A Social Interpolation model of group problem-solving

Sabina J. Sloman[1*], Robert L. Goldstone[2] and Cleotilde Gonzalez[1]

[1] Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, U.S.A.

[2] Department of Psychological and Brain Sciences and Program in Cognitive Science, Indiana University, Bloomington, Indiana, U.S.A.

[*] Corresponding author. Mailing address: Social and Decision Sciences, Dietrich College of Humanities and Social Sciences, 5000 Forbes Avenue, BP 208, Pittsburgh, PA, 15213. Email: ssloman@andrew.cmu.edu. ORCiD ID: 0000-0001-7479-4286

Keywords: exploration–exploitation trade-off; social networks; social learning; group dynamics; computational modelling

Abstract

How do people use information from others to solve complex problems? Prior work has addressed this question by placing people in social learning situations where the problems they were asked to solve required varying degrees of exploration. This past work uncovered important interactions between groups' *connectivity* and the problem's *complexity*: the advantage of less connected networks over more connected networks increased as exploration was increasingly required for optimally solving the problem at hand. We propose the Social Interpolation Model (SIM), an agent-based model to explore the cognitive mechanisms that can underlie exploratory behavior in groups. Through results from simulation experiments, we conclude that "exploration" may not be a single cognitive property, but rather the emergent result of three distinct behavioral and cognitive mechanisms, namely 1) breadth of generalization, 2) quality of prior expectation and 3) relative valuation of self-obtained information. We formalize these mechanisms in the SIM, and explore their effects on group dynamics and success at solving different kinds of problems. Our main finding is that broad generalization and high quality of prior expectation facilitate successful search in environments where exploration is important, and hinder successful search in environments where exploitation alone is sufficient.

1 Introduction

Imagine that you are in a new city, trying to choose a restaurant for dinner. Do you go back to the restaurant you stumbled across last night, which you know will be pretty good, or do you take your chances by trying something new, a decision which could result in a spectacular meal, a disastrous meal or anything in between?

This scenario is a canonical example of an *explore-exploit dilemma* (Mehlhorn et al., 2015; Schulz, Konstantinidis, & Speekenbrink, 2018; Sloman, Goldstone, & Gonzalez, 2019). In the context of decisions under uncertainty, decision-makers are typically understood to face an exploration–exploitation trade-off: *exploitation* refers to the strategy of repeatedly taking an action already observed to yield relatively favorable outcomes, but potentially missing out on even better payoffs somewhere not yet looked. *Exploration* refers to the strategy of taking actions whose outcomes are as yet unknown, forgoing the opportunity for guaranteed payoffs.

However, odds are you are not making a decision about where to go to dinner in the total absence of information from your past experiences or from others. The social accumulation of information is a large component of our decision-making processes. When you consider how risky choosing a particular restaurant is, you may consider experiences recounted by others or available reviews. If all travelers rely on online reviews—in which everyone who posts can see what everyone else posted—they may all gravitate towards a single good restaurant that gets increasingly many reviews. By contrast, relying on recommendations from different local connections who don't know each other may not converge on a ringing endorsement of a single restaurant. In other words, social information can come in many forms, and these forms can matter. In addition, we use social information to solve problems of varying degrees of complexity: is there a well-known best restaurant, or will your personal favorite be tucked away somewhere none of your connections may have yet visited?

We examine how people navigate search tasks when both individual and social information are available. Our work is a direct follow-up to a human experiment conducted by Mason, Jones, and Goldstone (2008), in which participants navigated such a search task. We present a novel computational model of search in the context of social information, the Social Interpolation Model (SIM), which specifies three points of possible variation in individuals' cognitive styles. We explore the effect of this variation on success in Mason et al. (2008)'s task. While we are able to somewhat replicate the negative interaction between network connectivity and problem complexity found by Mason et al. (2008), which we discuss further in section 2.3, we do not see the demonstration that this replication is possible as our primary contribution. Rather, we see our primary contribution as the exploration of how variations in cognitive style affect group-level dynamics in spatial search tasks.

We first review extant literature on social learning. We then summarize Mason et al. (2008)'s experiment and results before introducing the SIM.

1.1 Social learning

There is an extensive literature on social learning (Laland, 2004; Kendal et al., 2018). Perhaps some of the most seminal work has been done by Boyd and Richerson, e.g., Boyd and Richerson (2005), whose early work discusses the conditions under which social and individual learning will prove more or less effective. More recent work has examined the subtleties of social learning strategies, e.g., establishing cognitive phenomena like role differentiation (Roberts & Goldstone, 2011; Molleman, van den Berg, & Weissing, 2014), sophisticated reasoning processes about the cognitive trajectories of others (Frey, Albino, & Williams, 2018; Frey & Goldstone, 2018), sensitivity to perceived prestige (Brand, Heap, Morgan, & Mesoudi, 2020), and dynamic adjustment of the weight of others' outcomes (Gonzalez, Ben-Asher, Martin, & Dutt, 2015).

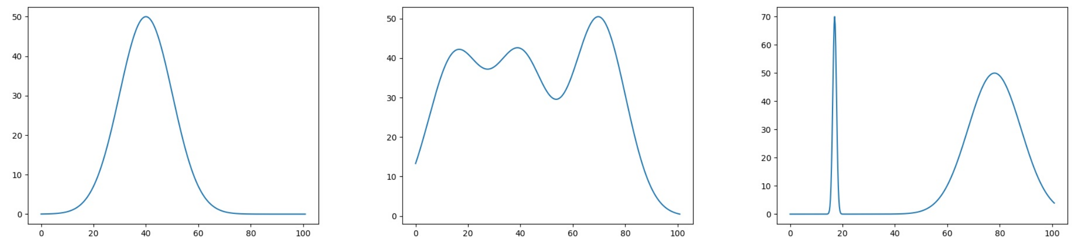
Other work more closely related to ours has discussed social learning strategies in networked spatial search environments similar to the one studied here (Rogers, 1988; Enquist, Eriksson, & Ghirlanda, 2007; Lazer & Friedman, 2007; Mason & Watts, 2012; Molleman et al., 2014; Toyokawa, Kim, & Kameda, 2014; Barkoczi, Analytis, & Wu, 2016; Barkoczi & Galesic, 2016; Toyokawa, Whalen, & Laland, 2019). For example, Toyokawa et al. (2014) construct a social learning environment in which participants have access to both the frequency with which others selected a particular option and to others' subjective evaluation of the option. They find that providing participants with choice frequency information both reduces exploration and enhances collective performance, while providing subjective evaluation information leads to a decline in group performance. In other related work, Toyokawa et al. (2019) show how aspects of the environment, namely uncertainty and group size, modulate the effectiveness of social learning strategies like imitation.

Unlike the papers cited above, we do not focus on discrete, identifiable social learning strategies. For example, while many of the papers above consider copying or imitation discretely (for a given action, agents either copy another or they don't), agents in our model do not make an explicit choice between individual and social learning. Rather, agents use the information at their disposal to interpolate the shape of the fitness landscape, and choose actions on the basis of this interpolation. Social learning strategies are thus represented not as discrete choices, but as the weight an individual assigns to their own vs. others' experienced outcomes (see section 3.3). In this sense, our work aligns with recent work on social sampling (Galesic, Olsson, & Rieskamp, 2018; Broomell, 2020; Krafft, Shmueli, Griffiths, Tenenbaum, & Pentland, 2021). Like us, this literature posits that people sample from their social environments, and aggregate the sampled information using simple, invariant heuristics. While Galesic et al. (2018) and Broomell (2020) focus on the judgment biases this can introduce, Krafft et al. (2021) highlight the potential for certain aggregation rules to lead to judgments that resemble perfect Bayesian inference. Like Krafft et al. (2021) and unlike Galesic et al. (2018) and Broomell (2020), we do not model differences between agents other than those introduced by the immediate informational environment. The version of the model we present therefore does not incorporate notions like group homophily. Unlike Krafft et al. (2021), we do not examine potential normative properties of our model. Rather, we report the results of empirical investigations of environments in which social sampling and interpolation can be helpful, and environments in which it can lead to systematic failures.

The primary contribution of our work is our examination of the effect of variation in the parameters of one's approach to informational sampling and aggregation—what we refer to as "cognitive mechanisms." Unlike prior work on social learning, our model is not specific to the social learning environment, but extends naturally from work on generalization and interpolation. We embed representations of more general individual differences in information-processing, such as breadth of generalization or quality of prior expectation, in a model of social learning. We explore the extent to which differences in behavior exhibited in a social learning context can be explained by basic cognitive mechanisms, potentially facilitating the unification of work on social learning with other areas of cognitive science.

2 Structure of Mason et al. (2008)'s experiment

Participants in Mason et al. (2008)'s experiment were brought into the lab, where they were told to guess numbers between 0 and 100 over 15 rounds. Participants' guesses were converted to payoffs via "fitness functions" which were unknown to the participants. All of these functions were smooth and had a unique global maximum. To maximize their cumulative payoffs, participants had to use the outcome information to guess numbers they thought would translate into high payoffs, and to uncover high-paying parts of the function landscape they did not yet have information about. Participants had access to not only their own outcome information—their guesses and the payoffs they had received from these guesses—but also to the outcome information of their "network neighbors." Mason et al. (2008) varied both the complexity of the fitness functions and the structure of the participants' networks.



(a) Unimodal landscape. (b) Trimodal landscape. (c) Needle landscape.

Figure 1: The fitness functions generating the payoffs in Mason et al. (2008)'s experiments.

2.1 Fitness functions

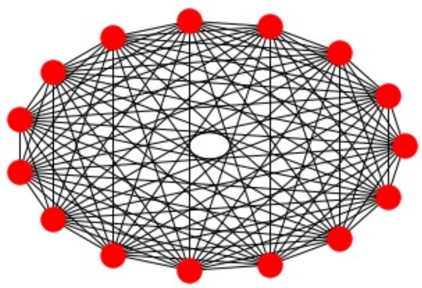
The fitness functions varied in terms of complexity, or difficulty of finding the global maximum using simple heuristics such as hill-climbing—i.e., without exploration. The least complex function—the function that required the least amount of exploration—was a unimodal function with a single, broad maximum. The next most difficult function was a trimodal function. The trimodal function's global maximum was as broad at the unimodal function's; however, in addition to the global maximum, the trimodal function had two local maxima that were almost but not as tall as the global maximum. Hill-climbers would likely get stuck on one of these two local maxima. Finally, the function that required the most exploration was a "needle" function. The needle function has a local maximum of the same height and width as the unimodal and trimodal functions' global maxima. However, in addition, the needle had a very thin global maximum. The fitness functions are shown in Figure 1.

2.2 Network structure

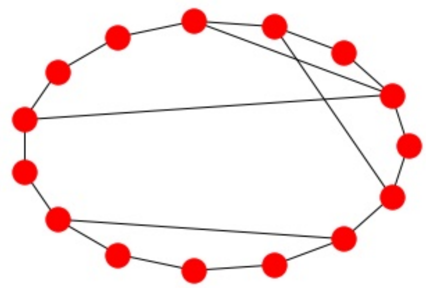
Participants were also assigned to different network structures which differed in terms of the efficiency of information transmission. Participants in the most highly-connected network, a fully-connected network, could see outcome information from everyone else in their group on every round.[1] Participants in the least-connected network, the lattice, could only see outcome information from their two immediate neighbors. Participants could also be assigned to a "small-world" network, in which information travelled quickly but not as quickly as in the fully-connected network, or a baseline network that randomly generated links between participants. Figure 2 shows the four different network structures. For the sake of brevity, we omit details of how the networks were generated, but interested readers can consult Mason et al. (2008) or Sloman et al. (2019).

It should be noted that participants in the fully-connected network have, on average, many more neighbors than participants in the other three networks. The result of this is that on each round, these participants view much more information.

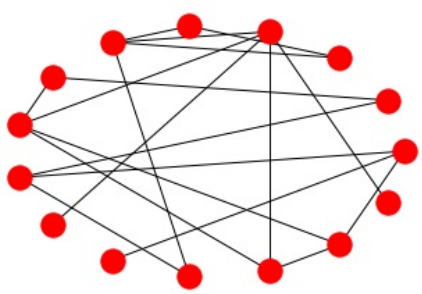
[1] Groups ranged in size from 5 to 19 players.



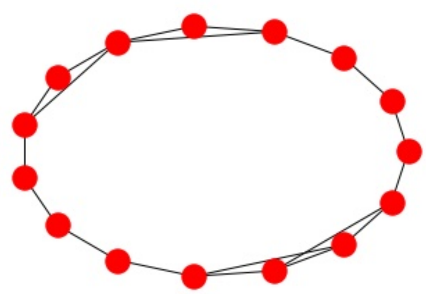
(a) Fully-connected network.



(b) Small-world network.



(c) Random network.



(d) Lattice network.

Figure 2: Examples of each of the possible network structures in Mason et al. (2008)'s experiments. Nodes represent participants, while edges link participants who could see each other's outcome information (all participants could see their own outcome information).

2.3 Experimental results

Here, we highlight Mason et al. (2008)'s three main findings about group behavior. In section 4, we discuss the extent to which our proposed model can account for these findings.

1. On the unimodal landscape, groups arranged in a fully-connected network outperformed groups in the more dispersed networks.

2. On the trimodal landscape, groups arranged in a small-world network outperformed other groups.

3. On the needle landscape, groups arranged in a lattice network outperformed groups in the more broadly interconnected groups.

In other words, broadly interconnected groups—groups in which members could see and imitate the actions and outcomes of other members—excelled at problems in which the solution was easy to find. Dispersed groups, in which individuals had little information about others' actions and outcomes, excelled at more complex problems.

2.3.1 cWSLS behavior

In a later reanalysis of Mason et al. (2008)'s data, we found that while participants in Mason et al. (2008)'s experiment tended to move around the problem space, they would move less when they were relatively successful, and more when they were relatively unsuccessful (Sloman et al., 2019). *Win-stay, lose-shift* (WSLS) refers to the policy that when you experience a gain ("win"), you stay with your current strategy, but when you experience a loss, you shift to another strategy (Robbins, 1952; Nowak & Sigmund, 1993; Bonawitz, Denison, Gopnik, & Griffiths, 2014; Billinger, Stieglitz, & Schumacher, 2018; Zhang, Moisan, & Gonzalez, 2021). While the traditional win-stay, lose-shift strategy operates over a small number of discrete strategies, we found that participants in Mason et al. (2008)'s experiment behaved as if they were using a graded, continuous version of win-stay, lose-shift (cWSLS, or continuous win-stay, lose-shift).

Guided by these observations, we developed the SIM, an agent-based model in which agents face the same task as Mason et al. (2008)'s participants. The next section describes the SIM. The SIM specifies three points of individual difference, encoded in the model as free parameters. Section 4 explores the effect of these parameters on group performance. In Appendix A, we demonstrate that the SIM can replicate participants' cWSLS behavior.

3 The Social Interpolation Model (SIM)

Agents in the SIM take actions, receive feedback from the environment on the outcomes of their actions and the outcomes of the actions of those in their social vicinity, and use that information to inform subsequent actions. In particular, they use this information to compute a valuation associated with each possible action, and then select an action with a probability proportional to its imputed valuation. In the context of the task described above, agents in the SIM *generalize* from the outcome information they have observed to unseen parts of the landscape.

To model perceptions of similarity, psychologists have used *generalization gradients*, or functions that describe the relationship between inputs and how similar these inputs are perceived to be to each other (Shepard, 1987). A central assumption embedded in most generalization gradients is that the perceived similarity of two inputs decays with the distance between them. In our context, the application of a generalization gradient implies that participants will think the number 0 is most similar to—i.e. has the closest expected associated payoff to—the number 1, is slightly less similar to the number 2, and is the least similar to the number 100. Generalization gradients typically also have the property that the rate of decay decreases as the distance from the focal input increases. For participants, the similarity between 0 and 11 is perhaps noticeably less than the similarity between 0 and 1, while the similarity between 0 and 100 is essentially the same as the similarity between 0 and 90.

One of the first cognitive models of function learning posits that people apply a generalization gradient to observed inputs to interpolate unseen parts of a function landscape (Busemeyer, Byun, Delosh, & McDaniel, 1997). More recently, the basic idea of similarity-based generalization has been applied to other models of function learning (Lucas, Griffiths, Williams, & Kalish, 2015), spatial search (Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018) and other exploration-exploitation tasks (Schulz, Speekenbrink, & Krause, 2018) in the form of Gaussian process regression. Gaussian process regression is a powerful form of regression that models the relationship between independent and dependent variables as a distribution over possible functions. One critical simplifying assumption that makes this approach computationally feasible is that the covariance between points is specified by a known function called a *kernel function*. Perhaps the most common kernel function is a radial basis function (RBF) kernel. Wu et al. (2018) show that the RBF kernel is a good approximation to human generalization in a spatial search task. The functional form of the RBF kernel, shown Equation 1, has been used as a generalization gradient in a number of psychological studies (Nosofsky, 1986; Busemeyer et al., 1997). Following this work, we also adopt the functional form shown in Equation 1 to model the generalization gradient, or *similarity function*, participants use to infer the expected payoffs at unseen parts of the fitness landscape.

$$S(i, j, c) = e^{-\left(\frac{i-j}{c}\right)^2} \tag{1}$$

c is a free parameter of the SIM which sets the degree of generalization (see section 3.1 for more detail). (For conciseness, we omit notation indicating dependence on c where c is specified by context.)

SIM agents combine their existing beliefs with the similarity function to form updated beliefs about the payoff $\hat{P}_i$ associated with each guess i:

$$\hat{P}_i = \frac{\sum_{j \in \{0 \ldots 100\}} w_j S(i, j) \tilde{P}(j)}{\sum_{j \in \{0 \ldots 100\}} w_j S(i, j)} \tag{2}$$

where $\tilde{P}(j)$ is the agent's expectation about the payoff corresponding to guess j. w_j is the weight the agents assigns to the information from observation j. We discuss this further in section 3.3.

All agents incorporate a subset of the outcome information they directly observed, either because they or one of their network neighbors guessed the number, by adopting the observed payoff as their expectation. For example, for an agent who observes a guess of 50 yield a payoff of 20, $\tilde{P}(50) = 20$.[2] The probability with which directly observed outcome information is incorporated is a function of how recently that information was observed: observations from the most recent round are incorporated with probability 1, while observations from previous rounds are incorporated with some probability $p(t) = e^{-bt}$, where t is the number of rounds that have passed since the observation occurred and b is a parameter that can be set at the discretion of the modeller. This function is a special case of the exponential function of memory decay (Wixted & Ebbesen, 1991; Rubin & Wenzel, 1996) where the probability of recalling an observation one just experienced is 1. While this element of the SIM allows for agents characterized by varying degrees of myopia, the agents in the simulations explored in the following section incorporate outcome information from only the previous round. This is equivalent to setting b to an extremely high value, resulting in agents who forget all the information they received in prior rounds. We made this choice for model tractability, and in order to better isolate the effects of the free parameters highlighted in the following subsections, all of which represent psychological constructs we found of more theoretical interest to understanding the dynamics of the task described above. However, the more general model in which b is allowed to vary facilitates both future exploration of the effect of memory decay, and application of our model to contexts in which decision-makers are more likely to rely on a history of outcomes (e.g., experimental contexts in which participants have the opportunity to record previously observed outcome information).

The SIM allows for individual-level variation in agents' beliefs about unobserved guesses; see section 3.2 for more detail.

Once they have computed these estimated payoffs, agents select probabilistically between all possible guesses. Probabilistic choice is one of the core foundations of decision-making research (Thurstone, 1927; Luce, 1963; Rieskamp, 2008; Busemeyer & Rieskamp, 2014). In particular, the probability that an agent selects a particular number i is proportional to its estimated payoff $\hat{P}_i$:

$$\pi_i = \frac{\hat{P}_i^2}{\sum_{j \in \{0,\ldots,100\}} \hat{P}_j^2} \tag{3}$$

The form of the decision function is taken from the decision stage of the Generalized Context Model (GCM) of category choice (Nosofsky, 1986; Ashby & Maddox, 1993; Nosofsky & Zaki, 2002). Like our model, the GCM posits that decisions are made probabilistically using weights computed on the basis of a generalization and inference process.

[2] If the agent observes outcome information for a number j twice, e.g., because two of their neighbors made the same guess, the agent forms two separate "expectations" $\tilde{P}(j_1)$ and $\tilde{P}(j_2)$. Note that $\tilde{P}(j_1)$ and $\tilde{P}(j_2)$ will not usually be equal because of the random noise added to payoffs.

The SIM incorporates three points of individual difference: the breadth of an agent's generalization, the nature of their beliefs about unobserved numbers, and the degree to which they rely on their own outcome information vs. information from others.

See section 3.4 for a worked example of how an agent evaluates and decides between possible actions.

3.1 Breadth of generalization

Agents' generalization gradient, shown in Equation 1, includes a free parameter c. The higher the value of c, the more observations that are distant from a number i influence the estimated payoff of i. (Agents whose degree of generalization corresponds exactly to the slope of the fitness function would have a $c = 14.29$.[3]) Agents with a high value of c can be thought of as more willing to generalize a good (or bad) experience in one place to other places that may appear dissimilar to others. To return to the restaurant goer we met in the introduction, a high c might imply that they are willing to generalize a good experience from one vegetarian restaurant to all other vegetarian restaurants, while other travelers may require the two restaurants to be more similar before they are willing to make such a generalization.

3.2 Quality of prior expectation

Equation 2 shows that agents rely on observed outcome information to compute the payoff they expect each number to yield. But what about numbers for which the agent *doesn't* directly observe payoffs? For these numbers, agents effectively pretend they have observed that number yield a payoff Λ. Λ is a free parameter that can vary at the discretion of the modeller. Using the terminology in Equation 2, $\tilde{P}(j) = \Lambda$ for all j for which the agent has not observed or not recalled any outcome information.

We believe that Λ is most naturally interpreted in terms of agents' uninformed expectations about the payoff they will receive from an arbitrary selection. Under this interpretation, agents with a high Λ are more optimistic: in the absence of other information, they assume actions yield positive rewards. Such optimism could stem from a generally positive disposition, the influence of an unobservable anchor (Epley & Gilovich, 2006), motivated reasoning derived from a particular aspiration level (Kunda, 1990), or any other factor that systematically influences the agent's uninformed priors.

Alternatively, Λ could be construed as the agent's valuation of novelty itself. Perhaps the traveller does not believe the food at an unexplored restaurant will be particularly good, but derives pleasure simply from having tried something new. This construal is related to curiosity, the drive to reduce uncertainty and seek new information, an important construct in understanding human behavior (Loewenstein, 1994; Mikulincer, 1997; Wojtowicz & Loewenstein, 2020) and designing artificial reinforcement learning systems (Bellemare et al., 2016; Pathak, Agrawal, Efros, & Darrell, 2017).

[3]This gradient captures the variance of the unimodal and trimodal functions' global maxima, and the needle function's local maximum.

We favor the previous interpretation because of Λ's direct relationship to $\tilde{P}(\cdot)$ in Equation 2, which represents the agent's prior valuation of a particular guess (see also the worked example in section 3.4). More specifically, $\tilde{P}(j) = \Lambda$ in case the agent does not have any outcome information about j. Like expectations based on the accrual of outcome information, Λ is generalized to other nearby guesses. While its role in the generalization process is a natural property of an expectation, it seems less natural to generalize novelty to other actions merely on the basis of spatial similarity, some of which one may have much information about. To reflect this, we refer to the cognitive property associated with Λ as the quality of an agent's uninformed prior or quality of prior expectation, but note that other interpretations are possible.

3.3 Reliance on information from own experiences

One of the main contributions of the present work is embedding existing work on generalization and interpolation in a context where participants are learning not only from their own actions and outcomes, but from the actions and outcomes of others. One of the important findings of the literature on social learning has been that people often do not treat equally information collected by others and information they collected themselves: we often overweight our own outcome information (Yaniv & Milyavsky, 2007; Weizsäcker, 2010; Yaniv & Choshen-Hillel, 2012; Goldstone, Wisdom, Roberts, & Frey, 2013; Puskaric, von Helversen, & Rieskamp, 2017; Grimm & Mengel, 2020).

Equation 2, which shows how agents form expectations for the payoff a particular number will yield, contains a sum taken over all observations (both actual observations and observations imputed to reflect the degree of prior expectation). If agents privilege their own outcome information, we would expect some observations to "loom larger" in these computations than others. This is captured by w_j, which always equals 1 unless the observation resulted from the agent's own action, in which case it is β. In other words, β represents the factor by which the agent weights their own outcome information more highly than other observations. A perfectly unbiased agent has $\beta = 1$. If the traveller in the introduction has a high value of β, they are more likely to rely on their own experiences rather than online reviews or recommendations from friends when making guesses about the quality of new restaurants.

3.4 An example decision

This section provides a worked example of how a hypothetical agent Al would compute their updated valuation of guessing 40, $\hat{P}_{40}$, after receiving one round of outcome information. Al guessed 32 on the last round and received a payoff of 66. Al has two network neighbors, one of whom guessed 20 on the last round and received a payoff of -3, and one of whom guessed 80 and received a payoff of 45. Al's internal parameter values are $c = 10$, $\Lambda = 10$ and $\beta = 5$. The steps below compute Al's updated valuation of 40.

1. Set W to an empty vector

2. Set $\tilde{P}$ to an empty vector

3. For all integers $j \in \{0, \ldots, 100\}$

   (a) Set $w_j = \beta = 5$ if $j = 32$ else 1

   (b) Append $w_j \times S(40, j, c = 10)$ to W

   (c) If $j \in [32, 20, 80]$

      i. Append observed corresponding payoff to $\tilde{P}$

   (d) Else

      i. Append $\Lambda = 40$ to $\tilde{P}$

4. Return $\frac{W \cdot \tilde{P}}{\sum W}$

The vector $\tilde{P}$ contains Al's prior valuations of each possible action, i.e., each integer between 0 and 100. Their prior valuation is equal to the observed information for 20, 32 and 80 (the actions for which they directly observed outcome information) and $\Lambda = 10$ for all other actions. The vector W contains the weighted similarity between the target action, 40, and all possible actions ($w_j S(40, j)$ for all $j \in \{0, \ldots, 100\}$). $\hat{P}_{40}$ is the normalized vector product of W and $\tilde{P}$, 17.43.

Al similarly updates their valuations of all possible actions, 0 through 100. Figure 3 shows the resulting valuation surface. Also shown are the effects of toggling each of the individual parameter values. For example, the red curve shows the valuation surface of an agent identical to Al, but who generalizes slightly more broadly.

Al then selects an action probabilistically according to Equation 3.

4 Model behavior

In this section, we use the SIM to explore how group success in the task designed by Mason et al. (2008) is affected by players' breadth of generalization, quality of prior expectation and reliance on their own experiences. In the sections below, we show how the percentage of participants who arrived within one unit of variance of the landscape's global maximum (*pctmax*) is the result of interaction effects between not only the fitness landscape and the network structure, but also between the network structures and agents' values of c, Λ and β.

For each combination of landscape, network structure and parameter values, we simulated 100 games, each composed of 15 agents endowed with the respective parameter values interacting over 15 rounds. This analysis shows that the direction of the parameters' effects is not consistent across the three landscapes. In the remainder of this section, we unpack the marginal effects of each of the three free parameters on performance in the different conditions—in other words, the effect of each parameter when the other two are held at baseline values. This allows us to isolate the effect of the three distinct cognitive mechanisms, and does not rely on additional researcher degrees of freedom in the choice of ranges and intervals at which to

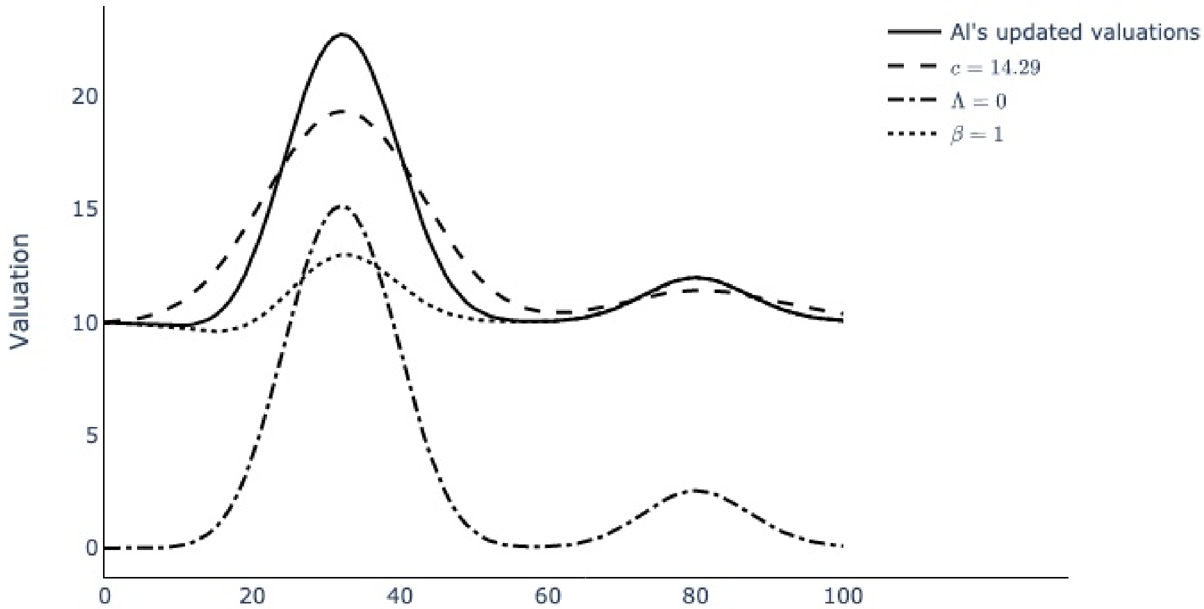


Figure 3: Al's updated valuations of every possible action (see main text for parameterization of this specific example). Also shown are the valuation surfaces created by altering each parameter to its baseline value (and keeping all other parameters unchanged).

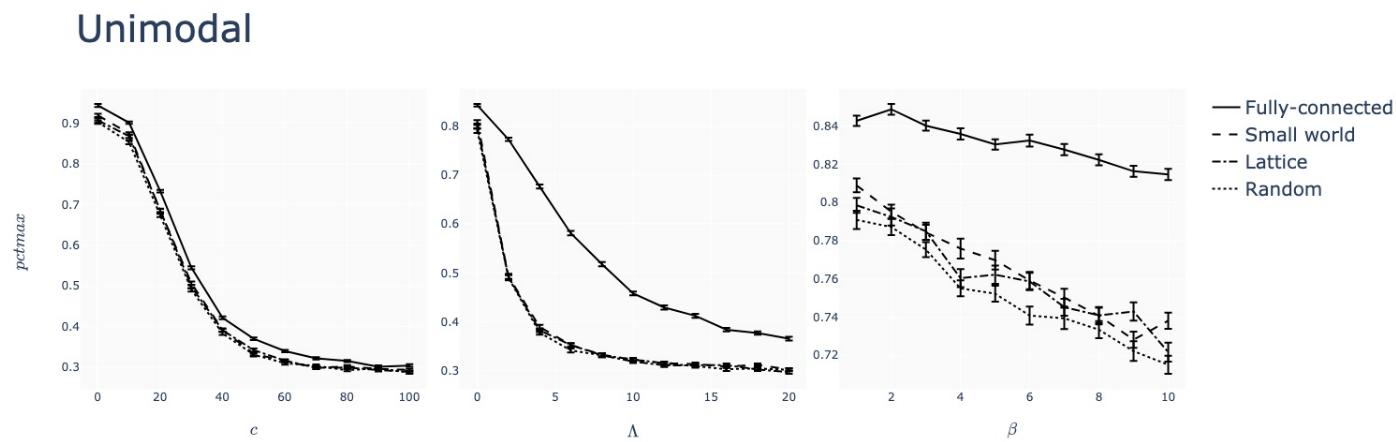
explore the other parameter settings. However, we also find important parameter interactions. Appendix B specifies a full regression model to capture the parameter effects and presents each parameter's main effects, collapsed across all other values of the other two parameters.

Here, we refer to Figure 4, which shows how *pctmax* changes with variation in each cognitive mechanism for each combination of landscape and network structure. Our main findings are as follows:

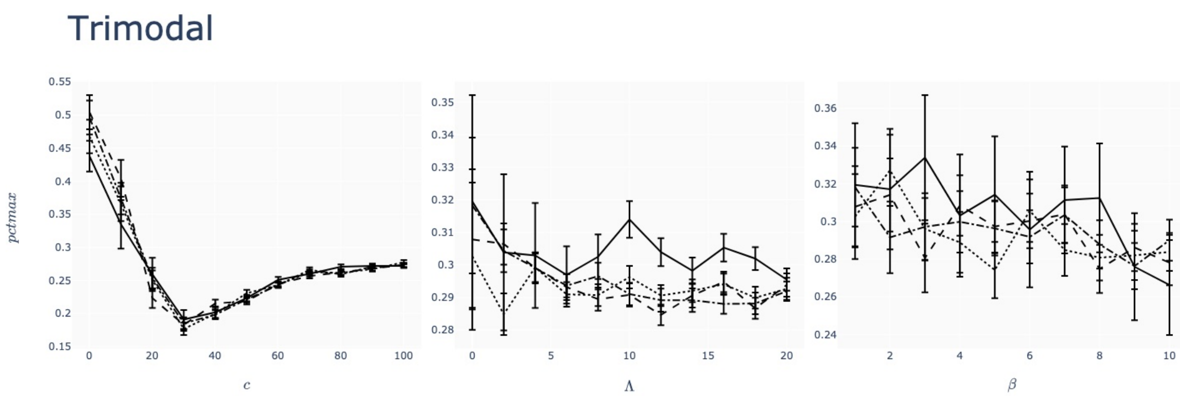
1. High c and high Λ help the most on the needle landscape, where exploration is important.

2. High c and high Λ are especially detrimental to performance on the unimodal landscape, where exploiting known high-yielding options is usually sufficient.

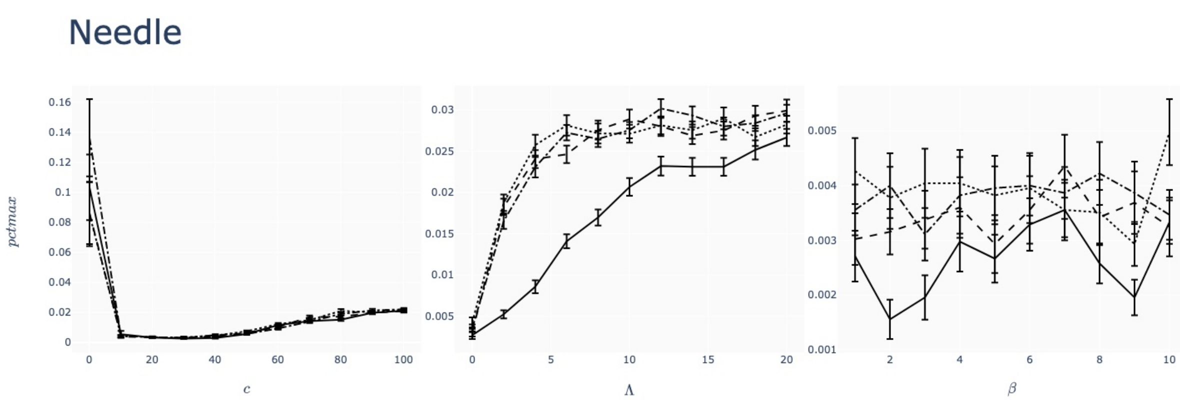
3. High β does not help performance on the more complex landscapes. While in general it is helpful on the unimodal landscapes, this benefit disappears in the absence of a mechanism for exploration.



(a)



(b)



(c)

Figure 4: Percentage of simulated agents' guesses that were within one unit of variance of the corresponding landscape's global maximum (*pctmax*) as a function of the agents' value of c (left), Λ (center) and β (right). The lines show the means across 100 trials (each consisting of 15 rounds of interaction between 15 agents, all endowed with the same parameter values). Error bars show one standard error of that mean. For plots on the left, Λ is fixed to 0 (a prior expectation of no payoff) and β is fixed to 1 (no overweighting of own information). For plots in the center, c is fixed to 14.29 (corresponding to the gradient of the unimodal and trimodal functions' global maxima) and β is again fixed to 1. For plots on the right, c is fixed to 14.29 and Λ is fixed to 0.

Effect of c (breadth of generalization). In general, the effects of c are consistent with broad generalization functioning as a mechanism for more exploration: higher c leads to higher *pctmax* in contexts where exploration is important (i.e., on the needle landscape), and hinders performance in contexts where exploiting small amounts of information is usually sufficient (i.e., on the unimodal landscape).

On the unimodal landscape, low c leads to better outcomes (top panel of Figure 4). Recall that the unimodal landscape was designed so that players could find the global maximum using simple heuristics such as hill-climbing or noisy imitation of others. Broad generalization of one's findings can lead the searcher to venture into unexplored parts of the landscape, and neglect to reap benefits right under their nose. Interestingly, agents with c near zero perform better than agents whose c is appropriately calibrated to the landscape. By "appropriately calibrated" we mean cases in which the generalization gradient matches the gradient of the landscape's mode—i.e., cases in which the degree to which agents' payoffs provide information about payoffs obtainable in other parts of the landscape *matches* agents' "beliefs" about the nature of this information. Agents on the unimodal landscape who think their findings apply more narrowly than they actually do perform better, at least in the aggregate. By not over-generalizing, these agents do not risk occasionally over-stepping the bounds of the global maximum, and are sure to remain comfortably within the region of highest payoffs.

Figure 4 shows a non-monotonic effect of c in the trimodal and needle landscapes. We interpret this result as elucidating a tension between appropriately calibrated generalization and generalization as a mechanism for exploration. As discussed just above, narrow generalization prevents agents from jumping from the global maximum once they've found it. On the other side of the same coin, broad generalization helps push agents out of local maxima. We suspect the difference in location in the inflection points in the effect of c on the trimodal and needle landscapes reflects the difference in the gradient of the local maxima in the two landscapes: since the global maximum on the needle landscape is so much narrower than the global maximum on the trimodal landscape (see Figure 1), even mild generalization can be sufficient to push agents out of it.

While throughout this paper we adopt the interpretation of c consistent with the rich psychological literature on generalization, it is worth noting that c can also be interpreted more directly as the value of sampling in a new location. For agents with low c, sampling somewhere new will only update their valuation of a narrow range of values. For agents with high c, sampling somewhere new will update their valuation around a much broader range of values. This interpretation provides additional intuition for c as a mechanism for exploration.

Effect of Λ (quality of prior expectation). As explained in section 3.2, agents with higher Λ look more favorably on unseen parts of the landscape. Like broad generalization, high prior expectations induce exploration: agents with high Λ are more likely to visit unseen areas because they attribute high payoffs to those areas.

The top panel of Figure 4 shows that low Λ leads to better outcomes on the unimodal landscape. Because this landscape has a single peak, players who jump around too much are likely to forsake the successful solutions they have already found.

On the other hand, the bottom panel of Figure 4 shows that higher Λ leads to better outcomes on the needle landscape. We speculate that high expectations lead to better performance on the needle landscape for the same reason they hinder performance on the unimodal landscape: high expectations induce exploration, which is important to find the needle's thin global maximum.

Interpreting both c and Λ as mechanisms for exploration is consistent with the strong negative interaction effect between c and Λ (Table 1 in Appendix B). For agents who generalize broadly, the effect of high expectations will be muted by the spillover of observed payoff information across the landscape. Put another way, in cases where the nature of one's expectations do not induce one to try new things, broad generalization is an important alternative mechanism for exploration. Agents must strike a balance between exploration and exploitation, and are most successful in the presence of counterbalancing mechanisms to search broadly on the one hand and to take certain gains on the other.

Effect of β (relative weight on personal experience). The rightmost plots in Figure 4 show that the marginal effect of β is difficult to detect in the trimodal and needle landscapes. In other words, when the other parameters are held at their baseline values, overweighting of one's own information does not have a noticeable effect in the more complex landscapes. Interestingly, the least complex landscape, the unimodal landscape, is the only one which exhibits a detectable, negative, marginal effect of β.

This was surprising to us: unlike the unimodal landscape, the trimodal and needle landscapes were constructed such that social information would be important to help participants avoid getting stuck on local maxima. The results presented in Appendix B are consistent with this intuition: the main effect of β (i.e., the average effect of β across all values of c and Λ we simulated) is distinctly negative on the trimodal and needle landscapes, and positive on the unimodal landscape. This suggested to us that there is something important about low prior expectations and/or relatively narrow generalization driving the counterintuitive marginal effects of β.

Further investigation showed that all instances where this qualitative pattern occurred—i.e., the effect of β was negative in the unimodal landscape and flat in the trimodal and needle landscapes at constant values of c and Λ—were characterized by $\Lambda = 0$. When $\Lambda = 0$, agents have little incentive to visit unexplored options. A high β would lead agents to stay in the neighborhood of their initial guess. On the unimodal landscape, this would be detrimental: chances are, one of an agents' neighbors would have stumbled on, or close to, the global maximum, and blindness to that information would lead agents to do especially poorly. On the other two landscapes, where exploration is needed to find the global maximum, circling around one's own guess may lead one to do no worse than looking to neighbors who themselves will get trapped in local maxima due to their own lack of exploration. According to this account, the observed pattern on the trimodal and needle landscapes is an artifact of the homogeneity of the simulated groups. We predict that the marginal effect of β would be negative for any individual with $\Lambda = 0$ in cases where the individual's social network contained agents with $\Lambda > 0$. We leave this more complex analysis to future work.

Effect of network structure. Among our simulated groups, the fully-connected network outperformed the other network structures on the unimodal landscape and underperformed compared to the other networks on the needle landscape. There was no clear best-performing network on the trimodal landscape, and performance between the other three networks did not appear to differ much on the other two landscapes.

These patterns can somewhat be seen in Figure 4: the line corresponding to the fully-connected network is clearly above the other lines in Figure 4a and below the other lines in the middle panel of Figure 4c. In Appendix B, we present more evidence supporting this general trend.

This pattern somewhat replicates the findings of Mason et al. (2008) listed in section 2.3. While the overall trends among the ranges of parameter values we chose to explore do not replicate the exact ordering of the networks in each condition, we also find that more connected networks do well on the least complex landscape, and comparatively worse on the more complex landscapes.

Later work by Mason and colleagues similarly complicates the straightforward negative interaction between connectivity and complexity found by Mason et al. (2008): Mason and Watts (2012) found that participants in more connected networks tended to outperform those in less connected networks even on complex, multimodal fitness functions. Intriguingly, they find that the mechanism for this is that participants were more likely to copy their neighbors if two or more neighbors guessed in the same place. In the inefficient networks, participants' neighbors were more likely to be neighbors with each other and thus there was substantial correlation between the information a participant's neighbors had access to, leading them more often to the same guesses. The SIM implicitly encodes this mechanism since each time the same—or similar—guess is made, it augments the valuation surface at that location (as long as the observed payoff is greater than Λ). One noteworthy feature of Mason and Watts (2012)'s work is that they keep the number of neighbors each participant has constant in each network, while in Mason et al. (2008)'s study, participants in the lattice networks had substantially fewer neighbors than participants in the fully-connected networks. Indeed, all participants in the fully-connected networks shared a neighborhood, since all participants were connected to all other participants. Combined with the higher average degree overall, we suspect that participants in the fully-connected networks actually observed more duplicate observations than participants in the lattice networks. Thus, we see the fact that both the empirical and simulation results suggest that the lattice networks offer an advantage over the fully-connected networks on the more rugged landscapes as supporting evidence for, rather than as inconsistent with, Mason and Watts (2012)'s findings.

5 Discussion

The SIM offers a novel characterization of problem-solving that postulates differences in 1) how broadly agents generalize information to unseen areas of a problem space, 2) the quality of their prior expectations, and 3) how much they value information about their own vs. others' outcomes. We laid out testable predictions of these cognitive mechanisms at work.

We show that broad generalization and high prior expectations can lead agents to succeed in environments where exploration is important—and fail in environments where exploitation alone is sufficient. In other words, individual differences in the basic cognitive process of similarity perception can have dramatic effects on a seemingly unrelated process: trading off between exploration and exploitation. This suggests to us that differences in explore–exploit trade-offs may stem from individual differences in breadth of generalization, in addition to or instead of from differences in preference for new information.

The SIM replicates Mason et al. (2008)'s finding that the fully-connected network has a distinct comparative advantage on the unimodal landscape, and performs comparatively worse in the more complex networks. However, as we stress in the introduction, our investigation of the SIM's behavior was not a replication attempt. In line with the intent of the present investigation, our findings introduce nuance into their results. For example, while Mason et al. (2008) found that participants in the lattice network outperformed participants in the fully-connected network on the needle landscape, in our simulations groups composed of agents with sufficiently high prior expectations did as well in the fully-connected network as in the lattice network. This suggests that different network structures may be optimal not only for different problem structures, but also for different cognitive styles. For example, if agents are highly connected to one another, then they may be able to find solutions to complex problems (e.g., the global maximum on the needle landscape) if they tend to view unseen options optimistically. It also suggests that for many complex problems, there may be multiple interventions or strategies that could lead to the same desired outcome. Armed with multiple structural and behavioral interventions that achieve a desired outcome, decision-makers and team leaders could adopt strategies that satisfy other constraints. For example, while reducing information exchange and social interaction can prevent bandwagoning, it can have other, likely obvious, detrimental effects. Rather than relying on reducing opportunities for social interaction, team leaders could consider ways to provide their teams with information or incentives that would change their valuation of unseen options—i.e., interventions that would mimic adjusting the Λ parameter—rather than change the network structure.

5.1 Areas for future work

Our work is *not* intended as definitive support for the SIM. While we have synthesized previous work into a *possible* model, we have not conducted a formal model comparison. In addition, our analyses are post-hoc. Future work could test the adequacy of the model via, e.g., selective influence tests, and compare the model to plausible competitors, e.g., models that posit forms of Bayesian inference (Krafft et al., 2021) or models that explicitly incorporate the trade-off between exploitation and uncertainty reduction via decision rules like Upper Confidence Bound sampling (Schulz, Speekenbrink, & Krause, 2018; Wu et al., 2018). Further development of the SIM could incorporate additional components to represent features of human cognition such as a ceiling on the number of observations that can be retained and used for inference (Miller, 1956).

Another immediate avenue for future work is to explore the effects of group size[4] or agent heterogeneity. Our results are based on the extremely unrealistic assumption that all agents in a group share the same cognitive style. Previous work has shown substantial heterogeneity in individual behavior in networked search and problem-solving tasks (Roberts & Goldstone, 2011; Toyokawa et al., 2014, 2019; Sloman et al., 2019). Incorporating agent heterogeneity into future work could reveal novel and important synergies between different cognitive styles.

Additional future directions could involve applying or extending our model to other spatial search contexts from available literature. For example, one extension we mentioned in the introduction, which is especially reminiscent of our motivating example of restaurant reviews, would be to model social information as subjective valuations made by other agents, rather than as objective outcome information (Toyokawa et al., 2014). Other immediate extensions could involve the complexity of the options themselves: while we examine a static search context in which the payoff distribution remains fixed, Toyokawa et al. (2019) demonstrate how to construct a dynamic context in which payoffs distributions change over time. In addition, while we study options characterized by only a single numerical dimension, in most ecological search contexts, options are multi-dimensional. Like us, Mason and Watts (2012) and Wu et al. (2018) study environments with spatially-correlated rewards, but extend the search space to two dimensions. Our model could naturally be extended to incorporate multiple dimensions by a simple modification of the similarity function (Equation 1) to account for distances across multiple dimensions. If i and j in Equation 1 referred to vector rather than scalar values, the usual approach in the psychological literature would be to replace the scalar difference $i - j$ with the city-block distance (a special case of the Minkowski distance) between the vectors i and j (Shepard, 1964; Nosofsky, 1986; Shepard, 1987).

In a paradigm known as the "NK landscape" paradigm, researchers can systematically vary both the dimensionality of the search space and the number of local maxima (Kauffman, 1995; Gavetti & Levinthal, 2000; Lazer & Friedman, 2007; Billinger et al., 2018). Future work could use existing experimental results in this paradigm mentioned in the introduction (Billinger et al., 2018) to constrain the parameter values of our model.

A final paradigm to which our model could be extended is search contexts in which the environment is characterized by time-varying features which define a context the learner can use to condition their predictions, so called "contextual multi-armed bandit" (CMAB) problems (Stojic, Analytis, & Speekenbrink, 2015; Schulz, Konstantinidis, & Speekenbrink, 2018). This effectively requires learning several payoff functions:

[4]Preliminary results suggest group size may interact with the SIM's parameter values in interesting and counter-intuitive ways.

one for each possible context. The mechanism of similarity-based interpolation can be extended to a problem with a hierarchical structure like CMABs: in the same way similarity between actions can be used to predict similarity of outcomes, similarity of context may also be a predictor of similarity of payoff function. While in the context focused on in this paper an observation was defined by an action (numerical guess) and payoff, in a CMAB problem, an observation is defined by an action, a set of contextual features and a payoff. Our model could incorporate this by combining both the similarity between actions and the similarity between contexts into a scalar similarity function that operates over observations, either by averaging the two similarity scores together or by treating the context as an additional dimension and aggregating similarity across dimensions as explained above.

6 Conclusion

We constructed a model from basic cognitive operations, and demonstrated that differences in exploratory behavior and group success can emerge from individual- and group-level differences in information-processing mechanisms. We hope that future work on how humans navigate the explore-exploit dilemma takes steps to identify which of the several cognitive processes could account for the observed results in new tasks.

Conflict of interest

The authors have no conflicts to disclose.

Data availability

The code to reproduce all the results reported in this paper is available at https://osf.io/xwvja.

Funding

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

References

Ashby, F. G., & Maddox, W. T. (1993). Relations between Prototype, Exemplar, and Decision Bound Models of Categorization. *Journal of Mathematical Psychology*, *37*.

Barkoczi, D., Analytis, P. 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.* Cognitive Science Society.

Barkoczi, D., & Galesic, M. (2016). Social learning strategies modify the effect of network structure on group performance. *Nature Communications*, *7*(1), 13109. doi: 10.1038/ncomms13109

Bellemare, M. G., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., & Munos, R. (2016). Unifying Count-Based Exploration and Intrinsic Motivation. *arXiv:1606.01868 [cs, stat]*.

Billinger, S., Stieglitz, N., & Schumacher, T. R. (2018). Search on Rugged Landscapes: An Experimental Study. *Organization Science*, *25*(1), 93–108.

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*, 35–65. doi: 10.1016/j.cogpsych.2014.06.003

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures.* Oxford ; New York: Oxford University Press.

Brand, C. O., Heap, S., Morgan, T. J. H., & Mesoudi, A. (2020). The emergence and adaptive use of prestige in an online social learning task. *Scientific Reports*, *10*(1). doi: 10.1038/s41598-020-68982-4

Broomell, S. B. (2020). Global–Local Incompatibility: The Misperception of Reliability in Judgment Regarding Global Variables. *Cognitive Science*, *44*(4). doi: 10.1111/cogs.12831

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997). Learning Functional Relations Based on Experience With Input-Output Pairs by Humans and Artificial Neural Networks. In K. Lamberts & D. R. Shanks (Eds.), *Knowledge, concepts, and categories* (1st MIT Press ed ed.). Cambridge, Mass: MIT Press.

Busemeyer, J. R., & Rieskamp, J. (2014). Psychological research and theories on preferential choice. In *Handbook of choice modeling* (pp. 49–72).

Enquist, M., Eriksson, K., & Ghirlanda, S. (2007). Critical Social Learning: A Solution to Rogers's Paradox of Nonadaptive Culture. *American Anthropologist*, *109*(4), 727–734. doi: 10.1525/aa.2007.109.4.727

Epley, N., & Gilovich, T. (2006). The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient. *Psychological Science*, *17*(4), 311–318. doi: 10.1111/j.1467-9280.2006.01704.x

Frey, S., Albino, D. K., & Williams, P. L. (2018). Synergistic Information Processing Encrypts Strategic Reasoning in Poker. *Cognitive Science*, *42*(5), 1457–1476. doi: 10.1111/cogs.12632

Frey, S., & Goldstone, R. L. (2018). Cognitive mechanisms for human flocking dynamics. *Journal of Computational Social Science*, *1*(2), 349–375. doi: 10.1007/s42001-018-0017-x

Galesic, M., Olsson, H., & Rieskamp, J. (2018). A sampling model of social judgment. *Psychological Review*, *125*(3), 363–390. doi: 10.1037/rev0000096

Gavetti, G., & Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search. *Administrative Science Quarterly*, *45*(1), 113–137. doi: 10.2307/2666981

Goldstone, R. L., Wisdom, T. N., Roberts, M. E., & Frey, S. (2013). Learning Along With Others. In *Psychology of Learning and Motivation* (Vol. 58, pp. 1–45). Elsevier. doi: 10.1016/B978-0-12-407237-4.00001-3

Gonzalez, C., Ben-Asher, N., Martin, J. M., & Dutt, V. (2015). A Cognitive Model of Dynamic Cooperation With Varied Interdependency Information. *Cognitive Science*, *39*(3), 457–495. doi: 10.1111/cogs.12170

Grimm, V., & Mengel, F. (2020). Experiments on Belief Formation in Networks. *Journal of the European Economic Association*, *18*(1), 49–82. doi: 10.1093/jeea/jvy038

Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity.* Oxford University Press.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences*, *22*(7), 651–665. doi: 10.1016/j.tics.2018.04.003

Krafft, P., Shmueli, E., Griffiths, T. L., Tenenbaum, J. B., & Pentland, A. S. (2021). Bayesian collective learning emerges from heuristic social learning. *Cognition*, *212*, 104469. doi: 10.1016/j.cognition.2020.104469

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Laland, K. N. (2004). Social learning strategies. *Learning & Behavior*, *32*(1).

Lazer, D., & Friedman, A. (2007). The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, *52*(4), 667–694.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. doi: 10.1037/0033-2909.116.1.75

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*(5), 1193–1215. doi: 10.3758/s13423-015-0808-5

Luce, R. D. (1963). Detection and recognition. In *Handbook of mathematical psychology* (pp. 103–189). New York: WIley.

Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*(3), 764–769. doi: 10.1073/pnas.1110069108

Mason, W. A., Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, *137*(3), 422–433. doi: 10.1037/a0012798

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., ... Gonzalez, C. (2015). Unpacking the Exploration–Exploitation Tradeoff: A Synthesis of Human and Animal Literatures. *Decision*, *2*(3).

Mikulincer, M. (1997). Adult Attachment Style and Information Processing: Individual Differences in Curiosity and Cognitive Closure. *Journal of Personality and Social Psychology*, *72*(5).

Miller, G. A. (1956). The Magical Number Seven. *The Psychological Review*, *63*.

Molleman, L., van den Berg, P., & Weissing, F. J. (2014). Consistent individual differences in human social learning strategies. *Nature Communications*, *5*(1). doi: 10.1038/ncomms4570

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal* *of Experimental Psychology: General*, *115*(1).

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924–940. doi: 10.1037/0278-7393.28.5.924

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*(6432), 56–58. doi: 10.1038/364056a0

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-Driven Exploration by Self-Supervised Prediction. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 488–489). Honolulu, HI, USA: IEEE. doi: 10.1109/CVPRW.2017.70

Puskaric, M., von Helversen, B., & Rieskamp, J. (2017). How Social and Non-Social Information Influence Classification Decisions: A Computational Modelling Approach. *Quarterly Journal of Experimental Psychology*, *70*(8), 1516–1534. doi: 10.1080/17470218.2016.1192209

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446–1465. doi: 10.1037/a0013646

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*(5).

Roberts, M. E., & Goldstone, R. L. (2011). Adaptive Group Coordination and Role Differentiation. *PLoS ONE*, *6*(7). doi: 10.1371/journal.pone.0022377

Rogers, A. R. (1988). Does Biology Constrain Culture. *American Anthropologist*, *90*(4), 819–831. doi: 10.1525/aa.1988.90.4.02a00030

Rubin, D. C., & Wenzel, A. E. (1996). One Hundred Years of Forgetting: A Quantitative Description of Retention. *Psychological Review*, *103*(4).

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2018). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6).

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. doi: 10.1016/j.jmp.2018.03.001

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87. doi: 10.1016/0022-2496(64)90017-3

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. doi: 10.1126/science.3629243

Sloman, S. J., Goldstone, R. L., & Gonzalez, C. (2019). Complex exploration dynamics from simple heuristics in a collective learning environment. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society.* Montreal, QC.

Stojic, H., Analytis, P. P., & Speekenbrink, M. (2015). Human behavior in contextual multi-armed bandit problems. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society.* Pasadena, CA.

Thurstone, L. L. (1927). Psychophysical Analysis. *The American Journal of Psychology*, *38*(3).

Toyokawa, W., Kim, H.-r., & Kameda, T. (2014). Human Collective Intelligence under Dual Exploration-Exploitation Dilemmas. *PLoS ONE*, *9*(4).

Toyokawa, W., Whalen, A., & Laland, K. N. (2019). Social learning strategies regulate the wisdom and madness of interactive crowds. *Nature Human Behaviour*, *3*(2), 183–193. doi: 10.1038/s41562-018-0518-x

Weizsäcker, G. (2010). Do We Follow Others when We Should? a Simple Test of Rational Expectations. *American Economic Review*, *100*(5), 2340–2360. doi: 10.1257/aer.100.5.2340

Wixted, J. T., & Ebbesen, E. B. (1991). On the Form of Forgetting. *Psychological Science*, *2*(6), 409–415. doi: 10.1111/j.1467-9280.1991.tb00175.x

Wojtowicz, Z., & Loewenstein, G. (2020). Curiosity and the economics of attention. *Current Opinion in Behavioral Sciences*, *35*, 135–140. doi: 10.1016/j.cobeha.2020.09.002

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*(12), 915–924. doi: 10.1038/s41562-018-0467-4

Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the Wisdom of Others to Make Better Decisions: Suspending Judgment Reduces Egocentrism and Increases Accuracy: Exploiting the Wisdom of Others. *Journal of Behavioral Decision Making*, *25*(5), 427–434. doi: 10.1002/bdm.740

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104–120. doi: 10.1016/j.obhdp.2006.05.006

Zhang, H., Moisan, F., & Gonzalez, C. (2021). Rock-Paper-Scissors Play: Beyond the Win-Stay/Lose-Change Strategy. *Games*, *12*(3).

Appendices

A Replicating cWSLS behavior

We hypothesized that the SIM would exhibit cWSLS behavior, because by design agents in the SIM gravitate towards higher payoffs. To examine whether the SIM did indeed exhibit cWSLS behavior, we simulated the game described in section 2 for each combination of a "low" and a "high" value of each of the three free parameters of the SIM: $c \in \{.1, 100.1\}$, $\Lambda \in \{0, 20\}$ and $\beta \in \{1, 10\}$. For each parameter combination, we simulated 100 games, each with 15 rounds and 15 "players," in each of the 12 combinations of fitness landscapes and network structures ({unimodal, trimodal, needle} × {fully-connected, small world, random, lattice}).

Figure 5 shows how agents' jump size—the distance between subsequent guesses—varies as a function of the difference in payoffs between these guesses, for each of these simulations. The thick black line shows the behavioral trend reported in Sloman et al. (2019): when participants in the experiment experienced relative *gains* (wins), their jump size decreased. By contrast, participants who experienced relative *losses* did not temper their movement.

Figure 5 also shows that the SIM is able to replicate this pattern. Each of the thinner colored lines shows the same trend for simulations of groups of SIM agents using a particular combination of parameter values. In general, the simulations exhibit the same qualitative trend as the empirical data for all parameter values. Readers may notice that the dashed lines tend to fall above the others for payoff differences greater than 0: when these agents experience gains, they are more hesitant to converge towards the high-yielding guess. These lines represent groups composed of agents with extremely high values of c, i.e., agents who generalize extremely broadly. When these agents experience or observe a high payoff, they are more willing to believe this payoff will generalize to other guesses further away, and thus do not reduce their jump size much. This is consistent with our finding, discussed further in section 4, that broad generalization functions as a mechanism for exploration.

B Further exploration of parameter effects

Section 4 explores the *marginal* effects of each of the SIM's three parameter values—i.e., the effect of each when the other two were held at their baseline values. This appendix presents the results of a complete regression model of the parameters' effects on *pctmax*, and an exploration of the parameters' effects when taking into account interactions with other parameter values.

As we stated in section 4, we simulated 100 games for each combination of landscape, network structure and parameter values, each composed of 15 agents endowed with the respective parameter values interacting over 15 rounds. Table 1 shows the effects of the properties of these simulated trials on each trial's corre-

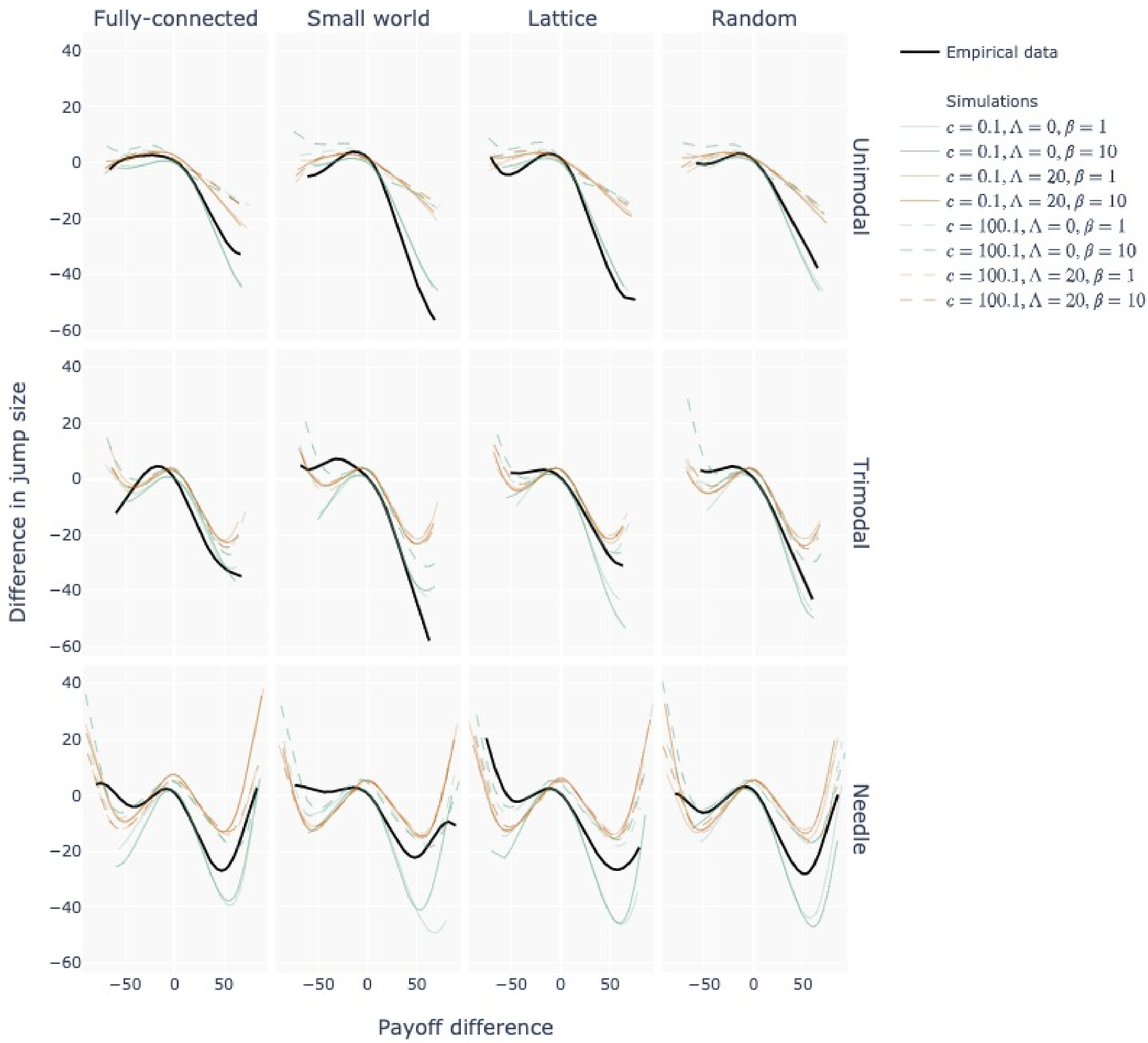


Figure 5: Jump size as a function of relative payoffs, broken down by fitness landscape (rows) and network structure (columns). Each line shows the relationship between the experienced difference in payoffs ($p(t) - p(t-1)$) (x-axis) and the difference in jump size ($|g(t+1) - g(t)| - |g(t) - g(t-1)|$) ($y$-axis), where $p(t)$ is the payoff experienced at time t, and $g(t)$ is the corresponding guess made at time t. Each colored line shows a local linear regression over all observations with a single set of parameter values (all eight combinations of $c \in \{.1, 100.1\}$, $\Lambda \in \{0, 20\}$ and $\beta \in \{1, 10\}$). Each regression line is estimated using all observations from 100 simulated games, each with 15 agents and 15 consecutive rounds ($n = 100 \times 15 \times 13$; observations from the first and last rounds of each game are omitted as there is no data either at time $t-1$ or at time $t+1$). The thick black line shows the same regression over the behavioral data collected by Mason et al. (2008). We used a bandwidth of 16 for all regressions, but other bandwidths show the same qualitative result.

sponding *pctmax*, estimated using OLS linear regression. To meet the linear model's assumption that the dependent variable is continuous and unbounded, we converted the *pctmax* value corresponding to each trial into a $logodds(max) = log(\frac{pctmax}{1-pctmax})$ value, and regressed this against each group's endowed values of c, Λ and β, as well as indicators that specified the group's network structure.[5] We transformed the parameter values so they could be interpreted as deviations from baseline values: 14.29 is subtracted from c (the gradient of the global maximum in the unimodal and trimodal conditions) and 1 is subtracted from β (corresponding to no additional weighting on own information; Λ remains unchanged and can be interpreted as the difference from a baseline of a prior expectation of no payoff). We also divided the values of c, Λ and β by the distance between their respective minimum and maximum values.[6] Under this transformation, a unit change in the independent variables representing the parameter values corresponds to the effect of toggling the parameter between its minimum and maximum values.

The intercepts correspond to the predicted $logodds(max)$ value for a fully-connected network composed of agents with perfectly calibrated generalization gradients, a prior expectation of no payoff and no over-weighting of own information. Baseline performance was highest on the unimodal landscape and lowest on the needle landscape, which is consistent with the behavioral patterns found by Mason et al. (2008).

Table 1 also shows that the values of the free parameters accounted for the least amount of variance in the simulated *pctmax* value on the trimodal landscape. In other words, manipulating the SIM's free parameters did not affect its predictions when operating on the trimodal landscape as much as when operating on the other two landscapes. We speculate that this is because the trimodal landscape yields more uniform payoffs than the other two landscapes (see Figure 1): guesses on different parts of the landscape do not result in dramatically different payoffs, and payoffs from other parts of the landscape may be sufficient in and of themselves to induce exploration. Stochasticity in initial conditions and payoffs thus likely plays a larger role in a group's outcomes than it does on the unimodal and needle landscapes.

Figure 6 plots the main effects of each of the three parameters in the different landscapes. It can be interpreted in exactly the same way as Figure 4, with the important exception that rather than fixing the other parameters, we calculate means and standard errors across all simulated trials. For example, points corresponding to a particular value of c show the means of 100×11 values of $\Lambda \times 10$ values of $\beta = 11{,}000$ trials.

Figure 6 gives a richer picture of the main effects shown in Table 1: in Figure 6a, which plots the effects of the parameters on the unimodal landscape, the lines slope downwards for c and Λ and upwards for β. This is consistent with the negative coefficients associated with c and Λ and positive coefficient associated with β in Table 1.[7] Section 4 discusses our interpretation of these directional effects. The sign of the coefficients

[5] To avoid discontinuities that arise at *pctmax* values of 0, we added 2×10^{-16} to each *pctmax* value.

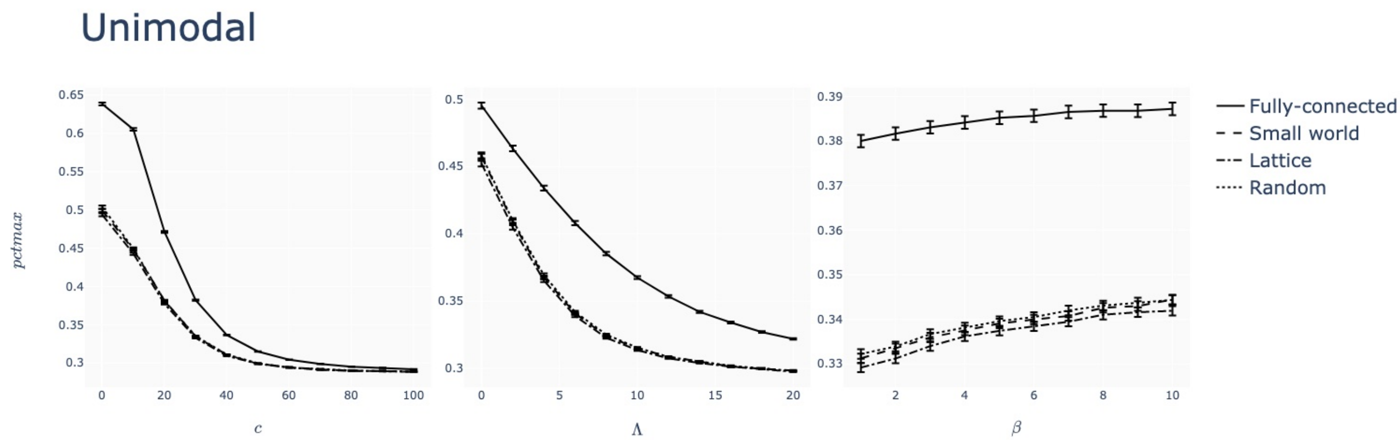
[6] Before the previous transformation, this was .1 and 100.1 for c, 0 and 20 for Λ and 1 and 10 for β.

[7] Readers may note that the negative marginal effect of β shown in Figure 4a contrasts with the positive *main* effect of β shown in Table 1 and Figure 6a. In other words, for agents with c and Λ fixed at their baseline values (14.29 and 0, respectively), β has a negative effect. However, when the overall effect of β is averaged across all agents with values of c and Λ within the considered range, it is positive, indicating that β only benefits agents with values of c or Λ that deviate from their baseline

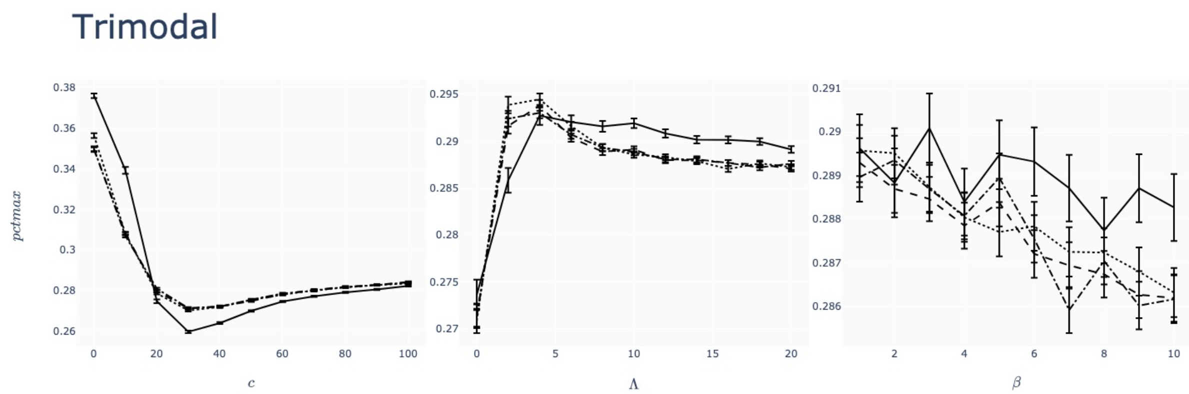
Table 1: Estimated coefficients on a linear regression of the *logodds*(*max*) of our simulations.[a]

	Unimodal	Trimodal	Needle
Intercept	.65	-.89	-12.30
Network structure			
Small world	-.21	.01	.99
Random	-.22	.01	1.04
Lattice	-.21	.01	1.02
c	-2.21	-.23	10.63
Λ	-1.59	.02	11.00
β	.08	-.01	-.33
$c \times \Lambda$	2.51	.18	-14.84
$c \times \beta$	-.13	-.02 (n.s.)	.48
$\Lambda \times \beta$	.04	.01 (n.s.)	.15 (n.s.)
$c \times \Lambda \times \beta$	-.03 (n.s.)	.00 (n.s.)	-.34 (n.s.)
Adjusted R^2	.56	.02	.18

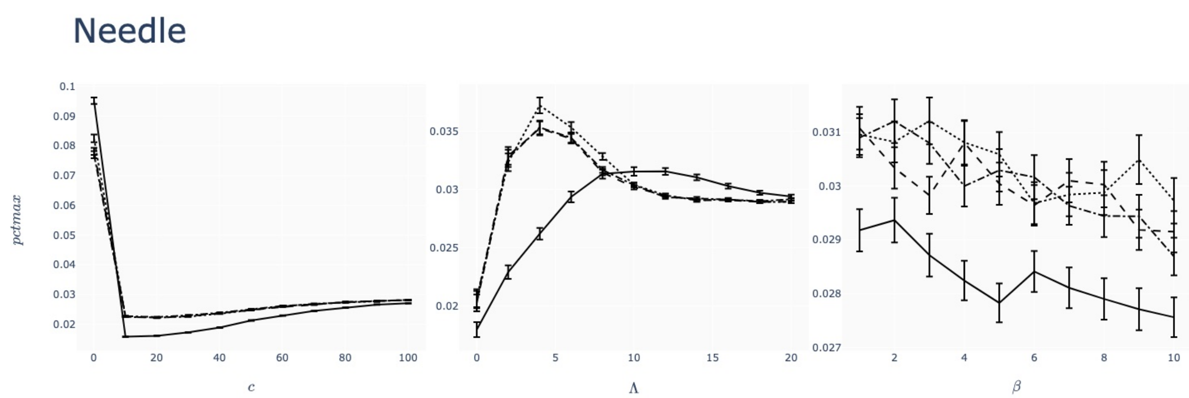
[a] Observations are trials of 15 players interacting over 15 rounds (100 trials per combination of network structure and parameter values; n = 100 × 4 network structures × 11 values of c × 11 values of Λ × 10 values of β = 484,000). A higher *logodds*(*max*) indicates agents in that trial performed better. With six exceptions, all coefficients are significant at at least the $p = .05$ level. The exceptions are indicated in the table by (n.s.).



(a)



(b)



(c)

Figure 6: Interpretation is the same as Figure 4. Unlike in Figure 4, means and standard errors are computed across simulations with varying levels of the other parameter values. For example, plots on the left show the means of 100×11 levels of $\Lambda \times 10$ levels of β simulated trials.

also generally match the shape of the curves shown in Figures 6b and 6c.

One notable difference between the results presented in Table 1 and Figure 6 is the effect of network structure: the coefficients on the *small world*, *random* and *lattice* indicators are negative in the models of the unimodal landscape, indicating that the fully-connected network outperforms the other networks on these landscapes. On the other hand, the coefficients on these indicators are positive in the models of the trimodal and needle landscapes, indicating that the fully-connected network *underperforms* relative to the other networks on the more complex landscape (although relative performance differs only very slightly on the trimodal landscape).

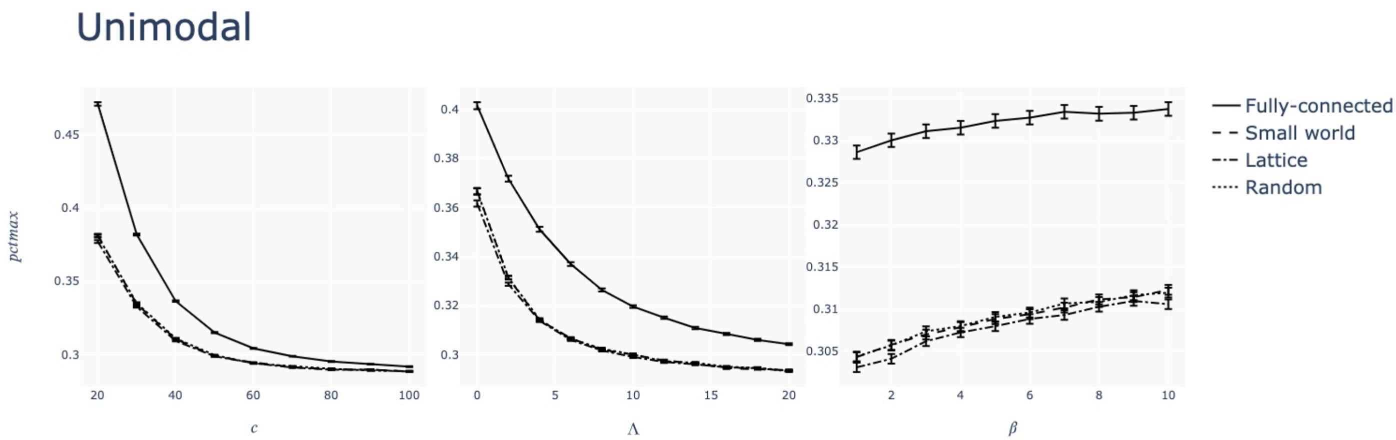
However, this pattern is not immediately apparent in Figure 6: for many parameter settings, the line corresponding to the fully-connected network is above the lines corresponding to the other networks, even on the needle landscape.

Readers may notice that the points in the leftmost plots of Figure 6 corresponding to very small c are extreme outliers, especially on the needle landscape. On this landscape, groups composed of agents with a very small c perform almost an order of magnitude better than other groups. Figure 7 shows the same analysis but with trials where $c < 14.29$ (the baseline value) removed. Once these outlying trials no longer obscure the prevalent effect, the visualized network ordering corresponds to what is suggested by the analysis in Table 1 and what we discuss in section 4: the fully-connected network outperforms the other networks on the unimodal landscape, and underperforms relative to the other networks on the needle landscape.

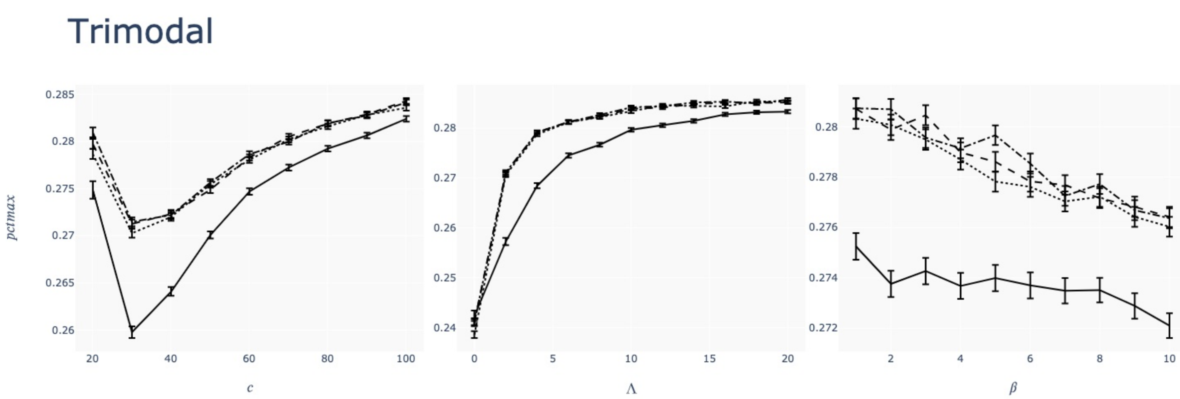
Section 4 discusses why agents who generalize extremely narrowly may perform well on all landscapes, regardless of complexity. But why is the effect so much more dramatic for agents in the fully-connected network? One possibility is that this is a function of the difference in the number of connections each individual has: while the other three networks have on average the same number of edges along which information can travel between agents, the fully-connected network has many more edges—namely, an edge for every possible pair of nodes. By virtue of having access to so many more pieces of outcome information, agents in these networks may both be more likely to stumble across the global maximum *and*, due to their narrow internal generalization gradient, unwilling to leave the global maximum once they've found it.

Finally, we stress the magnitude of the parameter interactions shown in Table 1. The parameters of the SIM, largely due to interdependencies imposed by the structure of the model, interact in both synergystic and counteracting ways in determining group success. Figures 8–10 show how *pctmax* varies as a function of two of the SIM's free parameters simultaneously. In plots of c vs. Λ, β is fixed to 1. In plots of c vs. β, Λ is fixed to 0. In plots of Λ vs. β, c is fixed to 14.29.

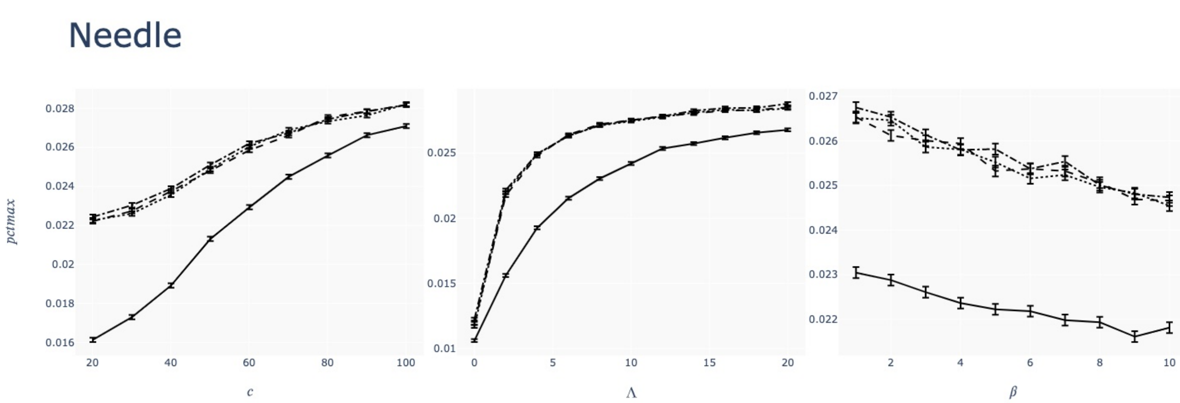
values.



(a)



(b)



(c)

Figure 7: Interpretation is the same as Figure 6. Unlike in Figure 6, we omit trials where $c < 14.29$.

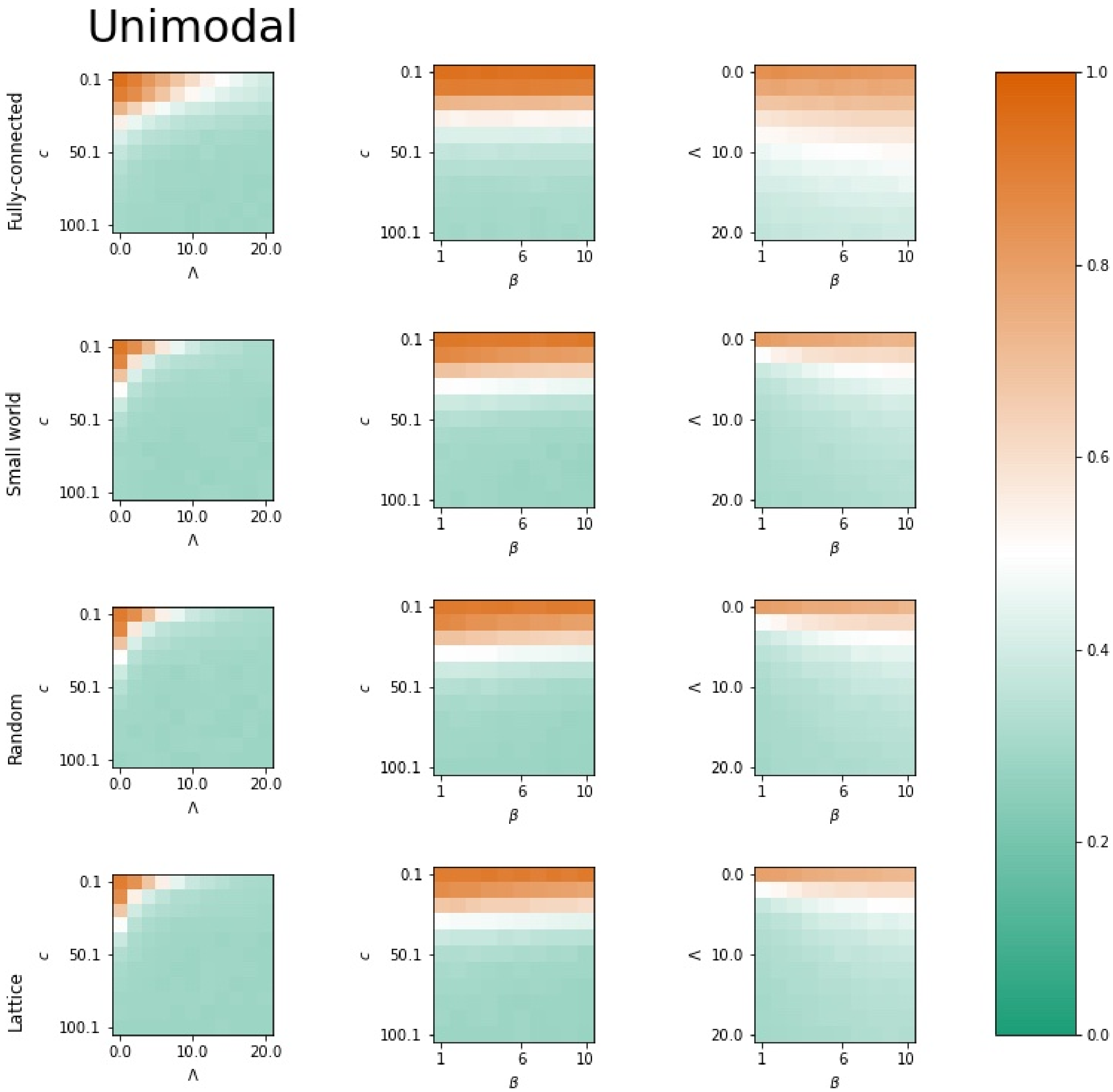


Figure 8: *pctmax* values on the unimodal landscape for different pairwise combinations of parameter values. Rows correspond to performance by groups in different network configurations, and columns correspond to sweeps across combinations of different pairs of parameters (e.g., the leftmost column shows variation in *pctmax* when c and Λ are varied). For plots on the left, β is fixed to 1 (no overweighting of own information). For plots in the center, Λ is fixed to 0 (a prior expectation of no payoff). For plots on the right, c is fixed to 14.29 (corresponding to the gradient of the unimodal and trimodal functions' global maxima).

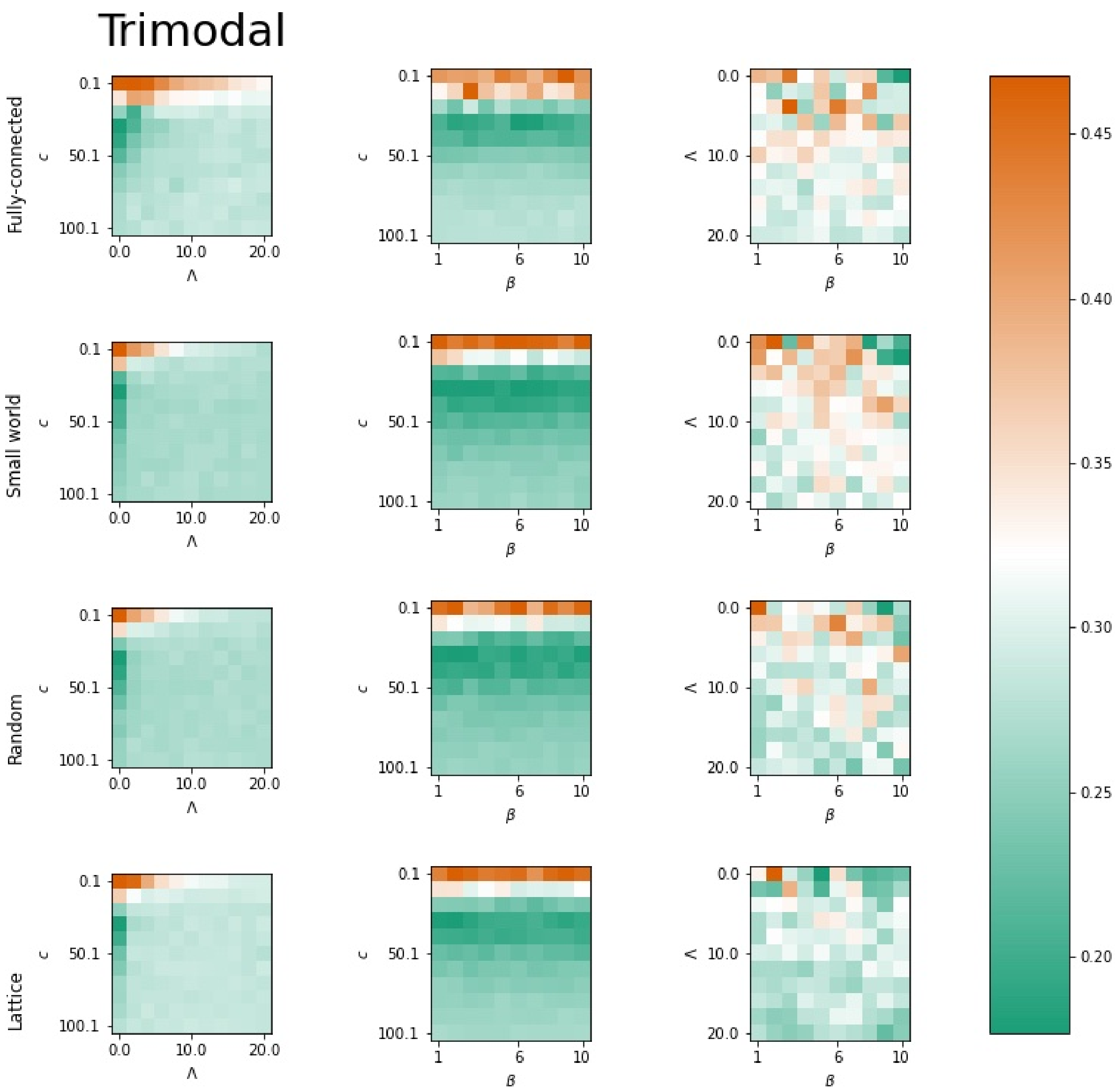


Figure 9: *pctmax* values on the trimodal landscape for different pairwise combinations of parameter values. Interpretation is the same as for Figure 8.

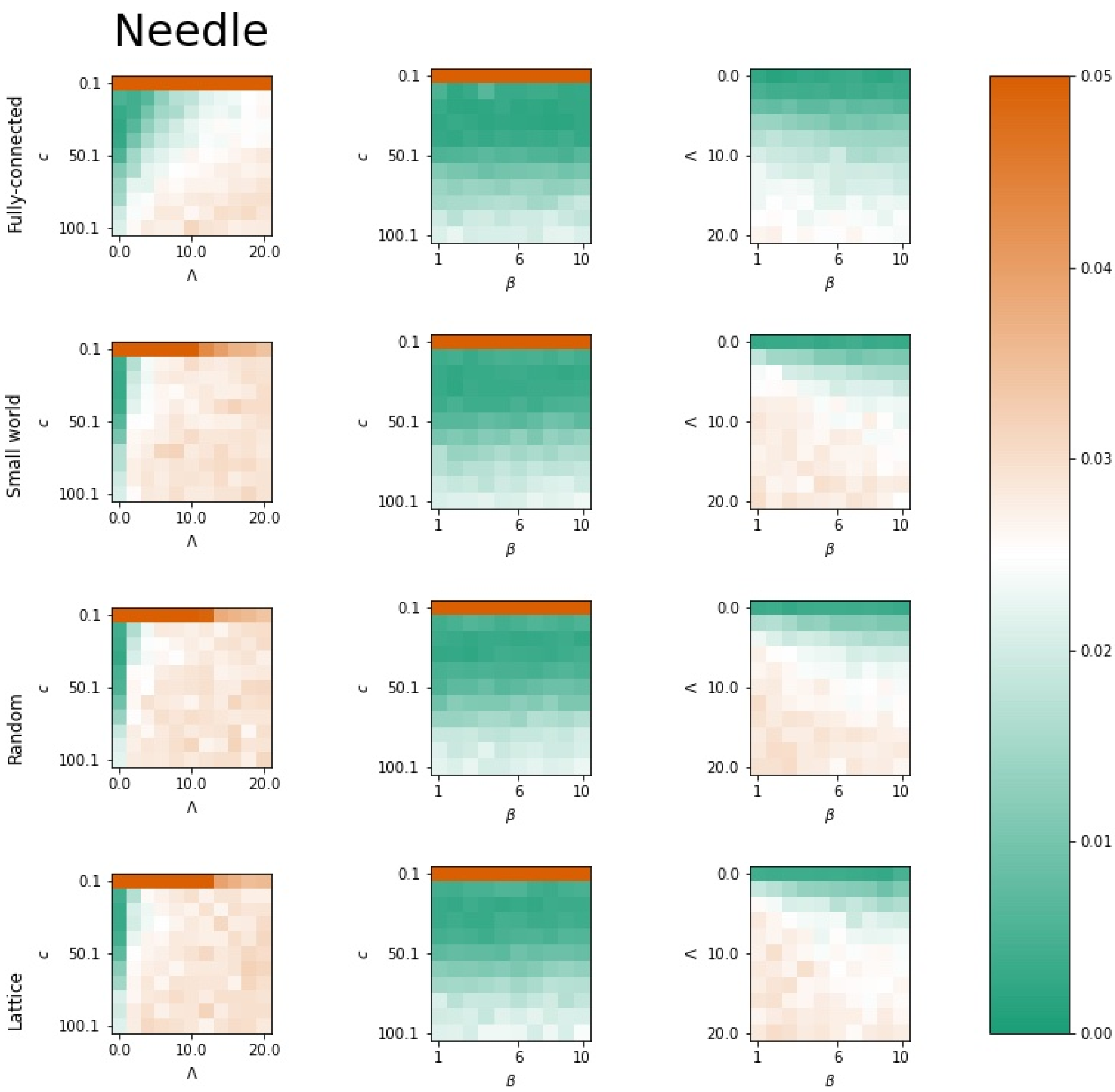


Figure 10: *pctmax* values on the needle landscape for different pairwise combinations of parameter values. Interpretation is the same as for Figure 8.