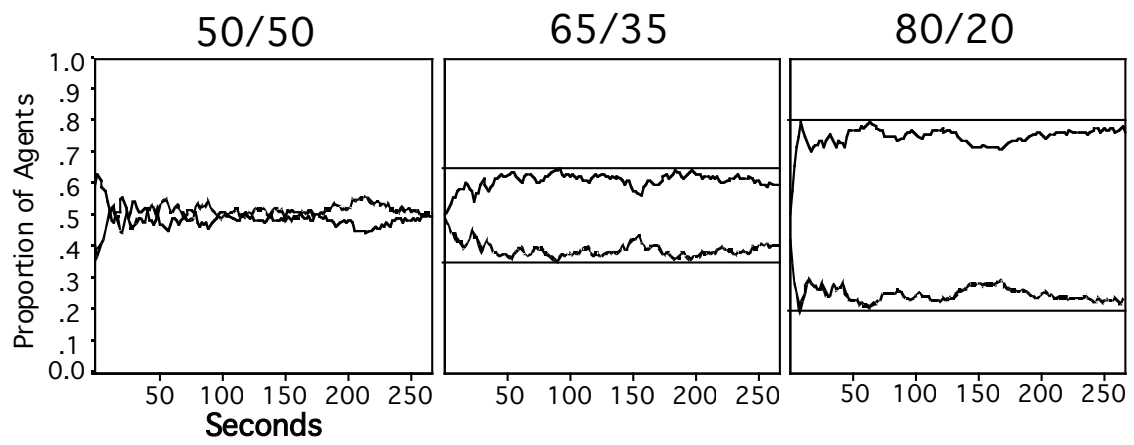
#### Author Notes

The authors wish to express thanks to Katy Borner, Rich Shiffrin, William Timberlake, Peter Todd, and Thomas Wisdom for helpful suggestions on this work. This research was funded by Department of Education, Institute of Education Sciences grant R305H050116, and National Science Foundation grant 0527920. More information about the laboratory and on-line versions of experiments related to the current work can be found at <http://cognitrn.psych.indiana.edu>.

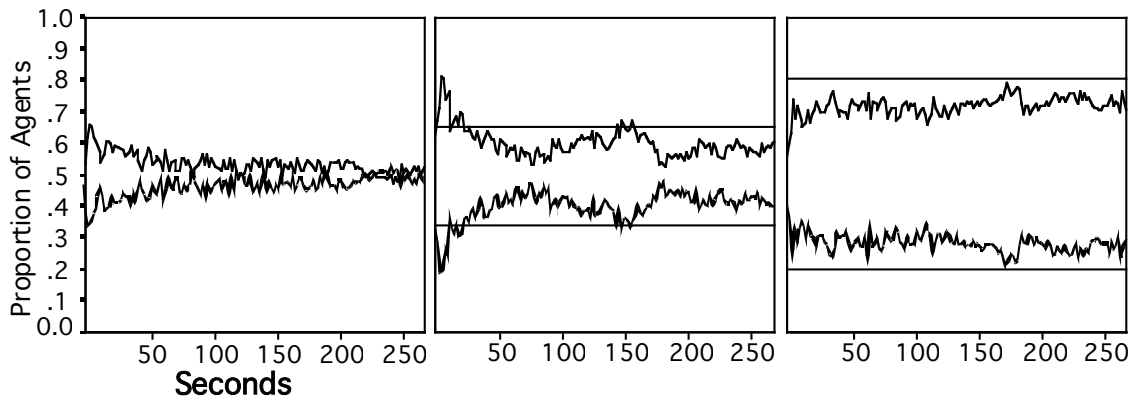
Correspondence concerning this article should be addressed to Robert Goldstone, Psychological and Brain Sciences, Indiana University, Bloomington, IN, 47405. Email: [rgoldsto@indiana.edu](mailto:rgoldsto@indiana.edu).



**Figure 1.** A frequency plot of participants' visits to each grid square in Experiment 1.

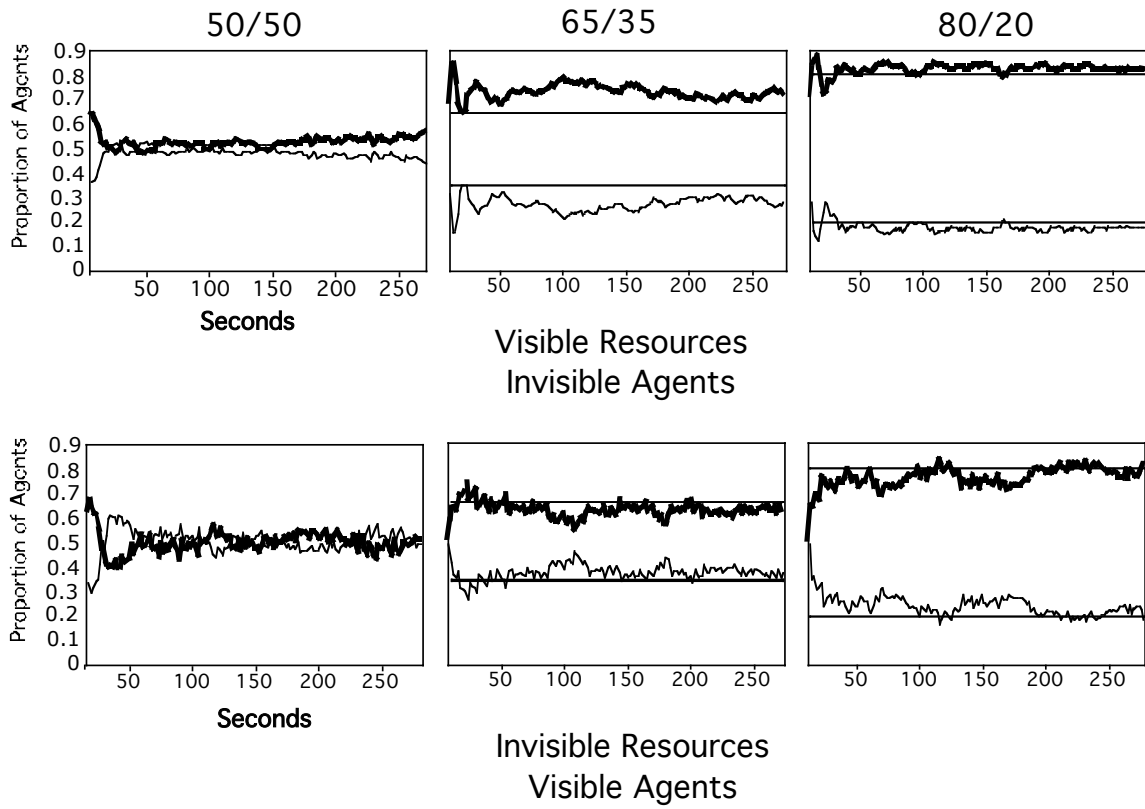


Visible



Invisible

**Figure 2.** Changes in group sizes over the course of a session in Experiment 1.



**Figure 3.** Changes in group sizes over the course of a session in Experiment 2. The thicker line represents the more productive resource pool.

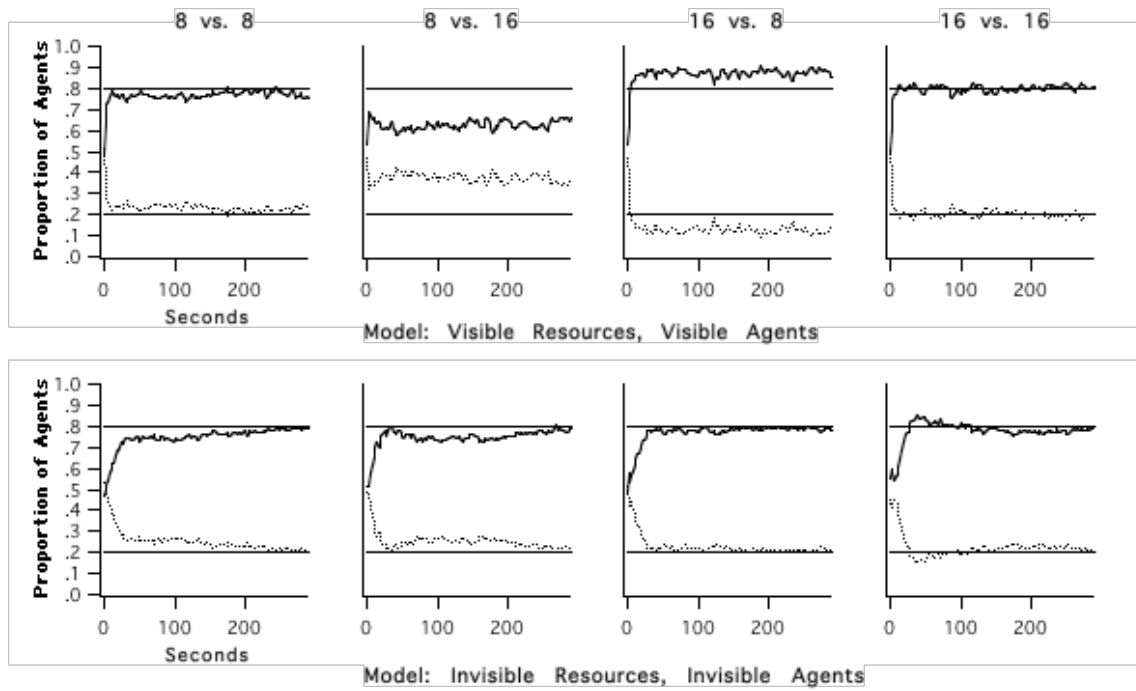
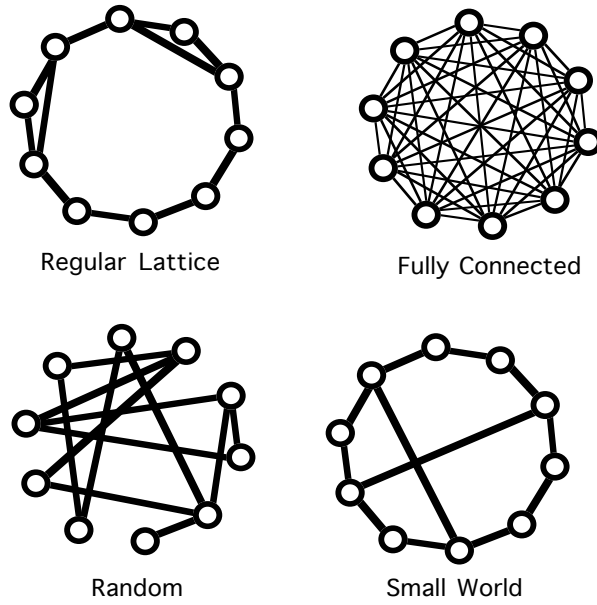
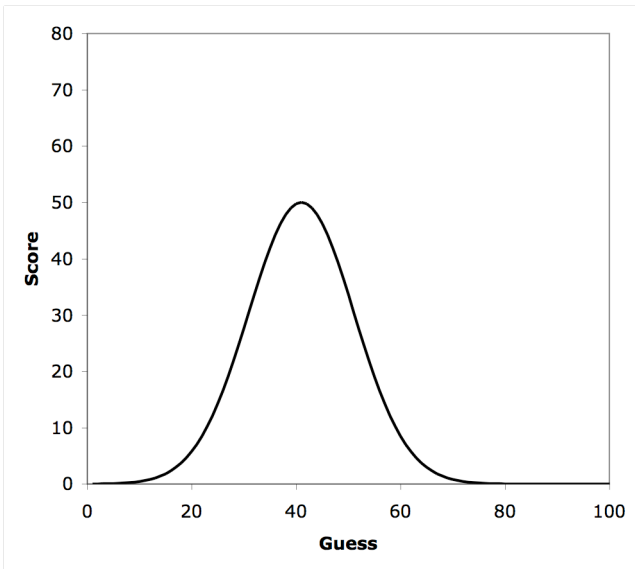


Figure 4: EPICURE matching results for the visible (top) and invisible (bottom) 80/20 food distribution conditions with different uniform variance pool sizes. In each condition, the first number at the top of the graphs indicates the radius of the 80% pool (square regions were used for simplicity), and the second number indicates the radius of the 20% pool. For example, 16 vs. 8 indicates that the 80% pool covers a total area of  $(16*2) * (16*2) = 1024$  cells while the 20% pool covers a total area of  $(8*2) * (8*2) = 256$  cells.



*Figure 5.* Examples of the different network structures for groups of 10 participants. Circles represent participants and lines indicate communication channels.

### Unimodal Fitness Function



### Multimodal Fitness Function

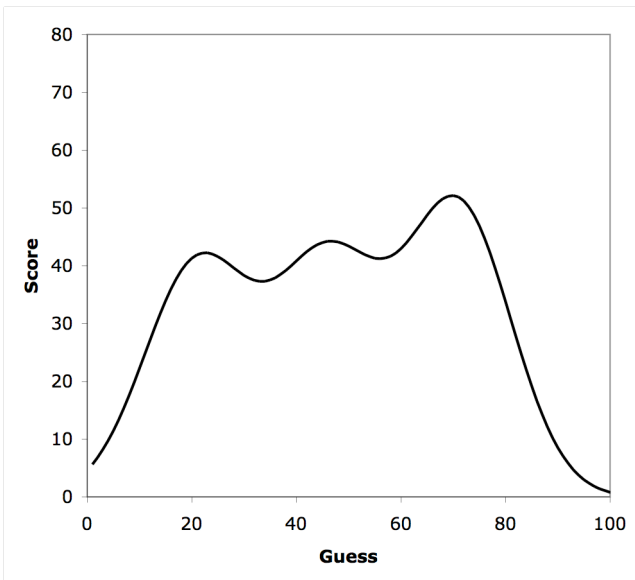


Figure 6. Examples of the unimodal and multimodal fitness functions that convert guesses into obtained points.

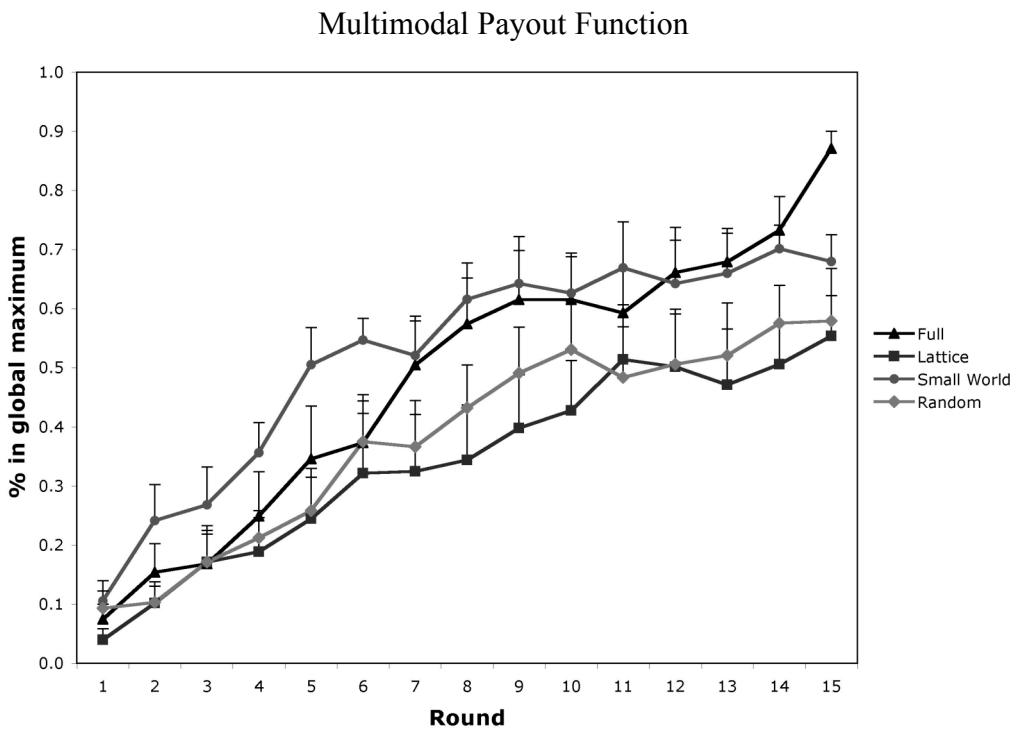
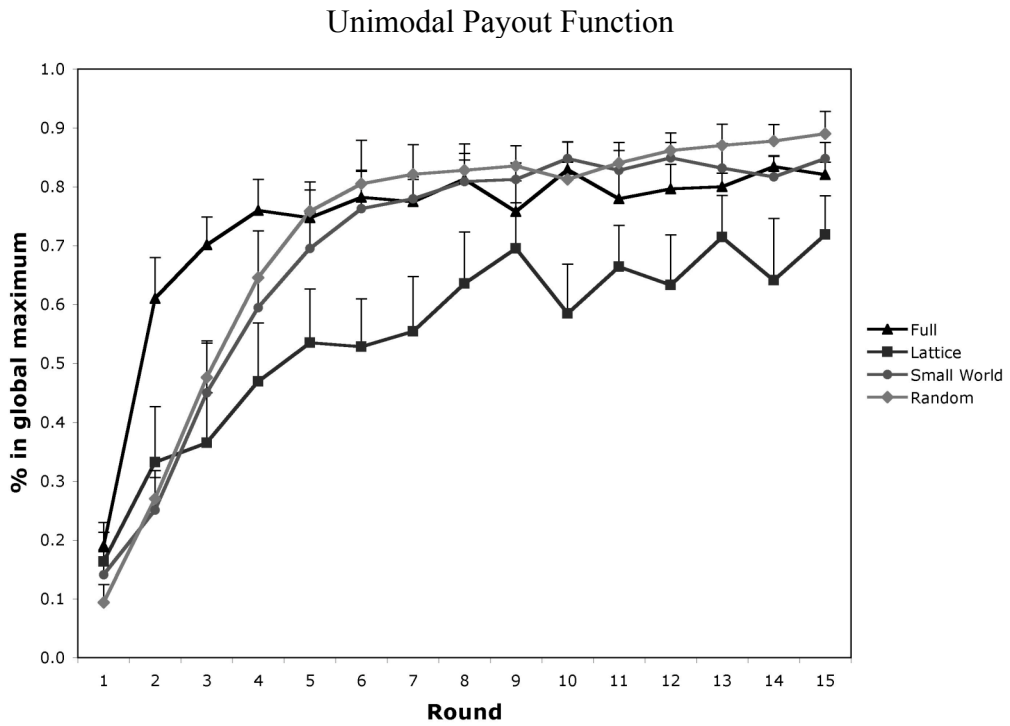


Figure 7. Percentage of participants within one standard deviation of the global maximum on each round for the unimodal and multimodal payout functions.

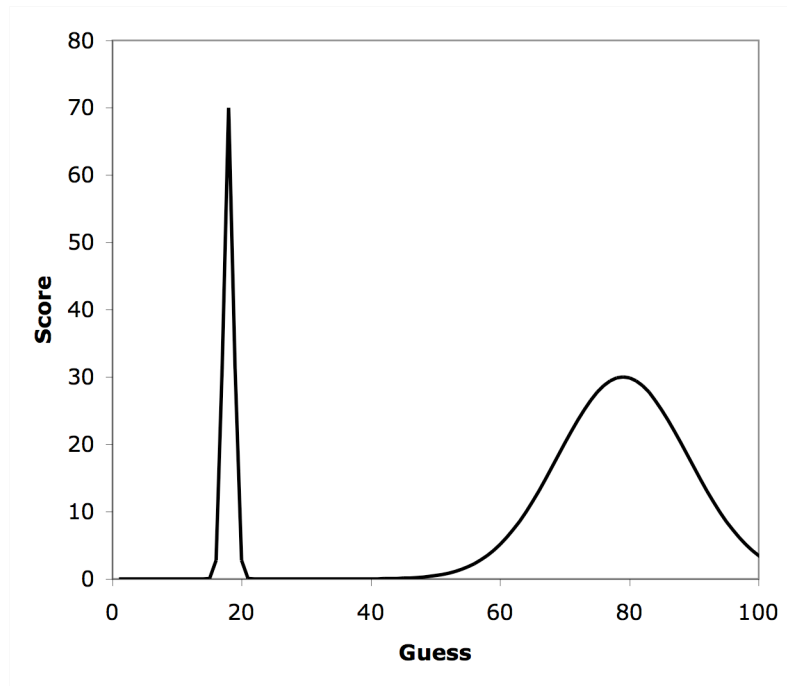


Figure 8. An example of the “needle” payout function. This function features one broad local maximum that is easy to find and one narrow global maximum that is difficult to find.

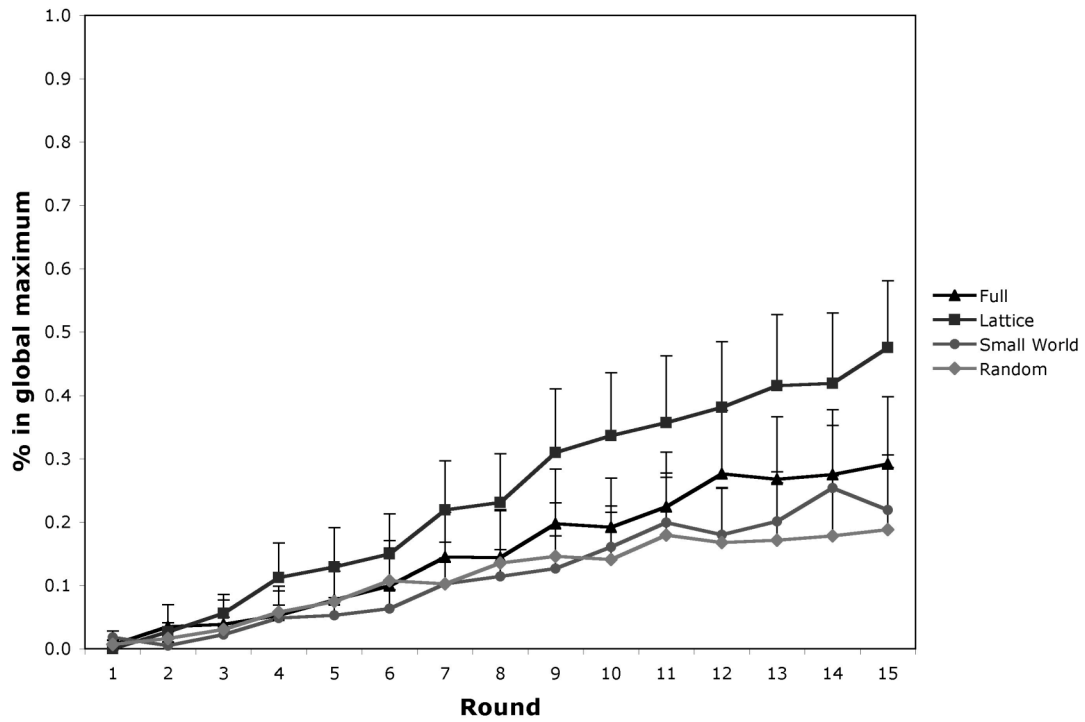


Figure 9. Performance for the four network structures with the Needle payout function. For this function, the lattice network performs better than the other three network types.

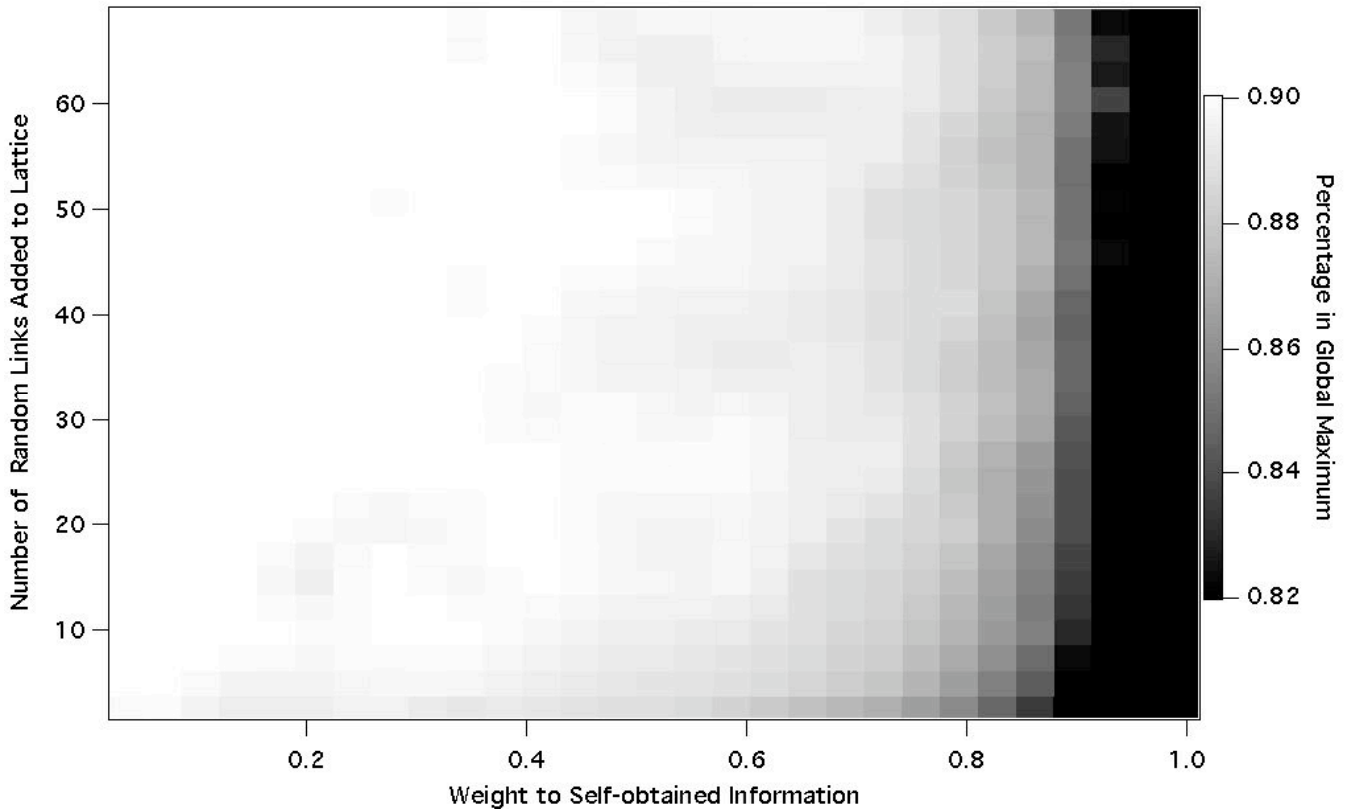


Figure 10. Group performance for the unimodal function. This graph shows the interaction between the bias for self- versus other-obtained information and the number of random links added to a regular lattice. Group performance is measured by the percentage of individuals within one standard deviation of the global maximum of the fitness function. The brightness of each square indicates the group's performance after 15 rounds of number guessing. For this simple problem space, group performance increases monotonically with increased reliance on others' information and network connectivity.

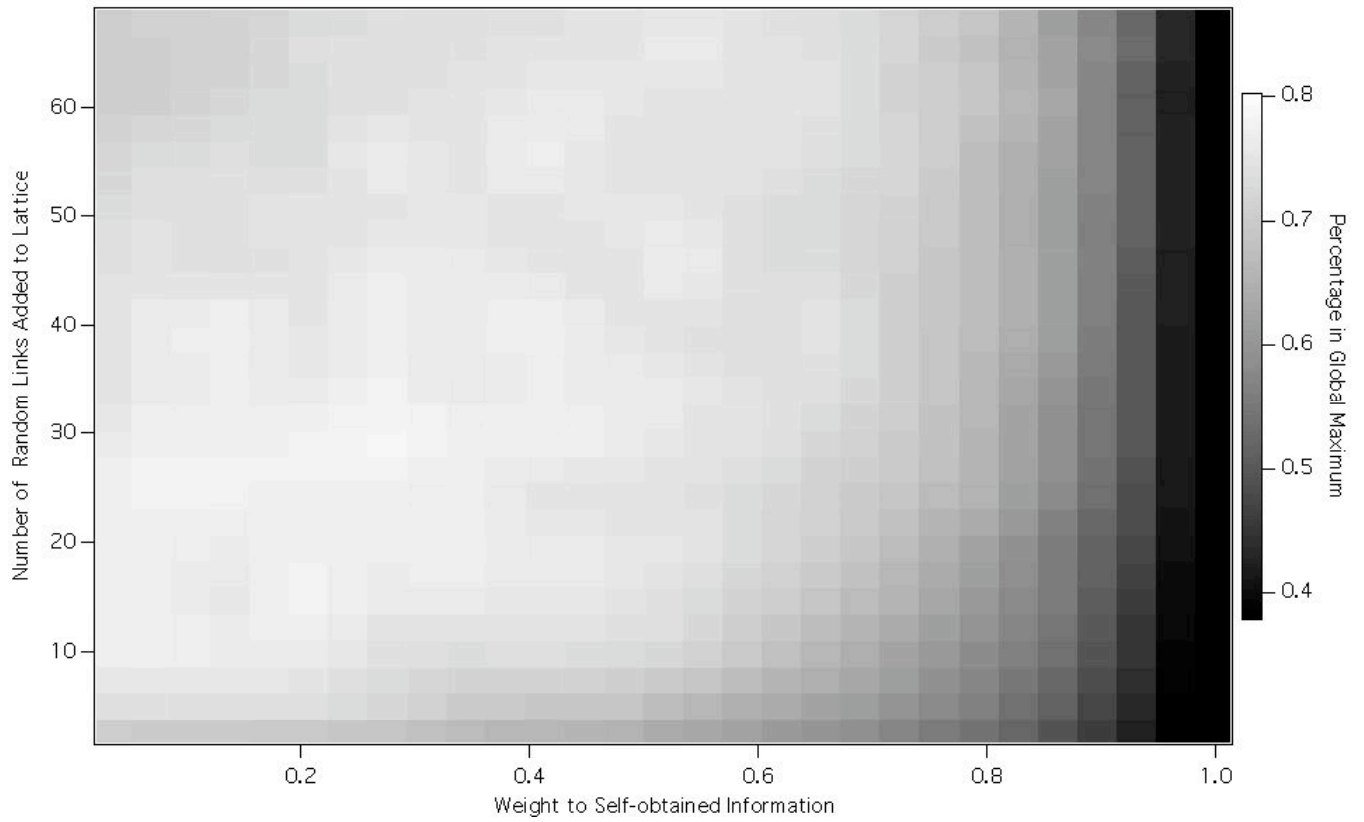


Figure 11. Group performance for the multimodal function. The best performance is found for a combination of using self- and other-obtained information, and for intermediate levels of network connectivity.

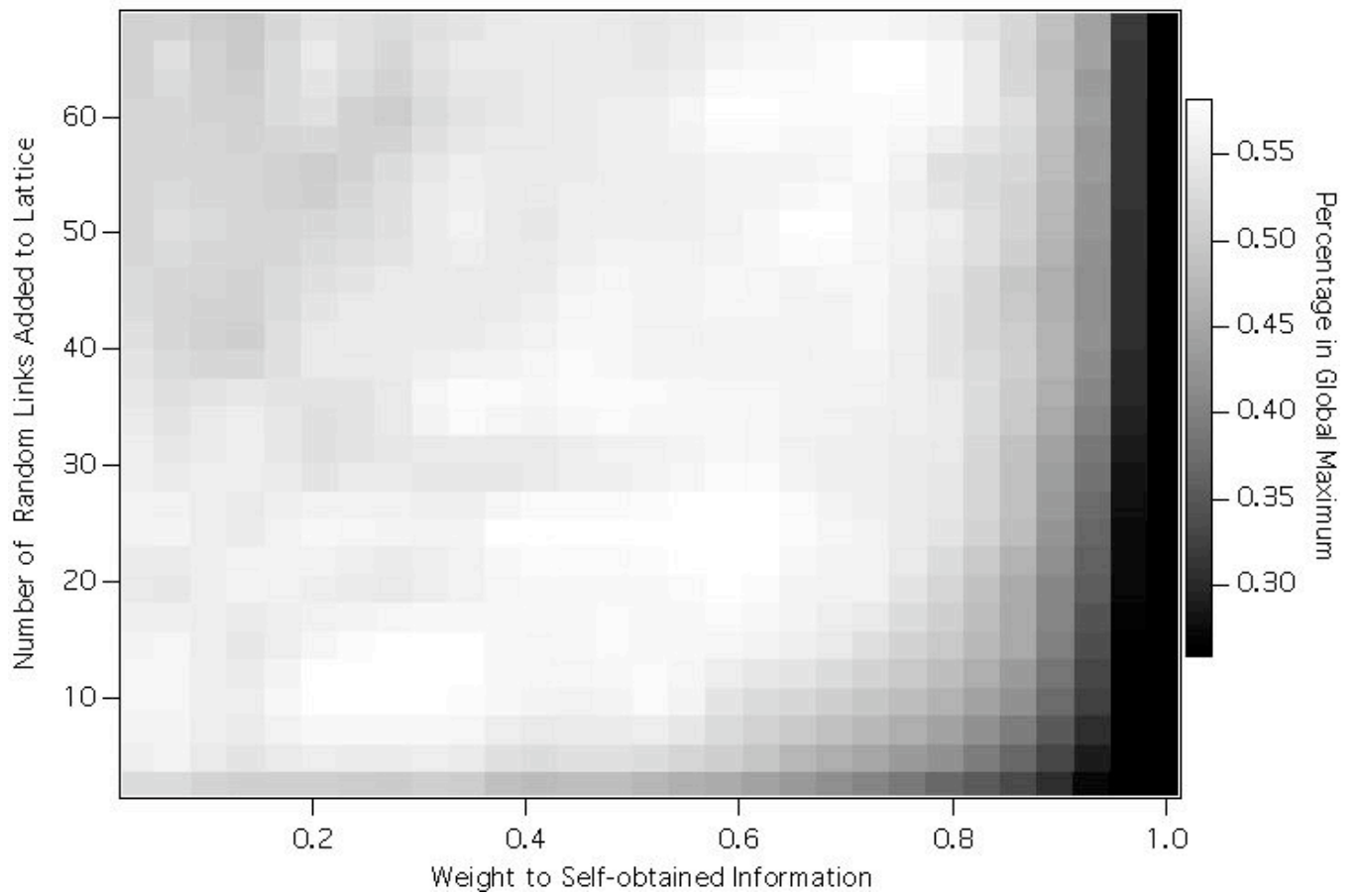


Figure 12. Group performance for the needle function. This function benefits from even greater reliance on self-obtained information and decreased global network connectivity.