Complex exploration dynamics from simple heuristics in a collective learning environment

Sabina J. Sloman (SSLOMAN@Andrew.Cmu.Edu)
Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Robert L. Goldstone (RGOLDSTO@Indiana.Edu)
Department of Psychological and Brain Sciences and Program in Cognitive Science, Indiana University, 1101 E. 10th St.
Bloomington, IN 47405 USA

Cleotilde Gonzalez (COTY@Cmu.Edu)
Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Abstract

Effective problem solving requires both *exploration* and *exploitation*. We analyze data from a group problem-solving task to gain insight into how people use information from past experiences and from others to achieve explore-exploit trade-offs in complex environments. The behavior we observe is consistent with the use of simple, reinforcement-based heuristics. Participants increase exploration immediately after experiencing a low payoff, and decrease exploration immediately after experiencing a high or improved payoff. We suggest that whether an outcome is perceived as "high" or "low" is a dynamic function of the outcome information available to participants. The degree to which the distribution of observed information reflects the true range of possible outcomes plays an important role in determining whether or not this heuristic is adaptive in a given environment.

Keywords: exploration; exploitation; networks; social learning

Introduction

Search—a dynamic maximization problem where outcomes depend on the agent's location in the problem space—is a fundamental part of our cognitive experience (Hills et al., 2015). When in a new city, we sample from different restaurants in order to find the best places to eat (Mehlhorn et al., 2015). When coming up with a new idea for a research project, the amount of intellectual and social "reward" we expect to experience is a function of whether the point in conceptual space we're interested in is novel and appreciated by others.

Effective search requires both *exploration*, or sampling from the space of outcomes to gain information about what's available, and *exploitation*, or taking advantage of the information available and resampling from places known to produce good outcomes. Should the traveller stick with the first decent restaurant she finds, or keep exploring her options? Should the scientist stick with her current line of work, or branch out into unchartered intellectual territory?

We analyze data from a group problem-solving task to gain insight into how participants use information from past experiences and from others to achieve explore-exploit trade-offs in rugged, networked environments. When the world is uncertain, complex and interconnected, the optimal trade-off between exploration and exploitation depends on the degree of complexity, on the extent of interconnectedness—and on the strategies individuals adopt to process and act on the information they encounter (Barkoczi, Analytis, & Wu, 2016; Barkoczi & Galesic, 2016; Toyokawa, Whalen, & Laland, 2019). In some cases, we may adapt our exploration level to the environment we're in, even when the shape of the environment is unknown to us (Mason & Watts, 2012).

We add to existing work that has looked at behavioral patterns of exploration in different environments, and examine the mechanisms that lead to the individual- and group-level patterns we observe. Our contributions are both methodological and theoretical. From a methodological perspective, we specify a generalization gradient and propose it as a useful measure of both individual- and group-level exploitation in smooth search spaces. From a theoretical perspective, we document exploration patterns and systematic behavioral responses to outcome information. We find that context-dependent explore-exploit trade-offs emerge even when participants are not told what kind of environment they're in, and speculate that differences in exploration patterns can be explained by differences in the outcome information available to participants.

Methods

Experimental paradigm

We analyzed data from the group search task designed and implemented by Mason, Jones, and Goldstone (2008). Each participant guessed numbers between 0 and 100 and a computer revealed to them how many points were obtained from the guess by consulting a hidden *fitness function*[1] that translates a guess into a number of points. Random noise was added to these points so that repeated sampling was necessary to accurately determine the underlying function relating guesses to scores. On each trial, a group of participants was assigned to one of several conditions (discussed below). Trials consisted of 15 rounds, over which each member of the

[1] We will use the terms "fitness function" and "fitness landscape" interchangeably.

group tried to maximize their total number of earned points. Importantly, on each round, participants got feedback not only on how well their own guess fared, but also had access to information about the actions and outcomes of their neighbors.

Two aspects of the environment were experimentally manipulated: the social network (the network topology that determines who counts as neighbors) and the complexity of the task (the shape of the fitness function that converted guesses to earned points). These are discussed in the sections below.

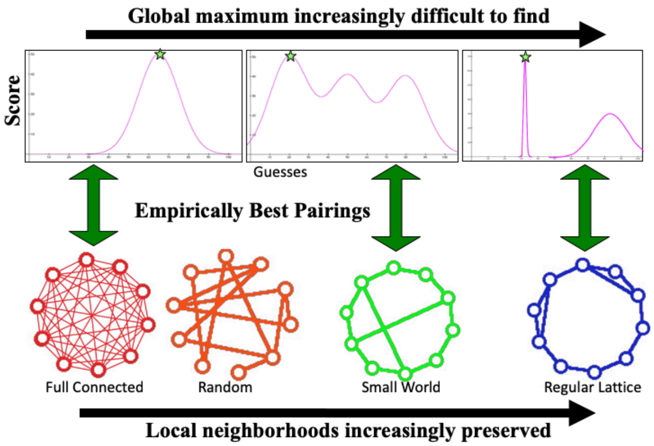
Global maximum increasingly difficult to find
Score
Guesses
Empirically Best Pairings
Full Connected
Random
Small World
Regular Lattice
Local neighborhoods increasingly preserved

Figure 1: The network structures and fitness functions used in Mason et al. (2008). Reproduced from Goldstone et al. (2013) with permission of the authors.

Social network structure Neighborhoods of participants were created to reflect *random*, *regular lattice*, *small world*, or *fully connected* networks. Examples of the graph topologies for groups of 10 participants are shown in Figure 1. In the random graph, connections are randomly created under the constraint that the resulting graph is connected. Participants in random graphs tend to be connected to others via relatively short paths.

The regular lattice configures 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": 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. The paths connecting people are much longer, on average, in lattice than in random graphs.

"Small world networks" have both cliques and a short average path length (Watts & Strogatz, 1998). From an information processing perspective, small-world networks are attractive 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).

A fourth network, a fully connected graph, allowed every participant to see the guesses and outcomes of every other participant.

	Full	**Small world**	**Random**	**Lattice**	**Total**
Unimodal	11	11	19	11	**52**
Trimodal	9	12	20	11	**52**
Needle	28	27	18	28	**101**
Total	**48**	**50**	**57**	**50**	**205**

Table 1: Number of trials of each condition in our data.

Environmental complexity Three hidden fitness functions for converting guessed numbers to points were tested across two experiments. The *unimodal* function has a single best solution that can eventually be found with a hill-climbing method. 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 search might not find the best possible solution. Finally, the *needle* function has one very broad local maximum, and one hard-to-find global maximum.[2] The height and variance of the global maximum in the unimodal conditions, global maximum in the trimodal conditions, and local maximum in the needle conditions are all equal (the height of these peaks is 50, while the height of the needle's global maximum is 70).

After excluding some trials due to apparently incomplete data, we used 205 trials in total for our analyses. The number of trials in each conditions is reported in Table 1. The number of players in each trial ranged from 5 to 19, with a mean of 11.89 ($SD = 4.05$).[3]

Measuring exploration

To measure the degree to which a sequence of guesses exploited a location of the search space (or, conversely, didn't explore the space), we developed a similarity metric (hereafter referred to as *similarity*) that captures the average degree of closeness of all pairwise combinations of the elements of a set of guesses along a generalization gradient[4] adapted to the specific problem space:

$$similarity(G_i, G_j) = e^{-(\frac{G_i - G_j}{c})^2}$$

[2] Mason et al. (2008) collected data on variations of the needle function in two separate experiments. In our analyses, when referring to the needle conditions we pooled data from the two experiments.

[3] Each group of participants was assigned to several conditions in sequence (for more details on the experimental procedure, see Mason et al. (2008)). We consider a "trial" to be uniquely specified by a combination of a group and a condition. In other words, if a group completed the task in n conditions, this is recorded in Table 1 as n distinct observations.

[4] A *generalization gradient* is a function that transforms distance in some space to distance in another—usually more psychologically interesting—space.

where $c = .07$ was set to reflect the variance of the global maxima on the unimodal and trimodal landscapes, and the local maxima on the needle landscapes. The total average similarity of a group of guesses G is

$$similarity(G) = \frac{\sum_{i,j} similarity(G_i, G_j) - n}{n^2 - n}$$

where $n = |G|$. We use $1 - similarity(G)$ as our measure of the degree to which G spans—or explores—the problem space.

While other measures, such as variance or the average volatility measure developed by Mason et al. (2008), capture the average distance between guesses, they do not directly capture the idea of the extent to which a set of guesses spans the problem space. Consider a participant A who alternates between guessing 0 and 100, and a participant B who guesses a number at every multiple of 10. We'd like to say that B is the better explorer, because their guesses are spread across the landscape. However, the variance and average volatility of participant A's guesses are much higher than the variance and volatility of B's guesses. By taking the average of *all* pairwise combinations of guesses, our similarity metric captures the *spread* of guesses, rather than simply the extent of their range.

In addition, our metric captures the intuition that similarity drops off steeply with the distance between two nearby solutions, but quickly flattens out (see Figure 2). The choices to jump 99 or 100 units away from where one is are considered effectively identical, while the choices to jump 0 or 1 unit away are much less similar.[5]

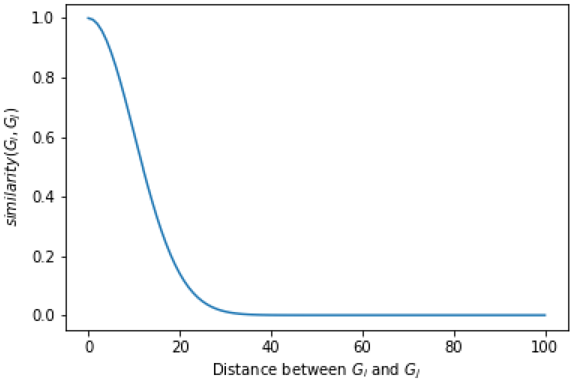

Figure 2: The shape of the generalization gradient underlying the *similarity* metric.

Some evidence suggests that a gradient of this form is a good approximation of how people make inferences about unseen locations in spatial search tasks (Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018). However, here we invoke *similarity* only to operationalize the degree of "exploratoryness" of a set of guesses, not to model participants' inferences.

Exploration patterns

Individual exploration

Figure 3 shows the heterogeneity in the amount of exploration between participants. Higher density on the right side of the histograms indicates that participants in that condition tended to distribute their guesses more evenly across the problem space.

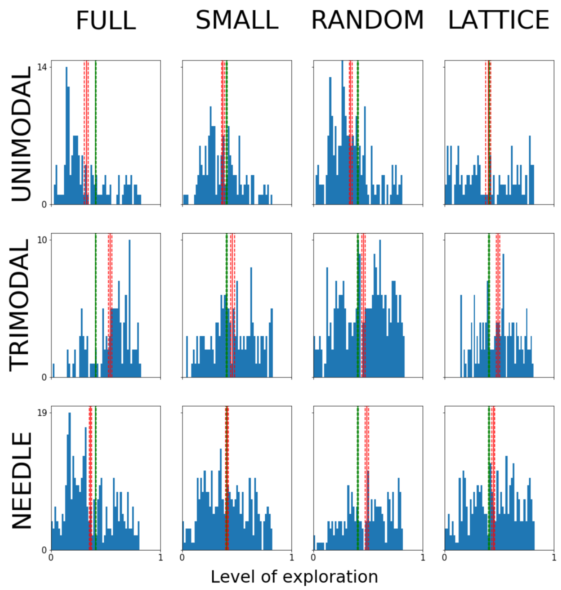

Figure 3: Histogram of participant-level exploration levels. A participant i who makes a sequence of guesses G_i has an exploration level equal to $1 - similarity(G_i)$. The green lines indicate the global mean and standard error of exploration levels across participants in all conditions (.408 (SE = .004)). The red lines indicate the mean and standard error of exploration levels across all participants *within* the respective condition. If a person participated in several conditions, they are treated as a separate participant in each condition.

Individuals tend to explore less than average on the unimodal landscapes, which were explicitly constructed so that their global maxima were easy to find. When the best solution can be found with very little exploration, more extensive search may just lead to foregone payoffs rather than valuable information. We tested this intuition by looking at the correlation between participants' exploration levels and their average payoffs. As expected, this correlation is much lower on the unimodal landscapes than on the more complex landscapes.

[5] The Gaussian shape of the fitness functions is compatible with the Gaussian similarity drop-off gradient we used. While this captures many of the same intuitions, it differs from the well-known exponential similarity function (Shepard, 1987). All our results are robust to the use of an exponential similarity function.

Exploration levels tend to be lower than average in one other condition: the fully-connected network on the needle landscape. Mason et al. (2008) found that when confronted with the difficulty of the needle landscape, participants tended to do better when in the sparsely-connected lattice network (see Figure 1). They speculated that this was because distributing social information hindered bandwagoning, or collective convergence on the tempting local maximum. Our results corroborate this speculation: While participants in the fully-connected networks explore the needle landscape less than average, the mean exploration level in the lattice networks is higher than the global average.

Collective exploration

The similarity metric allows us to calculate the "exploratoryness" of an arbitrary sequence of guesses. In particular, we can also use it to measure *group-level*, or collective, exploration.

Figure 4 shows the evolution of collective exploration over rounds, alongside the proportion of participants who were within one standard deviation of the global maximum on each round. Collective exploration declines quickly on the unimodal and needle landscapes. While this coincides with more participants finding the global maximum on the unimodal landscape, the proportion of participants who find the global maximum in the needle condition remains relatively low. These patterns reflect dynamics analogous to the individual-level patterns we discussed in the previous section: The group explores less when there is a salient local maximum, and especially so when outcome information is rapidly broadcast throughout the network.

The consequences of early exploration

Our explanations for many of the results in the previous sections depend on our assumption that exploration is more important in some cases than in others. In some environments, low exploration may cause high payoffs; quick convergence on promising areas of the landscape may cause the average payoff to rise. In others, maintaining a high amount of exploration and broadly surveying the problem space could lead to subsequently higher payoffs. This section further unpacks the sequentially contingent relationship between exploration and expected reward in the different conditions.

Figure 5 plots the cross-correlations between average payoffs and the collective exploration level within a round. It's unsurprising that all the correlations are below zero; as shown Figure 4, collective exploration subsides while payoffs increase over time. More informative for our purposes is the *difference* between the correlation of early exploration with later payoffs, and the correlation of early payoffs with later exploration. An interpretation that exploration causes higher downstream payoffs would require that the former be higher than the latter. The insets of the plots shows this difference for each condition. When the blue line is above zero, this indicates that exploration now is more highly correlated with payoffs later, than payoffs now are with exploration later.

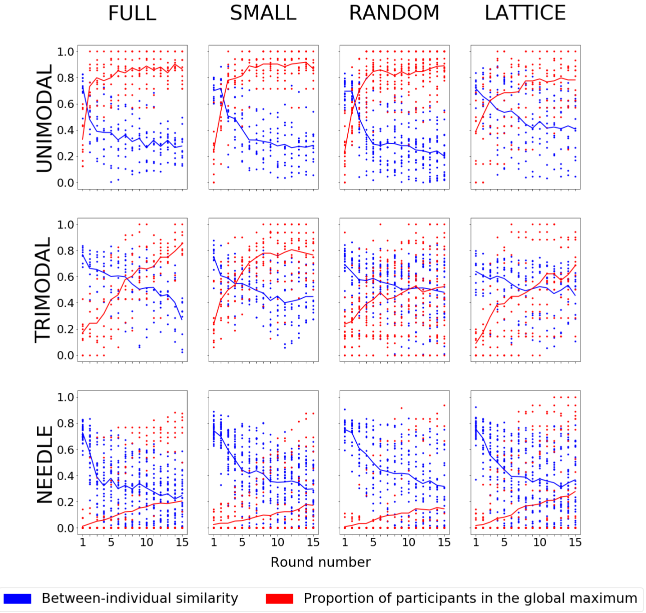

Figure 4: Collective exploration levels (blue) and proportion of players within one standard deviation of the global maximum (red) over rounds. Each dot represents one trial. On round t, group k has an exploration level equal to $1 - similarity(G_{k,t})$ where $G_{k,t}$ is the set of all guesses the members of group k made on that round. The blue line plots the mean exploration level across trials, and the red line plots the proportion of all players across trials who were within one standard deviation of the fitness function's global maximum.

The positive trend in the inset is most consistent across the needle and regular lattice conditions. When connectivity is low and finding the global maximum is especially difficult, group-level exploration leads to higher downstream payoffs. While this pattern is also consistent with an account that higher payoffs cause quicker collective convergence, the "exploration leads to downstream payoffs" account has the advantage that it predicts the particularly strong positive trend for the needle landscape, which is explicitly designed so that the global maximum is hard to find without considerable exploration.

WSLS: Win-shift-less, lose-shift-more

Win-stay, lose-shift (WSLS) is a heuristic applicable to search tasks by adaptive biological and artificial systems. The rule is simple: When you're successful, stay close to where you currently are. When you're unsuccessful, move further away (Bonawitz, Denison, Gopnik, & Griffiths, 2014; Nowak & Sigmund, 1993; Robbins, 1952).

WSLS is usually applied in contexts with discrete binary outcomes that can be easily dichotomized into wins and losses. However, the problem space facing the currently considered participants, like many real-world problem spaces, is both smooth—similarity of actions predicts similarity of

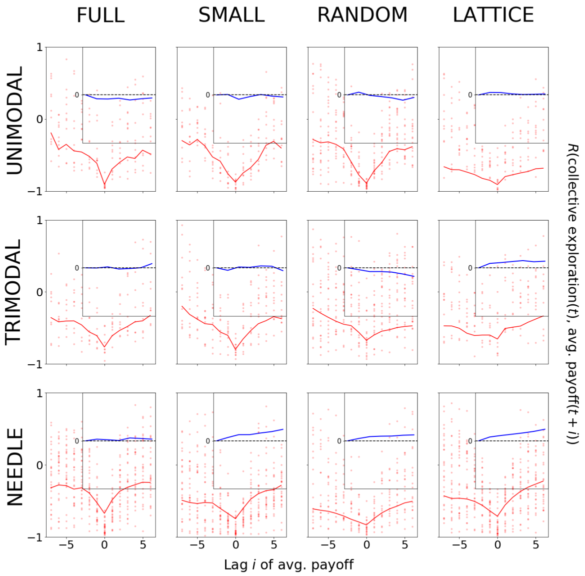

Figure 5: Cross-correlations between group exploration level and payoffs. On round t, group k has an exploration level equal to $1 - similarity(G_{k,t})$ where $G_{k,t}$ is the set of all guesses the members of group k made on that round. Payoff observations are the average payoff participants experienced on round t. A lag of i on the x-axis indicates the correlation between group exploration level at time t and average payoffs at time $t+i$. Each dot corresponds to the correlation using the data from one trial. The red lines plot the correlations across trials. Insets show the difference between the cross-correlation coefficients at lag i and lag $-i$ for $0 \leq i \leq 7$.

outcomes—and continuous. In this section, we show that participant behavior is consistent with a generalization of WSLS: The degree to which participants stray from promising locations varies with both the absolute and relative amount of reward they've experienced there. Participants shift less when they win, and shift more when they lose.

In many situations, WSLS or close variants can lead to approximately optimal search behavior on intractable problem spaces (Bonawitz et al., 2014; Robbins, 1952). To the participants facing the task at hand, the range of possible outcomes is unknown. We suggest that they dynamically incorporate outcome information into their understanding of what's a "win" and what's a "loss". When good outcomes are easy to find, the outcome information participants accumulate accurately reflects the range of attainable payoffs. In these cases, the application of WSLS-like rules may lead to adaptive explore-exploit trade-offs. But when the best outcomes are difficult to find, participants do not get full outcome information about the range of possible payoffs. They fail to appropriately calibrate their "shift-more" and "shift-less" responses. On the needle landscape, WSLS-like rules may lead participants to prematurely converge on the local maximum. In short, when good outcomes are hard to find, information flow is reduced, and individuals cannot appropriately tune their behavior to the relevant search space, resulting in suboptimal individual- and group-level outcomes.

Absolute "wins": Responses to high payoffs

Figure 6 shows how the similarities between participants' preceding guesses (blue) and subsequent guesses (red) covary with the payoffs they experience. Recall that the similarity of two guesses is a measure of the closeness of the guesses. If a participant's guesses on round t and round $t+1$ are more similar than their guesses on round t and round $t-1$, we say they are *exploiting* more on round $t+1$ than on round t.

In all conditions, there is some payoff value above which participants tend to exploit more than explore. The blue vertical lines mark the normalized payoff values where the trend in participants' future level of convergence dips below their past level of convergence—participants begin to shift more (explore). The red vertical lines indicate payoff values where the reverse switch occurs—participants begin to shift less (exploit). In general, participants shift more when payoffs are low, and shift less when payoffs are high.

Where this switch occurs varies by landscape. We speculate that these differences are a direct effect of differences in the outcome information available to participants, and how they adjust their beliefs about the range of possible payoffs based on their observations (Parducci, 1965). In the trimodal conditions, the "switch point" is shifted to the right (participants wait for relatively high payoffs before they begin to settle down), but so is the density of payoff observations. As shown in Figure 1, payoffs on the trimodal landscape remain relatively high even when participants stray from the global maximum. When they observe that locations that lead to "wins" are distributed widely across the landscape, participants are more reluctant to settle down.

By contrast, in the needle conditions, both the switch point and the bulk of the density is shifted towards the left side of the plot. Few participants stumble upon the narrow global maximum, and most experienced payoffs are a smaller proportion of the highest possible payoff. The emergent patterns resemble those in the unimodal conditions because most participants do not have outcome information to suggest that they *are not* on a well-behaved landscape with a similar payoff distribution.

Relative "wins": Responses to improving payoffs

Figure 7 shows how the similarity of participants' guesses changes as a function of the difference between their most recently experienced payoffs. Points to the right of the y-axis correspond to instances where a player had just experienced an immediate increase in payoff. Points above the x-axis correspond to instances where the player's round-to-round exploration level decreased. In general, immediate gains lead to convergence, and losses lead to continued exploration.

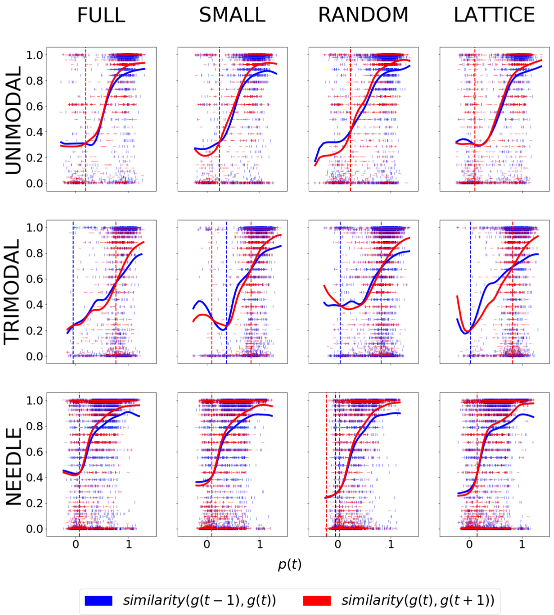

Figure 6: Experienced payoffs against similarity of guesses. $p(t)$ denotes the payoff a player experienced at time t (normalized by the height of the global maximum in each condition), and $similarity(g(t), g(t'))$ denotes the similarity between a guess made at time t and a guess made at time t'. Each dash corresponds to one experienced outcome. Solid lines show the estimated Gaussian kernel regressions. Vertical lines mark shifts between exploitation and exploration (see text).

Note the inverted U-shaped trend recovered by the kernel regression across the needle landscapes. Participants who have just experienced an exceptionally large improvement tend to shift more than those who have just experienced a moderate improvement. This is consistent with our understanding of WSLS as a dynamic process: Participants who stumble upon the global maximum dynamically adjust their understanding of the range of possible payoffs, and are less willing than before to settle with what they have.

Discussion

We argued that the behavioral patterns we observe are consistent with the application of a dynamic, continuous variant of win-stay, lose-shift. While participants tend to "stay" in areas where they've experienced both high and improving payoffs, they use information from themselves and others to adapt their willingness to "stay" and "shift" to their environment.

One phenomenon we have only briefly addressed is Mason et al. (2008)'s finding that participants in the lattice network were more likely than participants on other networks to find the needle landscapes' global maxima. Our central claim is that reduced information flow can lead to suboptimal outcomes when participants do not have full information about the range of possible payoff values. Why would the network that restricted information flow the most perform the best when the search task is especially hard?

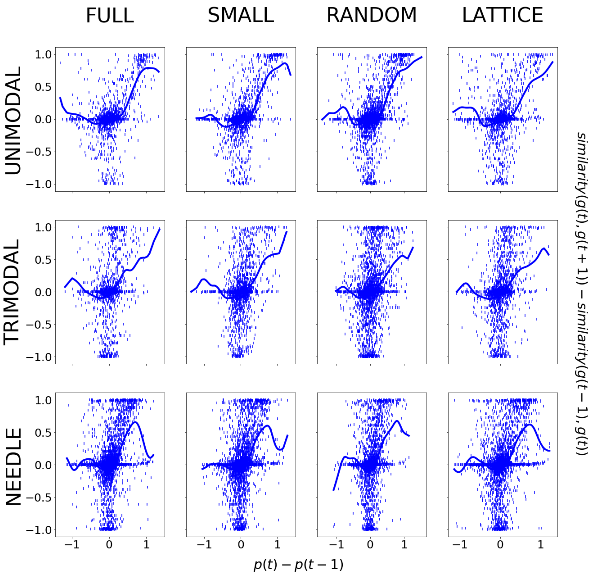

Figure 7: *Differences in* experienced payoffs against *differences in* similarity of guesses. Notation is the same as in Figure 6. Each dash corresponds to one experienced outcome. Solid lines show the estimated Gaussian kernel regressions.

Visual inspection of Figure 6 suggests that the payoff values at which participants switch from exploration to exploitation do not vary much as a function of the network structure, but depend more on the underlying fitness function. The lattice network's structural restriction of information flow could mean that it takes participants even longer to reach their threshold value or "switch point". Participants may search longer for "wins", resulting in more exploration where it matters the most.

While our analyses were motivated by our desire to understand the relationship between individual- and group-level exploration dynamics, we did not assume that participants make this trade-off explicitly. Rather, we suggested that participants may be using simple heuristics from which an apparent trade-off emerges. We adopted an information processing framework (Oppenheimer & Kelso, 2015): The environment affects behavior and outcomes via its effect on the information group members receive and broadcast to others. By analyzing participant behavior through the lens of *information flow*, we can come closer to understanding what determines the search conditions under which we do well, and the conditions under which we could do much better.

Acknowledgements

Cleotilde Gonzalez was supported by the Army Research Office, Network Science Program, Award Number:W911NF1710431.

Author note

A limited version of this work will appear as an abstract in the proceedings for ACM Collective Intelligence 2019.

References

Barkoczi, D., Analytis, P., & Wu, C. M. (2016). Collective search on rugged landscapes: A cross-environmental analysis. In *Proceedings of the 38th annual conference of the cognitive science society.*

Barkoczi, D., & Galesic, M. (2016). Social learning strategies modify the effect of network structure on group performance. *Nature Communications*, *7*.

Bonawitz, E., Denison, E., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference [Journal Article]. *Cognitive Psychology*. doi: 10.1016/j.cogpsych.2014.06.003

Goldstone, R., Wisdom, T., Roberts, M., & Frey, S. (2013). Learning along with others. *Psychology of Learning and Motivation*, *58*, 1–45.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & the Cognitive Search Research Group. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*(1), 46–54.

Kleinberg, J. (2000). Navigation in a small world. *Nature*, *406*.

Mason, W., Jones, A., & Goldstone, R. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, *137*, 422–433.

Mason, W., & Watts, D. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*(3), 764–769.

Mehlhorn, K., Newell, B. R., Lee, M., Morgan, K., Braithwaite, V. A., Hausmann, D., ... Gonzalez, C. (2015). Unpacking the explorationexploitation tradeoff: A synthesis of human and animal literatures [Journal Article]. *Decision*. doi: 10.1037/dec0000033

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, *364*.

Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making [Journal Article]. *The Annual Review of Psychology*, *66*, 277-294. doi: 10.1146/annurev-psych-010814-015148

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*(6), 407-418.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*, 527-535.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science [Journal Article]. *Science*, *237*(4820), 1317-1323.

Toyokawa, W., Whalen, A., & Laland, K. N. (2019). Social learning strategies regulate the wisdom and madness of interactive crowds. *Nature Human Behavior*, *3*, 183-193.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*.

Wu, C., Schulz, E., Speekenbrink, M., Nelson, J., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behavior*, *2*, 915-924.