The Bayesian Superorganism I: collective probability estimation

Edmund R. Hunt[1], Nigel R. Franks[1], Roland J. Baddeley[2]

[1]School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK

[2]School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, UK

ORCIDs: ERH, 0000-0002-9647-124X; NRF, 0000-0001-8139-9604; RJB, 0000-0002-7431-6580

Authors for correspondence: email: edmund.hunt@bristol.ac.uk, roland.baddeley@bristol.ac.uk

Abstract

Superorganisms such as social insect colonies are very successful relative to their non-social counterparts. Powerful emergent information processing capabilities would seem to contribute to their abundance, as they explore and exploit their environment collectively. In this series of three papers, we develop a Bayesian model of collective information processing, starting here with nest-finding, then examining foraging (part II) and externalised memories (pheromone territory markers) in part III. House-hunting *Temnothorax* ants are adept at discovering and choosing the best available nest site for their colony. Essentially, we propose that they estimate the probability each choice is best, and then choose the highest probability. Viewed this way, we propose that their behavioural algorithm can be understood as a sophisticated statistical method that predates recent mathematical advances by some tens of millions of years. Here, we develop a model of their nest finding that incorporates insights from approximate Bayesian computation as a model of collective estimation of alternative choices; and Thompson sampling, as an effective regret-minimising decision-making rule by viewing nest choice in terms of a multi-armed bandit problem (Robbins, 1952). Our Bayesian framework points to the potential for further bio-inspired statistical techniques. It also facilitates the generation of hypotheses regarding individual and collective movement behaviours when collective decisions must be made.

Keywords: superorganism, decision-making, approximate Bayesian computation, Thompson sampling, regret minimisation, multi-armed bandit

1. Introduction

Of fundamental importance to most animals is finding food, mating opportunities, and obtaining a good place to live. The nests of social insects can be considered as examples of extended phenotypes (Dawkins 1999), conferring considerable advantages to the colony's defence, shelter, microclimate, and organisation of function (Kleineidam and Roces 2000, Noirot and Darlington 2000, Pinter-Wollman et al. 2011). While social insects like termites or paper wasps build their nest from scratch, certain house-hunting ants search the environment for pre-existing cavities that can be adapted to their needs. This entails the evolution of an effective collective exploration and decision-making procedure that nevertheless relies on the following of simple rules by individuals making use of local information (Camazine et al. 2001). Two such collective house-hunting systems have been investigated in detail: honeybee swarms looking for a new hive (Seeley and Buhrman 2001) and *Temnothorax* ants (Franks et al. 2003). In the case of *Temnothorax albipennis* ants, even when the nest is intact, colonies still send out scouts looking for better alternatives with an intensity inversely proportional to the quality of their current site (Doran et al. 2013), moving the whole colony if an improvement is identified (Dornhaus et al. 2004). Scouting ants identify candidate nests and recruit other ants one-by-one to inspect it by a process known as tandem running; when enough ants are at a site, a quorum is reached and emigration begins (Pratt et al. 2002, Pratt 2005). Ants have clearly revealed preferences for what they like in a nest (Franks et al. 2003, Franks et al. 2006), and artificial nest choice experiments can be easily conducted in the laboratory. Colonies are very good at discerning even small differences in important attributes such as nest darkness (Sasaki et al. 2013).

In the house-hunting problem the ants need to make a *timely* one-off decision: discovering and moving to the maximum quality nest site. This is distinct from the challenge of optimal foraging, where animals need to make continuous decisions about where to move to next in the environment. The house-hunting decision problem can be fruitfully mapped to the mathematical framework known as the 'multi-armed bandit' problem (MAB), whereby a player (or social unit: the ant colony) faces a choice of projects (nest locations) of unknown value. The name derives from the nickname

55 ‘one-armed bandit’ for slot machines in a casino, where initially one does not know what the
56 probability of making a payoff is for each machine but can only estimate this by repeated play. The
57 MAB has been used to examine the foraging behaviour of animals such as birds (Krebs, Kacelnik et
58 al. 1978, Shettleworth and Plowright 1989, Sherratt 2011), bees (Keasar et al. 2002), fish (Thomas et
59 al. 1985), and slime moulds (Reid et al. 2016). Making choices sequentially, the objective is to
60 maximise the overall reward, which necessitates a trade-off between exploration (observing new
61 locations to ascertain their quality), and exploitation (staying at the current location and making use
62 of its benefits, but missing out on the potential gains to be had elsewhere). This trade-off is
63 ubiquitous across biological domains, and is a key evolutionary force in the development of
64 cognition (Hills et al. 2015). Ideally an ant colony would observe a few different locations, quickly
65 find the one of the highest available quality, and then move there, so that its process of foraging and
66 reproduction can continue apace.
67 The multi-armed bandit (MAB) is difficult to analyse as a theoretical problem to find the
68 optimum solution, though in certain special cases this can be computed (Bellman 1956, Gittins
69 1979). A ‘Gittins index’ may be calculated and then the problem simply becomes choosing the option
70 with the highest index value (Whittle 1980, Gittins et al. 2011). Yet there remains a possibility of
71 settling on playing the wrong arm forever, known as incomplete learning (Brezzi and Lai 2000). This
72 is analogous to the ants settling on what seems like a good quality nest, when an excellent one is
73 available lying undiscovered some short distance away.
74 Optimal solutions to the MAB are difficult to compute, especially when there are many
75 available options. Fortunately, rules of thumb are available to cope with the challenging
76 exploration–exploitation trade-off, and heuristics that approach the theoretical optimum are likely
77 to have been evolved by animals that are effective in their particular ecology. One decision-making
78 rule is known as Thompson sampling (Thompson 1933), also called randomised probability matching
79 (Scott 2010), which was originally developed in the context of medical trials. There, if one has two
80 experimental treatments, it is desirable to switch some of the patients from one treatment to the

other if it is showing promising results, but without giving up too quickly on the alternative treatment. It has been rediscovered several times independently in the context of reinforcement learning (Wyatt 1997, Strens 2000, Ortega and Braun 2010). Thompson sampling draws on Bayesian ideas, whereby the player (colony) has an assumed set of prior payoff distributions for each arm, and after making observations updates these to a posterior distribution using Bayes' rule. The use of Bayes' theorem and its assumption of prior 'opinions' about outcomes has been argued to be reasonable in the context of animal behaviour, given both animals' past individual experience of the environment, and the adaptation to that environment by previous generations (McNamara et al. 2006). In Thompson sampling, the player makes a random draw from these distributions, and chooses the arm (location) associated with the largest sample for its next observation. By making random draws from the player's own distributions, which become more accurate with further observations, this decision-making rule permits the player mostly to choose the arm it believes to be the best. At the same time, it leaves open the possibility of making further observations from what the player believes to be a lower quality arm, if by chance it throws up a high sample value. In the case of searching for a new nest site, scouting ants are looking for the global quality maximum, and thus do not want to waste time returning to locations they know to be of lesser quality relative to what they have already observed. In a world of noisy observations, however, it can be imprudent to rule out an option prematurely. The Thompson sampling strategy achieves an effective compromise for this problem, and minimises what is known as 'regret' – the performance relative to a theoretical optimum (Agrawal and Goyal 2012, 2013). This effectiveness is borne out in empirical studies of its effectiveness (Chapelle and Li 2011), and it has been claimed to be a good model of humans responding to the 'restless' MAB problem, when the payoffs change through time (Speekenbrink and Konstantinidis 2015). Thus, we hypothesize that ant behaviour should approach a strategy resembling Thompson sampling.

We develop a simplified model of house-hunting that does not include spatial traversal. Of course, in reality, ants must explore a pseudo-two-dimensional environment looking for potential

nest sites. Finding nest sites is a difficult challenge, and having to select between multiple options at the group level compounds the problem. Nevertheless, *T. albipennis* is an example of an ant species that does this very effectively: for example, they are able to choose a distant superior nest over an in-the-way poor one (Franks et al. 2008). *T. albipennis* makes use of a highly adaptive behaviour known as tandem running to share information about the location of a nest or food resources (Franks and Richardson 2006). In the vicinity of another ant, a leader ant with valuable private information about the position of such a resource releases a particular recruitment pheromone, and the naïve ant will follow it on a somewhat slow and tortuous route to the location of interest. When the follower ant arrives, it can make its own observations of the location, independently deciding whether it is indeed a high-quality resource such as a new nest site. If it is also enthusiastic about the location, it can start leading its own tandem runs, in a positive feedback mechanism that allows a significant number of ants to discover the resource much more quickly than if they were all independently exploring. Recent work (Sasaki et al. 2018) finds that *T. rugatulus* colony decision-making behaviour to be consistent with the *Sequential Choice* model, where competing options are evaluated in parallel, in a race to a decision threshold; whereas single ants match the *Tug of War* model where alternatives are directly compared and take longer as a result. This paper proposes an explanation for how this emergent change in behaviour is possible: how individual rejection sampling can lead to a global distribution of ants that is computationally meaningful.

We develop a sequence of models in turn, before presenting results from simulations of those models and discussing their features. These include (1) effective individual sampling (the Thompson strategy); (2) many such individuals sampling in parallel, with their emergent global distribution identifying a probability of an option being the best choice (approximate Bayesian computation); and (3) occasional communication between individuals when high-quality options are found (tandem running).

2. Models and simulations

We develop three models that include progressively more behavioural features of an ant colony and become more effective in quickly identifying the highest quality potential nest location.

Model 1 has (to motivate the problem) one ant trying to find the highest quality potential nest site out of three available. Model 2 then introduces the concept of approximate Bayesian computation (ABC) to explain the action of multiple house-hunting ants, and this model is found to identify the highest quality site more quickly than in the single ant case. Finally, in Model 3 tandem running is introduced whereby multiple ants explore in parallel, independently; but if a subjective, uniformly distributed quality threshold is exceeded in an ant's estimation, another ant (randomly chosen) also observes that location at the next time iteration. Along with the ABC character of the ants' decision-making process, this allows the highest quality location to be rapidly identified.

2.1 Model 1: Single ant, three locations

We begin with a simple model of the ant nest site selection process. There is a single ant and three potential locations for the ant to locate the nest. The existence of these three locations is known *a priori*, and there are no locations unknown to the ant. The ant begins with a set of prior estimates of the quality of each location, which is simply uniform to begin with. These are expressed as a separate mean and standard deviation for a normal distribution, μ_i and σ_i, for each location $i = 1,2,3$. The standard deviation expresses how confident the ant is in its estimate, and reduces with repeated observations. The locations have a true and constant quality q_i that is accessible to the ant by making repeated noisy observations. These observations are drawn from a normal distribution with a mean q_i and a standard deviation which is the observation error, σ_{obs}, a fixed parameter 'known' to the ant. This encapsulates the fact that the methods ants use to find and assess the characteristics of a potential nest are associated with noise and error. For instance, *T. albipennis* has been shown to estimate the floor area of potential nest sites using the 'Buffon's needle' algorithm (Mallon and Franks 2000). By laying an individual-specific pheromone trail inside a

potential nest site and later revisiting it one or more times, an ant can estimate its area by counting the number of times it intersects with previous trails (more crossed paths implies a smaller area). The overall quality of a potential nest site will be a combination of such noisy measurements, with repeated observations allowing the error or uncertainty in its estimation to be progressively reduced (this error, because of the Gaussian nature of the quality signal, reduces as a square root of the number of observations, but other noisy observation models would be easy to incorporate).

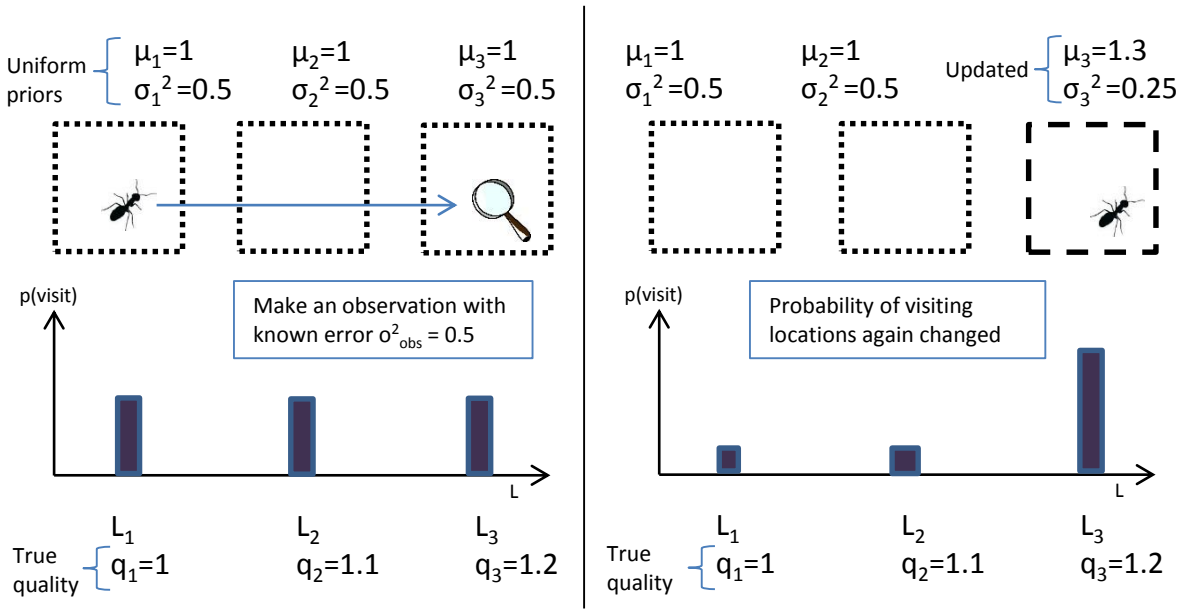

Figure 1. An ant initially has uniform prior beliefs about unobserved nest qualities; thus there is a uniform probability of visiting each of the three locations. After drawing a sample from each of these models, it goes to the location with the highest sample, and makes a noisy observation of its quality. It updates its estimate of that location's quality (Location 3 here), and the estimated mean quality, and its associated uncertainty with it is updated. Since Location 3 now has the highest estimated quality, it also has the highest probability of being observed on the next exploration trip.

The ant makes a sequence of observations of the different locations to try and find the one with the highest quality. To decide which location to travel to and observe next, it draws a sample simultaneously from each of its subjective location quality estimates, $Q_i \sim \mathcal{N}(\mu_{i_1}, \sigma_i)$, and moves to the location with the highest drawn sample. This is a form of Thompson sampling, or randomised probability matching. While this will tend to move the ant to the location with the highest quality estimate μ_i, its probabilistic nature means that sometimes a location with a lower quality estimate

will generate a sample higher than that from the better-quality estimates, especially if the uncertainty in that estimate σ_i is still relatively high.

When the ant makes an observation, for instance of Location 3, it updates its estimate of the location's quality using Bayes' rule (see Figure 1). With a known observation error σ_{obs}, after making an observation of Location 3, O_3, the ant's posterior distribution of nest quality Q_3 can be shown to be:

$$Q_3|O_3 \sim \mathcal{N}\left(\frac{\frac{O_3}{\sigma_{\text{obs}}^2}+\frac{\mu_3}{\sigma_3^2}}{\frac{1}{\sigma_3^2}+\frac{1}{\sigma_{\text{obs}}^2}}, \frac{1}{\frac{1}{\sigma_3^2}+\frac{1}{\sigma_{\text{obs}}^2}}\right)$$

This is a weighted average of the observed quality value O_3, with the ant's prior estimate of it μ_3, according to the associated observation error σ_{obs}, and the estimate of the error in μ_3 given by σ_3. If for instance the measurement error is low and the prior uncertainty was high, the posterior estimate of μ_3 will be close to the observed quality value. The above procedure of making a biased exploration based on current quality estimates, and updating those estimates based on observations of consequently chosen locations, can be iterated through a series of observations and will converge through time upon the true location quality Q_3. A review of empirical studies indicated that many animals (birds, mammals, fish, insects) exhibit behaviour consistent with such Bayesian updating of environmental parameter estimates (Valone 2006).

The ant's running estimates of μ_i and σ_i for each location $i = 1, 2, 3$ encodes a summary statistic that captures all that an ant knows about each of the qualities at each step in time, and given that the mean and standard deviation of a Gaussian are sufficient statistics for a normal distribution, no memory of the individual past observations is needed. As more observations are made, the uncertainty $\sigma_i$in the estimate of each location's quality reduces toward zero, and thus it becomes less and less likely that the sampling step will lead the ant to revisit locations of lower quality. Eventually the ant stays in a particular location, provided it has a quality higher than the other alternatives, and this represents a choice being made in favour of that location.

Although the ant in this model can make its choice fairly quickly, since there are only three locations to visit, it still requires several observations to be made. In practice, house-hunting ants like *T. albipennis* make their choice using several scouting ants, after each makes only a few observations of one or two of the potential locations: and it can still make accurate choices because it makes its decisions collectively with a quorum sensing mechanism (Pratt, Mallon et al. 2002). The next model shows how this is analogous to approximate Bayesian computation (ABC).

2.2. Model 2: Multiple ants, three locations

In this model, individual ants make observations in the same way as in the single ant model, and do so independently. However, their global distribution across the three locations is used as a running estimate of the posterior (post observation) probability that each location is the best available place to establish the colony's nest.

ABC methods seek to bypass the evaluation of the likelihood function in Bayes' rule in determining the posterior probability distribution. All ABC-based methods approximate the likelihood function by simulations, the outcomes of which are compared with observed data. The ABC rejection algorithm samples a set of parameter points from the prior distribution, simulates data sets under the statistical model specified by each of those parameters, and rejects those parameters whose simulated data set is too different from the observed data set (summarised by a measure such as its mean), based on some distance measure between them. In the case of ants, they make an observation of a location, and reject it if it is not sufficiently high quality and move elsewhere, or accept it (stay there) if it is high quality. *T. albipennis* ants do not need individually to make direct comparisons between alternative choices to collectively choose the best nest site (Robinson et al. 2009). An ant's current location (which is accepted or rejected) is like one of the sampled parameter values in the described ABC rejection algorithm, and the judgement of sufficiently high quality is like the comparison with the observed data set mean. The resulting

parameter sample (or ant locations) is an approximation of the desired posterior probability – in the ants' case, the probability that a location is the best, taking into account the ants' observations.

As an example, we can simulate five exploring ants, which for a period of time make observations and decide at each time step whether to 'accept' a site (stay in the location, exploitation) or 'reject' a site (move to a new location, exploration). At the end of the time period there are four ants at Location 3, one ant at Location 2, and no ants at Location 1 (see Figure 2). This corresponds to a probability of 80% (four out of five ants) that Location 3 is the best location; similarly 20% and 0% for locations 2 and 1. Psuedocode for this algorithm is provided in Table 1.

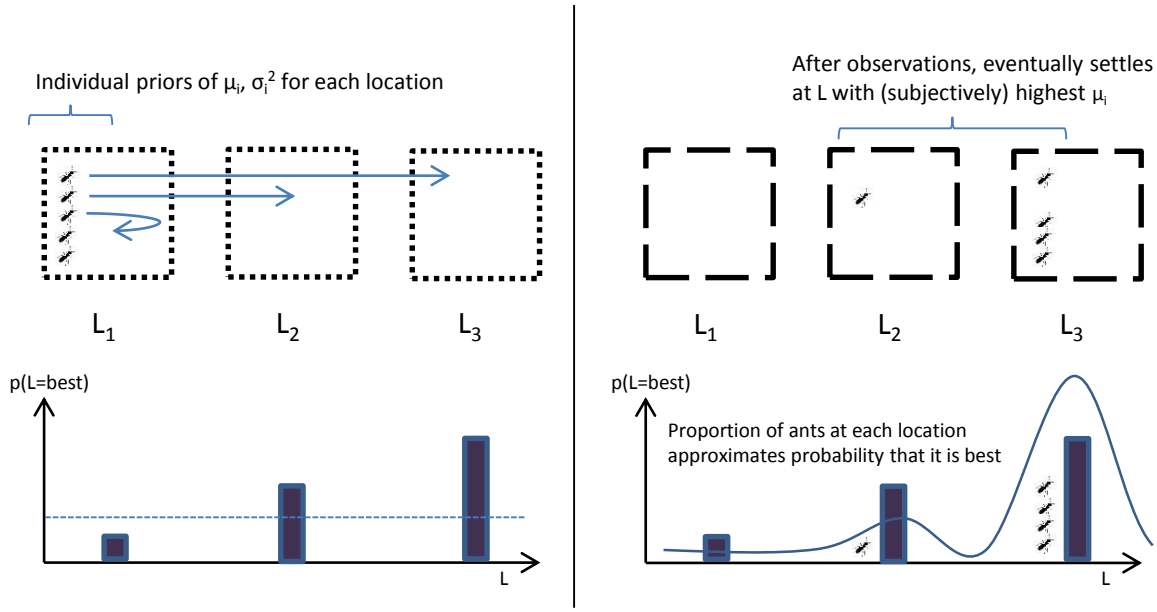

Figure 2. Multiple non-interacting ants engage in Thompson sampling until they settle on the location each believes is the highest quality. Their macroscopic distribution approximates the true posterior distribution of location qualities.

The global distribution of a sufficient number of ants (ten, for example) is an approximation of the posterior probability that each choice (potential nest site, for instance) is the best available; when a threshold probability (resource quality) is exceeded a decision can be said to have been reached. Allowing many ants to explore in parallel allows a decision to be taken more quickly, because while the total amount of ant exploration time (number of ants × time) may be similar, the actual time taken can be much less. This is important for biological feasibility, since an individual ant cannot explore indefinitely, and is unlikely to discover all available sites independently. This ABC

model is clearly more realistic than the single ant model in that it captures the emergent, statistical phenomena whereby the spatial distribution of scouting ants determines the collective decision. However, one obvious simplification made is that the ants are non-interacting, such that each ant has to engage in a costly search process (in terms of time and energy) and make its own observations of a location that may have already been visited many times by its nestmates. It is also possible (though unlikely in this model with only three locations) that an ant will fail to make (accurate) observations of the highest quality location, and thus make a suboptimal choice by settling on a lower quality location as its choice.

Table 1. Pseudocode for simple approximate Bayesian computation nest finding method.

Pseudocode for approximate Bayesian computation nest finding method for i nest locations
input: number of ants N, location qualities $q_1, q_2, \dots, q_i$ independent uniform priors μ_i, σ_i^2 for each ant, fixed observation error σ_{obs}^2, number of simulation time steps T output: posterior probability estimate that each location is the best, $p_1, p_2, \dots, p_i$

for t = $1{:}T$
  for ant = $1{:}N$
make random draw from each location's running quality estimate, $\mathcal{N}(\mu_{i_1}, \sigma_i)$
make a quality observation from the location with the highest random draw
make a precision-weighted update of quality estimate μ_i and its uncertainty σ_i^2
  end
end
calculate posterior probabilities that each location is best
$p_i = \sum(\text{ants at location } i)\ /N$

2.3. Model 3: Multiple ants, tandem running

The final model has multiple ants exploring the target distribution in parallel, whose global distribution estimates the probability that a particular location is highest in quality, as per the analogy with approximate Bayesian computation developed in the three-location model. It additionally includes a representation of the *T. albipennis* tandem running behaviour, whereby a scouting ant that has found a high-quality resource (like a potential new nest site) finds another worker ant and attempts to lead her to that spatial position. This is illustrated in Figure 3. The interaction between individuals is central to the emergence of collective sensing of the environment

(Berdahl et al. 2013), and so even simple interactions are an important feature of a realistic model. The tandem running behaviour is implemented in the model by assigning each scouting ant a threshold quality, whereby an observation above that threshold will trigger it to lead another scouting ant to that location. This threshold can also be interpreted as a probability, if one views the 'quality space' of potential nest sites instead as 'probability space' that each location is the best available: with a simple transformation the two concepts can be seen as equivalent. The tandem run behaviour can be understood as a form of 'particle filter' method (also known as Sequential Monte Carlo), which boosts sampling from higher quality (probability) space. The threshold p_{lead} is uniformly distributed in this model, but we also consider the possibility of it being heterogeneously distributed in the discussion. When an ant observes a high-quality location that exceeds p_{lead} another ant from all of those exploring is randomly chosen for an attempted recruitment. If that ant's estimate of its current location's quality is lower than or equal to the leader ant's estimate of the proposed location, it relocates to the leader ant's position and makes a new observation of it. If the quality of the tandem initiator's location, as estimated by the leader, is less than that estimated by the potential follower at its current position, it resists recruitment and stays put. Pairwise comparison between a proposed location and the current location has been referred to as a 'duelling bandits' scenario (Yue et al. 2012), although it is somewhat different in this context because we use two separate sources for the proposed and current locations (leader and follower ants).

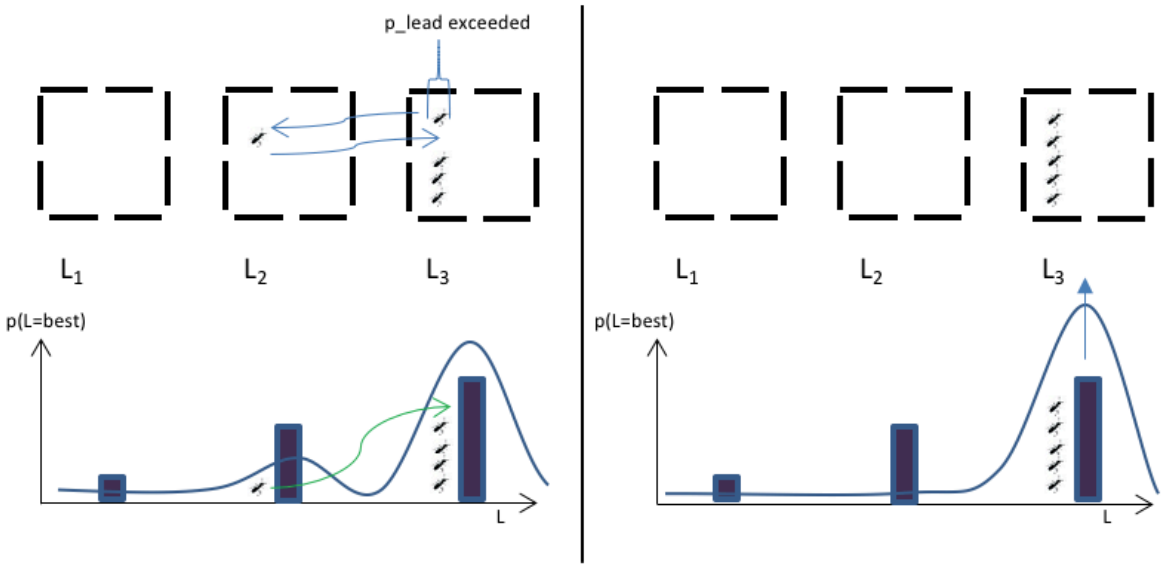

Figure 3. The tandem running behaviour acts like a particle filter, boosting sampling from higher quality locations. Ants, having discovered a high-quality location, recruit nestmates to also observe that location. This results in a posterior distribution (global distribution of ants) that is more likely to converge upon the true state of the environment (right pane).

3. Results

3.1 Model 1: Single ant, three locations

An example simulation of Model 1 is shown in Figure 4. The ant starts with uniform priors over the three locations, $\mu_{i=1:3} = 1$, $\sigma_{i=1:3} = 0.5$, makes observations with fixed, known error $\sigma_{\text{obs}} = 0.5$, while the actual location qualities are distributed as $q_1 = 1, q_1 = 1.1, q_1 = 1.2$. As per the Thompson sampling procedure, it draws samples from its estimated quality distributions, and moves to the location producing the highest sample, where it then makes an observation that decreases its uncertainty in the estimate. Over the course of 100 time steps in this example, its location changes several times, until it begins to narrow its search to locations 1 and 2 from t=14 onwards. It then settles on Location 3 as the highest quality from $t = 56$, when it does not move any more (the samples from its estimated quality distributions always produce Location 3 as the highest quality). Its running estimate of the location qualities is shown in the second pane in Figure 4; its estimate for locations 2 and 3 becomes accurate as it makes repeated observations, while its lack of observations for Location 3 means its quality is somewhat underestimated. The uncertainty in the ant's estimates

for locations 2 and 3 steadily decreases as new observations are made. Because the ant acts alone it must make several observations of each location before it can be confident that it has identified one with the highest available quality. In practice, an advantage of the colony's tight-knit social organisation is that multiple ants can explore and make observations of different location qualities in parallel, so that the process is greatly accelerated. Multiple parallel assessments are a feature of the next model.

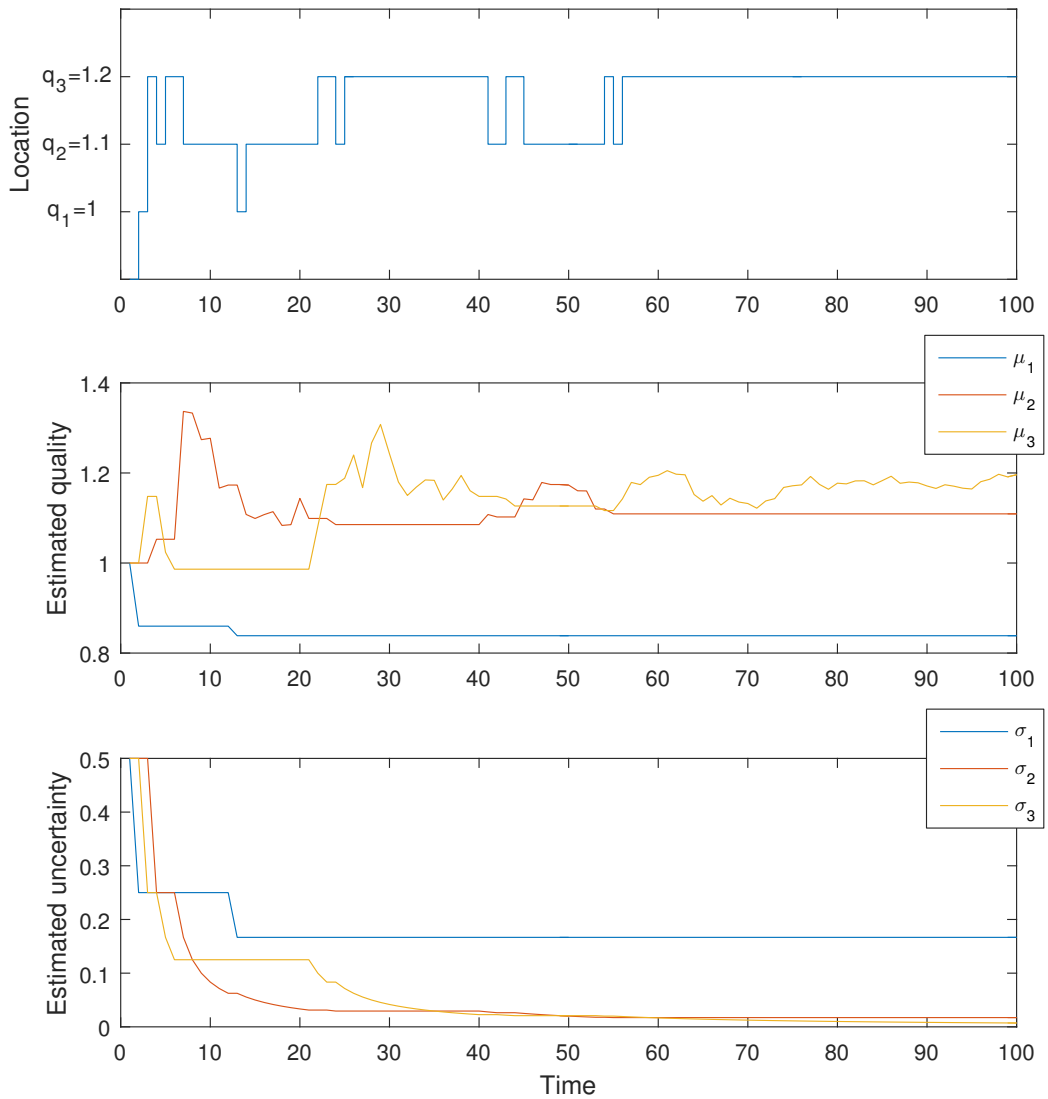

Figure 4. Model 1: Thompson sampling to find the best nest, 1-dimensional model, single ant. Location 3 has the highest quality, and after sampling the locations the ant makes several observations from it and accurately estimates its quality, $q_3 = 1.2$, and stays in that location after around 60 iterations.

3.2. Model 2: Multiple ants, three locations

With the Thompson sampling procedure working as before, this model initiates multiple ants exploring in parallel. Figure 5 shows the result of simulating ten ants for 100 time steps, with the number of ants in each location shown on the y-axis. This number corresponds to the probability that each location is the best, when normalised, from the level of the colony's 'collective cognition'. In this case, a high threshold of 60% probability (6 of 10 ants) is breached quickly at $t = 4$ for Location 2, and then 80% at Location 3 at t=6. This is followed by a clear lead for Location 3 at around $t = 30$. At the end of 100 simulated time steps, the model has settled on probabilities of 80%, 20% and 0% for each of locations 3, 2 and 1 being the best available quality.

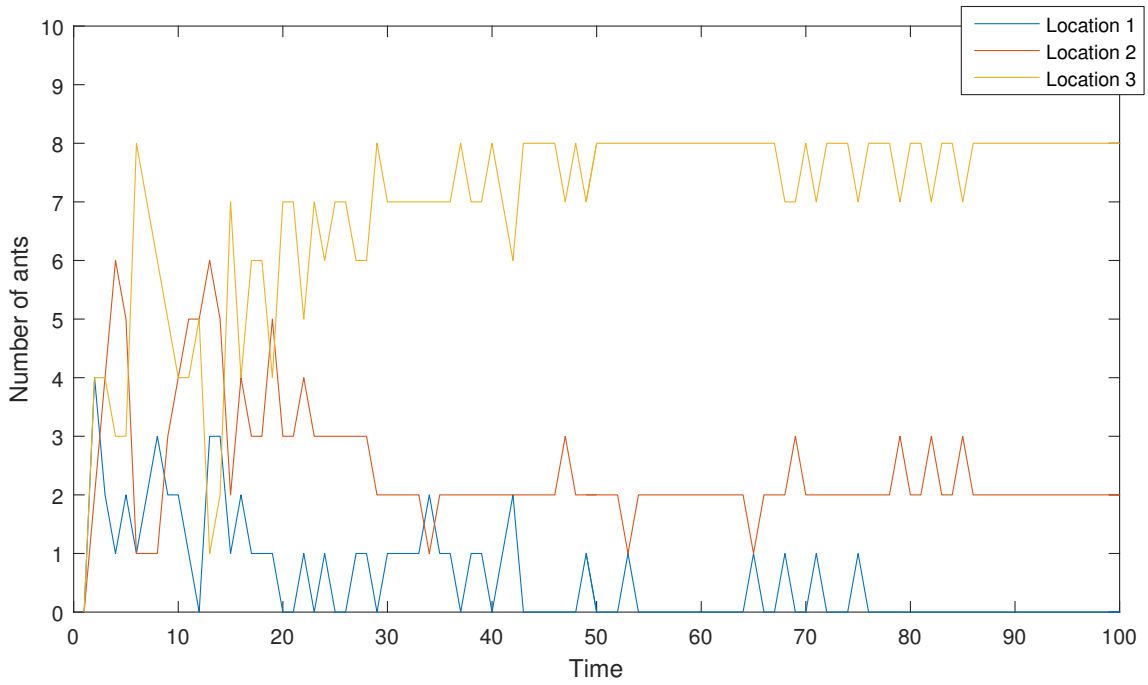

Figure 5. Model 2: Multiple (ten) ants independently exploring three locations. Their global distribution approximates the probability that each location is the highest quality available. At t=100, there are 8, 2, and 0 at locations 3, 2, and 1, or probabilities of 80%, 20%, and 0% that these are the best choices. A clear lead for Location 3 is established much earlier at around $t = 30$.

This model highlights two aspects of real ant colony decision-making: first, that working collectively allows the optimal choice to be identified more quickly; but second, that it is possible for premature decisions to be made in favour of a site of good quality (Location 2) over the site of the best quality

(Location 3) if a threshold is set too low or breached too quickly. For example, with a threshold of 60% probability a quorum in favour of Location 2 may have been met at $t = 4$. It is important, then, to ensure that available locations of high quality are identified in a timely way, since their discovery may otherwise come too late in the nest finding process; and to have a quorum threshold set at a suitable level.

In the final model multiple ants explore in parallel and also engage in a representation of the *T. albipennis* tandem running behaviour to preferentially sample from high quality regions.

3.3. Model 3: Multiple ants, tandem running

An example of the third model's output, for the same location qualities as before ($q_1 = 1, q_1 = 1.1$, $q_1 = 1.2$), is shown in Figure 6. A high estimated quality threshold of 1.25 is set in this case. The potential follower will move to the high-quality location only if the leader's estimate of that location's quality is higher than the follower's estimate of her current position's quality. After some initial volatility, as ants switch between location 2 and 3, a clearer preference for location 3 is established than in Model 2 (c.f. Figure 5) after around t=25. Because the tandem running threshold p_{lead} has been set with reference to what is available in the environment, it begs the question of how this threshold may be set in real ants that do not have any preview into their surrounding options. This question is explored in the discussion.

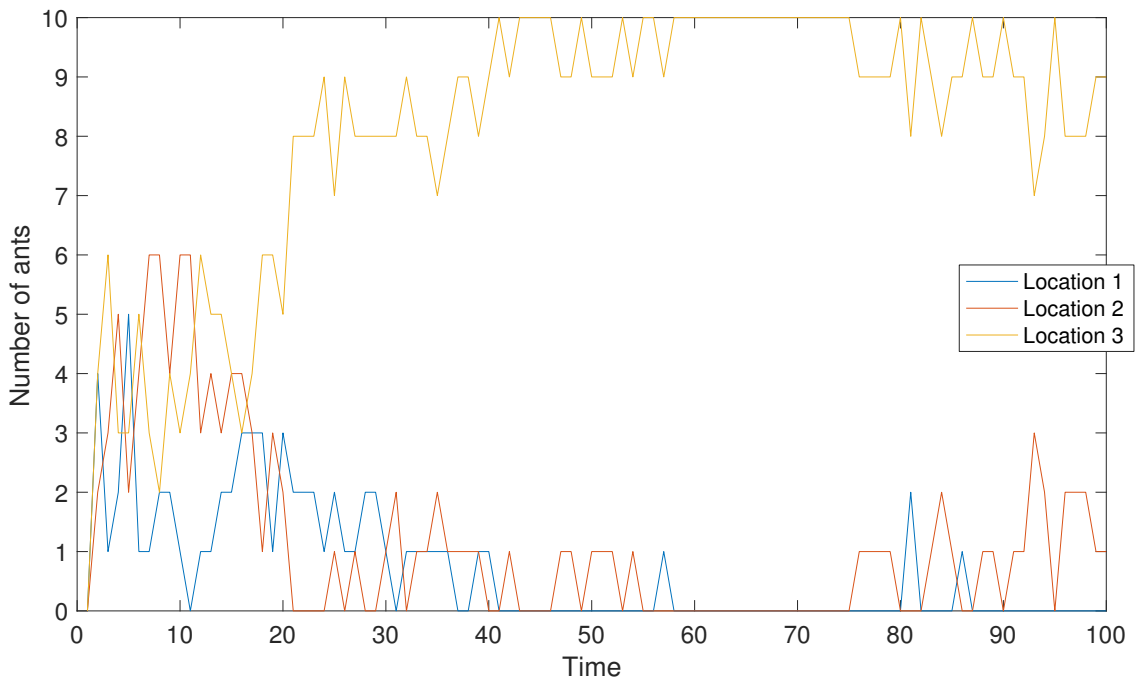

Figure 6. Multiple (ten) exploring ants, with a tandem running (particle filter) behaviour included. Each timestep, after all the ants have made their choice of location and updated their local quality estimate, those at a high-quality location are able to attempt recruitment. If their quality estimate is above a threshold q_lead, they attempt to lead a random ant to that location. If the leader's estimate of the proposed location is above the potential follower's local estimate, it relocates to the new position and makes a new observation of it.

4. Discussion

Three models of *Temnothorax albipennis* nest finding have been developed that progressively include more realistic features, and that are more efficient and effective at selecting the best available location for the new nest in a timely way. The three-option models use a Thompson sampling approach to solve them as a 'multi-armed bandit problem'. Tandem running is introduced in the final model with multiple ants to ensure that information about the position of high quality locations is shared among the ants. This necessitates the introduction of a new parameter p_{lead} , which is a threshold probability (quality) above which tandem runs are initiated by the observer. While an appropriate p_{lead} is set in this model by reference to the known overall distribution of qualities, in nature the ants do not know the distribution of location qualities a priori. However, the ants' evolutionary 'experience' should have favoured the setting of a p_{lead} that is appropriate to their typical environment, since colonies that have made beneficial use of the tandem running

behaviour to help them reproduce are likely to have avoided both the pitfalls of p_{lead} being too high (so the behaviour is never used) and p_{lead} being too low (the behaviour is used excessively to share low value information). In practice, it may be biologically difficult to set a uniform intrinsic activation threshold for tandem running, because although *T. albipennis* worker ants are highly related (0.75 relatedness in haplodiploid social insects that share the same mother and father) there is still some degree of genetic heterogeneity, not to mention differing life experience histories. A heterogeneous distribution of p_{lead} is therefore likely, and perhaps even desirable, since it allows information of variable value to be propagated through the colony and collectively assessed. The optimal and empirical distribution of thresholds in social insects is an active topic of research (Jones et al. 2004, Robinson et al. 2012, Ishii and Hasgeawa 2013) and models have been developed showing the effectiveness of heterogeneous thresholds (Masuda et al. 2015). In addition to genetic influence over behavioural thresholds, environmental experience is also likely to be an important contributing factor to thresholds. For example, current nest site quality could affect perceptions of what is generally available in the environment, and in turn lead to adjusted thresholds for what is a good nest site option worthy of initiating tandem runs.

One aspect of the nest finding process that is not explicitly included in the models is the ants' quorum sensing mechanism (Pratt et al. 2002), whereby a decision is taken in favour of a location when individual ants have a high enough encounter rate with other ants, indirectly estimating the number of ants at that location (Pratt 2005). In the case of the ABC-based models 2 and 4, the quorum threshold could be introduced simply by specifying a probability p_{quorum} , which is exceeded when enough ants are at a location. When the need to identify a new nest is urgent, for instance when the colony's nest is damaged, a high threshold will be a handicap. In the case where a new nest is needed quickly, the quorum threshold for a decision is lowered, sometimes at the expense of accuracy (in terms of choosing the best option) (Franks et al. 2003); this speed-accuracy trade-off is faced by animals in many contexts (Chittka et al. 2009). As a result the medium quality

site may be estimated as the best site. Nevertheless, the Thompson–ABC model presented here would still be a good representation of this aspect of the ants' behaviour.

Conclusion

We have developed a model of the *Temnothorax* nest choice process, which incorporates both the environmental exploration and collective decision phases. The technique of approximate Bayesian computation is used as an analogy to describe in informational terms how the ants collectively estimate the probability of a location being the best available; and a decision rule about where to move individually is inspired by Thompson sampling, in the context of the multi-armed and duelling bandits problems. This model provides important insights into how the collective cognition of the 'super-organism' takes place during the nest finding problem.

Alternatives to the Thompson sampling procedure have recently been developed, such as 'Optimistic Bayesian Sampling' (OBS) (May et al. 2012) in which the probability of playing an action increases with the uncertainty in the estimate of the action value. Simulations suggest that this may result in better-directed exploratory behaviour while outperforming Thompson sampling for the example problems presented (May et al. 2012). Ultimately, empirical aspects of the *Temnothorax* collective choice process can be further compared with theoretical optimal behaviour derived from new developments in sampling algorithms, to see whether natural selection has anticipated these decision-making rules with its own evolved heuristics. Similarly, the continued careful study of collective choice in various superorganisms may inspire new improvements to sampling methods for the multi-armed bandit.

Author contributions. ERH drafted the paper and ran the simulations, NRF advised on social insect biology, RJB originated and advised on the theoretical concepts; all authors contributed to the present draft.

Funding statement. E.R.H. thanks the UK Engineering and Physical Sciences Research Council (grant no. EP/I013717/1 to the Bristol Centre for Complexity Sciences).

Competing interests. We have no competing interests.

References

Agrawal, S. and N. Goyal (2012). Analysis of Thompson Sampling for the Multi-armed Bandit Problem. 25th Annual Conference on Learning Theory, Journal of Machine Learning Research.

Agrawal, S. and N. Goyal (2013). Further Optimal Regret Bounds for Thompson Sampling. AISTATS.

Bellman, R. (1956). "A PROBLEM IN THE SEQUENTIAL DESIGN OF EXPERIMENTS." Sankhya **16**: 221-229.

Berdahl, A., C. J. Torney, C. C. Ioannou, J. J. Faria and I. D. Couzin (2013). "Emergent Sensing of Complex Environments by Mobile Animal Groups." Science **339**(6119): 574-576.

Brezzi, M. and T. L. Lai (2000). "Incomplete Learning from Endogenous Data in Dynamic Allocation." Econometrica **68**(6): 1511-1516.

Camazine, S., J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz and E. Bonabeau (2001). Self-organization in biological systems, Princeton University Press.

Chapelle, O. and L. Li (2011). An empirical evaluation of thompson sampling. Advances in Neural Information Processing Systems.

Chittka, L., P. Skorupski and N. E. Raine (2009). "Speed-accuracy tradeoffs in animal decision making." Trends in Ecology & Evolution **24**(7): 400-407.

Dawkins, R. (1999). The extended phenotype: the long reach of the gene. Oxford, UK, Oxford University Press.

Doran, C., T. Pearce, A. Connor, T. Schlegel, E. Franklin, A. B. Sendova-Franks and N. R. Franks (2013). "Economic investment by ant colonies in searches for better homes." Biology Letters **9**(5).

Dornhaus, A., N. R. Franks, R. M. Hawkins and H. N. S. Shere (2004). "Ants move to improve: colonies of Leptothorax albipennis emigrate whenever they find a superior nest site." Animal Behaviour **67**: 959-963.

Franks, N. R., A. Dornhaus, C. S. Best and E. L. Jones (2006). "Decision making by small and large house-hunting ant colonies: one size fits all." Animal Behaviour **72**(3): 611-616.

Franks, N. R., A. Dornhaus, J. P. Fitzsimmons and M. Stevens (2003). "Speed versus accuracy in collective decision making." Proceedings of the Royal Society B-Biological Sciences **270**(1532): 2457-2463.

Franks, N. R., K. A. Hardcastle, S. Collins, F. D. Smith, K. M. E. Sullivan, E. J. H. Robinson and A. B. Sendova-Franks (2008). "Can ant colonies choose a far-and-away better nest over an in-the-way poor one?" Animal Behaviour **76**: 323-334.

Franks, N. R., E. B. Mallon, H. E. Bray, M. J. Hamilton and T. C. Mischler (2003). "Strategies for choosing between alternatives with different attributes: exemplified by house-hunting ants." Animal Behaviour **65**(1): 215-223.

Franks, N. R. and T. Richardson (2006). "Teaching in tandem-running ants." Nature **439**(7073): 153-153.

Gittins, J., K. Glazebrook and R. Weber (2011). Multi-armed bandit allocation indices, John Wiley & Sons.

Gittins, J. C. (1979). "Bandit processes and dynamic allocation indexes." Journal of the Royal Statistical Society Series B-Methodological **41**(2): 148-177.
Hills, T. T., P. M. Todd, D. Lazer, A. D. Redish, I. D. Couzin and G. Cognitive Search Res (2015). "Exploration versus exploitation in space, mind, and society." Trends in Cognitive Sciences **19**(1): 46-54.
Ishii, Y. and E. Hasgeawa (2013). "The mechanism underlying the regulation of work-related behaviors in the monomorphic ant, Myrmica kotokui." Journal of Ethology **31**(1): 61-69.
Jones, J. C., M. R. Myerscough, S. Graham and B. P. Oldroyd (2004). "Honey bee nest thermoregulation: Diversity promotes stability." Science **305**(5682): 402-404.
Keasar, T., E. Rashkovich, D. Cohen and A. Shmida (2002). "Bees in two-armed bandit situations: foraging choices and possible decision mechanisms." Behavioral Ecology **13**(6): 757-765.
Kleineidam, C. and F. Roces (2000). "Carbon dioxide concentrations and nest ventilation in nests of the leaf-cutting ant Atta vollenweideri." Insectes Sociaux **47**(3): 241-248.
Krebs, J. R., A. Kacelnik and P. Taylor (1978). "Test of optimal sampling by foraging great tits." Nature **275**(5675): 27-31.
Mallon, E. B. and N. R. Franks (2000). "Ants estimate area using Buffon's needle." Proceedings of the Royal Society B: Biological Sciences **267**(1445): 765-770.
Masuda, N., T. A. O'Shea-Wheller, C. Doran and N. R. Franks (2015). "Computational model of collective nest selection by ants with heterogeneous acceptance thresholds." Royal Society open science **2**(6): 140533.
May, B. C., N. Korda, A. Lee and D. S. Leslie (2012). "Optimistic Bayesian sampling in contextual-bandit problems." Journal of Machine Learning Research **13**(Jun): 2069-2106.
McNamara, J. M., R. F. Green and O. Olsson (2006). "Bayes' theorem and its applications in animal behaviour." Oikos **112**(2): 243-251.
Noirot, C. and J. P. E. C. Darlington (2000). Termite Nests: Architecture, Regulation and Defence. Termites: Evolution, Sociality, Symbioses, Ecology. T. Abe, D. E. Bignell and M. Higashi. Dordrecht, Springer Netherlands**:** 121-139.
Ortega, P. A. and D. A. Braun (2010). "A minimum relative entropy principle for learning and acting." Journal of Artificial Intelligence Research **38**: 475-511.
Pinter-Wollman, N., R. Wollman, A. Guetz, S. Holmes and D. M. Gordon (2011). "The effect of individual variation on the structure and function of interaction networks in harvester ants." Journal of The Royal Society Interface **8**(64): 1562-1573.
Pratt, S. C. (2005). "Quorum sensing by encounter rates in the ant Temnothorax albipennis." Behavioral Ecology **16**(2): 488-496.
Pratt, S. C., E. B. Mallon, D. J. T. Sumpter and N. R. Franks (2002). "Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant Leptothorax albipennis." Behavioral Ecology and Sociobiology **52**(2): 117-127.
Reid, C. R., H. MacDonald, R. P. Mann, J. A. R. Marshall, T. Latty and S. Garnier (2016). "Decision-making without a brain: how an amoeboid organism solves the two-armed bandit." Journal of the Royal Society, Interface / the Royal Society **13**(119).
Robbins, H. (1952). "Some aspects of the sequential design of experiments", Bulletin of the American Mathematical Society, **58**(5): 527–535.
Robinson, E. J. H., O. Feinerman and N. R. Franks (2012). "Experience, corpulence and decision making in ant foraging." Journal of Experimental Biology **215**(15): 2653-2659.
Robinson, E. J. H., F. D. Smith, K. M. E. Sullivan and N. R. Franks (2009). "Do ants make direct comparisons?" Proceedings of the Royal Society B-Biological Sciences **276**(1667): 2635-2641.
Sasaki, T., B. Granovskiy, R. P. Mann, D. J. T. Sumpter and S. C. Pratt (2013). "Ant colonies outperform individuals when a sensory discrimination task is difficult but not when it is easy." Proceedings of the National Academy of Sciences of the United States of America **110**(34): 13769-13773.

Sasaki, T., S.C. Pratt, A. Kacelnik (2018). "Paralell vs. comparative evaluation of alternative options by colonies and individuals of the ant *Temnothorax rugatulus*". Scientific Reports **8**: 12730.

Scott, S. L. (2010). "A modern Bayesian look at the multi-armed bandit." Applied Stochastic Models in Business and Industry **26**(6): 639-658.

Seeley, D. T. and C. S. Buhrman (2001). "Nest-site selection in honey bees: how well do swarms implement the "best-of-N" decision rule?" Behavioral Ecology and Sociobiology **49**(5): 416-427.

Sherratt, T. N. (2011). "The optimal sampling strategy for unfamiliar prey." Evolution **65**(7): 2014-2025.

Shettleworth, S. J. and C. M. S. Plowright (1989). "Time horizons of pigeons on a two-armed bandit." Animal Behaviour **37, Part 4**: 610-623.

Speekenbrink, M. and E. Konstantinidis (2015). "Uncertainty and Exploration in a Restless Bandit Problem." Topics in Cognitive Science **7**(2): 351-367.

Strens, M. (2000). A Bayesian framework for reinforcement learning. International Conference on Machine Learning.

Thomas, G., A. Kacelnik and J. Van Der Meulen (1985). "The three-spined stickleback and the two-armed bandit." Behaviour: 227-240.

Thompson, W. R. (1933). "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples." Biometrika **25**: 285-294.

Valone, T. J. (2006). "Are animals capable of Bayesian updating? An empirical review." Oikos **112**(2): 252-259.

Whittle, P. (1980). "Multi-armed bandits and the Gittins index." Journal of the Royal Statistical Society. Series B (Methodological): 143-149.

Wyatt, J. (1997). Exploration and inference in learning from reinforcement. PhD, University of Edinburgh.

Yue, Y., J. Broder, R. Kleinberg and T. Joachims (2012). "The K-armed dueling bandits problem." Journal of Computer and System Sciences **78**(5): 1538-1556.