.

# EVOLUTION OF OBJECTIVES CAN ENABLE PROSOCIAL BEHAVIOUR WITHOUT SOCIAL AWARENESS

**Bernd Meyer**
Department of Data Science and AI
Monash University
Melbourne, Australia
bernd.meyer@monash.edu

**Yue Yang**
Department of Data Science and AI
Monash University
Melbourne, Australia
yue.yang1@monash.edu

**ABSTRACT**

Division of labour is fundamental to the functioning of societies and socially living organisms. While it has been central to their study for decades, no complete picture has emerged yet. Some of the most fascinating questions arise in the context of self-organised societies, like those of social insects, which coordinate their behaviour with completely decentralised simple decision-making performed by individuals that only have local information at their disposal. Based on empirical evidence, these collectives appear to balance task engagement globally across their whole task network for the benefit of the colony overall. How can this pro-social coordination be achieved by independently acting individuals? How is a global workforce balancing possible based on only local perception with no knowledge of the global colony status or needs? How can information flow through the task network, so that a changed task demand in one part of the network can lead to adjustments in distant other parts? We detail a model that presents a potential answer to this conundrum. Our model is informed by evolutionary game theory and rests on the assumption that the perception of an individual's sensory input can evolve. We present a conceptual argument and simulation studies to show that pro-social behaviour will evolve in a collective of agents that adjust their behaviour using primitive reinforcement mechanisms if we assume an evolving perception function.

## 1 Introduction

Division of labour is fundamental to the functioning of social organisms and has been central to their study for decades [1]. The separation of tasks among different individuals or groups within a collective allows for the efficient use of resources and increases the chances of survival for the collective as a whole [2, 3, 4]. Studies have shown that division of labour is prevalent in many socially living organisms, such as ants, bees, termites, and even some mammals [5, 6, 7, 8, 9].

Colonies of social insect are one of the most important model systems for the study of collective behaviour in general, and division of labour in particular. They are well known for their intricate colony organization and their ability to handle a wide range of tasks simultaneously, including foraging, colony defence, nest construction, temperature regulation, and caring for offspring [6, 10]. The colony's ability to effectively allocate its workforce to these different competing tasks, adapting flexibly to changes in both external conditions and internal needs, is often cited as a key to their astonishing ecological success [11, 12, 13, 14]. Understanding the underlying mechanisms that enable this flexibility is thus fundamental to understanding these social organizations and the functioning of complex social systems in general.

It is well established that both developmental and genetic factors significantly influence the division of labour in social insects [15, 16, 17]. Additionally, studies have shown that faster and self-organized mechanisms for division of labour exist within colonies, enabling them to rapidly and adaptably respond to shifts in task requirements [18, 19, 20]. We adopt the term *task allocation* for this fast process to distinguish it from the more general phenomenon of division of

labour.[1] It is well established that the process of task allocation generally responds to altered environmental conditions such as food abundance, predation threats, weather and climate conditions [21, 22]. Empirical research has also emphasized the significance of social context and interactions in shaping the task preferences of individuals [23, 24, 25]. The fast dynamics of task allocation is thus driven by the interplay between workforce distribution, interactions structure, and environmental factors [26, 27, 28, 29, 30].

As this collective plasticity is ultimately based on flexible individual behaviour, understanding the rules that govern the task choice behaviour on the individual level is core to deciphering this collective behaviour.

Much of the research on task choice behaviour has focussed on the performance of a single task or a binary choice between a pair of tasks. However, the reality is far more complex: the colony is faced with a complex network of tasks, spanning foraging, brood care, defence, nest maintenance, thermoregulation, and many others. Fitness pressure dictates that the task execution has to be balanced globally across the whole task network for the benefit of the whole colony rather than just locally for the immediate benefit of a particular task as perceived from the perspective of an individual. Apart from the obvious evolutionary fitness argument for network-wide or colony-wide optimisation, empirical evidence of such network-wide balancing in real colonies exists [31] and has been known for almost 50 years [32, 33, 34] (summarised in [35]).

Arguably, the most puzzling aspects of global task allocation is that it must be managed by individuals that generally lack knowledge of the overall state of the colony,[2] so that their behavioural decisions must rely on only local information that can be readily sensed by them [37, 38] and potentially additional information obtained in interactions with other colony members [39, 40].

Fascinating questions about this global balancing arise from two different perspectives. The first perspective, which drives our work, concerns the proximate behavioural mechanisms: how is such global balancing even possible based on only local perception? How can information propagate through a task network, so that a changed task demand in one part of the task network can lead to adjustments in distant other parts? The second perspective it that of evolutionary theory, which asks how altruistic behaviour of individuals arises in the first place.

It turns out that the answers to both questions are tightly linked: they lie in the evolution of perception. What is missing in the overall picture is an explicit link between individual behaviour and integrative colony fitness which is not visible at the level of the individual. An evolving "perception function", which shapes how directly perceived sensory inputs are used in the process of behaviour adaptation, can establish this link. The single assumption of an evolving perception function is sufficient to let pro-social behaviour emerge within a collective.

To study this, we consider an abstract task allocation scenario [18, 41, 42] in which a group of agents has to balance global task engagement across a network of tasks. Individuals can migrate between tasks along the edges of this network. At any point of time, an individual is only concerned with a binary choice between two tasks and can only sense information about this pair of tasks.

Such binary choices are a paradigmatic study scenario, both in empirical and in modelling work. The core ingredient in the most widely accepted models of such binary task choice in social insects is stimulus intensity as formalised in the response threshold models [29, 43, 44] (for reviews see [45, 46, 37]). The general idea of response thresholds has been used in many different forms. It simply refers to the assumption that individuals have an internal threshold for task-related stimuli and that they will react to a task demand only if its stimulus exceeds this threshold. The more sophisticated probabilistic reinforced response threshold model (RRTM) [44, 47] very sensibly dispenses with the idea of a hard threshold and uses it only as a parameter that regulates the probability of responding to a task stimulus. This is one of the few models that have been fully mathematically formalized and arguably the most widely accepted one among these.

We use the same principle in a different mathematical formulation called cross learning [48]. Cross learning is essentially a continuous-valued version of one of the earliest mathematical models for behaviour adaptation in primitive animals, the so-called Tsetlin automaton [49]. Like the RRTM, cross learning is a very simple form of reinforcement learning.

In our model, each agent in the collective independently decides which tasks to execute and adapts their choice behaviour via cross learning. In the binary choice scenario, this is virtually the same as the classical model as cross learning can be shown to exhibit the same dynamics as the RRTM in the limit of constant stimuli [50]. However, expressing this as cross learning allows us to analyse the system with the much more powerful mathematical framework of evolutionary game theory [51]. This is because it is well known that cross learning exhibits the same dynamics as replicator dynamics [52], one of the core methods of evolutionary game theory.

[1] More correctly, it should be termed "task selection", since tasks are selected in a self-organised manner by the individuals performing them rather than being "allocated." However, "task allocation" is the established term.

[2] With a few exceptions, such as honeypots [36]

We first investigate a binary choice scenario and show how the agents learn to solve simple allocation problems using only locally perceived information and cross learning. However, they inevitably fail to act pro-socially (in the interest of the collective) with this behavioural mechanism in more difficult scenarios. We then introduce an evolving "perception function", modelled as an evolvable neural network, that shapes how the locally perceivable and directly sensed information is used as input to the cross learning process. Since this effectively amounts to the reshaping of the objective function of an optimisation problem (see Section 2), we term this process "evolution of objectives" or "objective evolution" for short.

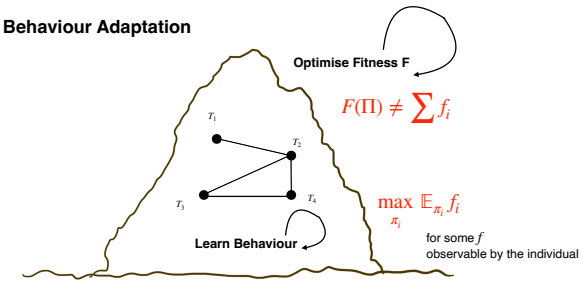


Figure 1: In the classical model, behaviour adaption amounts to individuals learning behavioural policies $\pi_i$ that optimise the expectation of their individual task reward (or performance feedback) $f_i$, for example, the amount or quality of food retrieved in foraging. However, there is no channel for information flow from the overall colony fitness $F$, which depends on the joint behaviour $\Pi$ of all agents, to the individual feedback. If $F$ is not simply additive across all tasks and individuals, this can lead to behaviour policies that are not pro-social.

For the binary choice scenario, it turns out that this simple and plausible addition enables the agents to learn pro-social behaviour, because it establishes a link between colony-level fitness pressures and individual behaviour. Next, we show that the ability to "correctly solve" binary task choices also resolves the network-wide balancing problem as it enables task-demand information to propagate unhindered through the network. Effectively, our model shows that we do not have to assume complex forms of information exchange between individuals, since global regulation can already be achieved on the basis of propagating task demands.[3] Finally, we show the applicability of our model to real biological scenarios in the context of the transport-optimising behaviour of foraging Army ants.

It is important to note that we do not claim to have biological evidence that this is the actual process taking place in social insect colonies. However, this is a parsimonious, consistent, and, we claim, plausible model of how global pro-social task balancing in a network of tasks can take place, a process for which currently no general model exists. We thus believe that it is a hypothesis worthy of further investigation. The point of the present paper is to detail this hypothesis and show through computational studies and mathematical analysis that it is consistent.

## 2 Model

Our model captures how task allocation can be regulated to achieve pro-social outcomes even if rational individuals independently adjust their task choices using only locally perceived information. We construct this model in two steps: In the first step, the model addresses a single binary choice between two tasks, while the complete model addresses a network of such binary choices, so that the whole task network of the collective can be captured.

### 2.1 Base model for binary choice

#### 2.1.1 Individual agent behaviour

We consider a set of $N_a$ agents $A = \{A_1, \ldots, A_{N_A}\}$ and a set of $N_t$ tasks $T = \{T_1, \ldots, T_{N_t}\}$ that these agents have to perform. Each agent $A_i$ decides independently in which task to engage based on a set of $N_p$ agent-specific properties $X_i \in \mathbb{R}^{N_p}$ ("continuous-valued traits" in evolutionary game theory terminology) and a behaviour policy $\pi : \mathbb{R}^{N_p} \to \mathbb{R}^{N_t}$ that maps its trait values to the probability of engaging in any of the tasks. For simplicity, we assume in the following that each agent has a single trait per task that directly models its response probability for this task, i.e. $N_p = N_t$ and $\pi(X_i) = X_i$. The joint profile of all agents' traits is $X = (X_1, \ldots, X_{N_A})$. We write $[\cdot]_j$ to select the $j$-th component of a vector, i.e. $[X]_j = X_j$ is the (possibly multivariate) trait of agent $A_j$.

Agents individually perceive task-specific performance feedback (or rewards), which could, for example, be the quality and/or the amount of food retrieved during foraging. We allow for the possibility of this feedback to be dependent on the task choices of other agents, and specifically to be frequency-dependent. Let $C = T^{N_a}$ be the joint action space (space of task choices for all agents) and $Z \in C$ be the joint action profile for all agents, ie. $[Z]_j$ is the task that agent $A_j$ engages in. The task feedback function $f : C \to \mathbb{R}^{N_t}$ with $f(Z) = (f_1, \ldots, f_{N_t})$ gives the feedback for all tasks,

[3] Of course, this is not saying that information exchange between individuals could not be used to improve self-organised task allocation.

which are specific for each task but the same for all agents engaging in the same task. Ultimately, we are interested in the social benefit (the benefit for the whole collective) that we take as a fitness proxy. This is defined by $F : C \to \mathbb{R}$.

Central to our model is the fact that individual agents cannot directly perceive the social benefit $F(Z)$. Individuals can only observe $N_S$ directly perceived sensory inputs. We denote the sensory input of agent $A_i$ as $S_i = (s_{i,1}, \ldots, s_{i,N_S})$ with $s_{i,j} \in \mathbb{R}$. Depending on the scenario modelled, the immediate task feedback, specifically for the task $[Z]_i$ that the agent is currently engaged in, would typically be included in $S_i$.

Crucially, we assume that these sensory inputs are not directly used to control behaviour adaptation but instead that they are first processed by a *perception function* $N_\theta : \mathbb{R}^{N_S} \to \mathbb{R}$ that we implement as a neural network with parameters $\theta$. $N$ is thus an universal function approximator.[4] In our model this function approximator evolves (see Sec. 2.1.2).

*It will turn out that, provided $S$ is sufficiently informative, the ability of $\theta$ to evolve is sufficient for the collective of agents to evolve pro-social behaviour.*

As a shorthand, let $\hat{S}_i = N_\theta(S_i) \in [0, 1]$ be the perception of agent $A_i$, ie. its sensory inputs processed through the perception function and normalised in the unit interval.

Agents continuously adjust their task choices based on their perception with a simple form of reinforcement learning, which is well established as a mechanism of behaviour adaptation in simple animals, including insects [53]. Our model progresses in discreet time steps $t_1, \ldots, t_n$. At each step each agent $A_i$ (1) chooses a task to engage in according to $\pi(X_i)$, (2) evaluates its perception $\hat{S}_i$ (including its task performance feedback), and (3) updates its traits $X_i$ accordingly.

Assuming that $A_i$ executes task $T_j$ in a given time step $t$ it updates its traits as

$$[X_i]_k \leftarrow \begin{cases} \hat{S}_i + (1 - \hat{S}_i)[X_i]_k & \text{for} \quad k = j \\ (1 - \hat{S}_i)[X_i]_k & \text{for all} \quad k \neq j \end{cases} \tag{1}$$

This update schema is known as cross-learning [48], and, in the limit of constant stimuli, exhibits the same dynamics as the reinforced response threshold model, arguably the best established model of behavioural adaptation in social insects [44].

### 2.1.2 Colony evolution

The model so far captures the fast behaviour adaptation by individuals on short time scales during their lifetime. Crucially, the parameters of the perception function can evolve. The dynamics of the system thus plays out on two different timescales. On the fast time scale (behaviour of the collective), we assume $\theta$ as fixed and let the agents' traits equilibrate resulting in the steady-state trait profile $X^\star$ and the associated joint behaviour policy $\Pi^\star = \Pi(X^\star) = (\pi(X_1^\star), \ldots, \pi(X_{N_A}^\star))$. On the slower timescale (evolution of the collective), we assume $\mathbb{E}_{\Pi^\star} F(Z)$ as the fitness at steady-state and let $\theta$ evolve accordingly through mutation and crossover using differential evolution [54].

### 2.1.3 Environment (binary choice)

Cross learning is well know to exhibit the same dynamics as replicator dynamics in expectation and in the limit of small time steps [48], so that it will converge on a Nash equilibrium. The task of pro-social reward optimisation is thus trivial if individual rationality leads to the pro-social optimum. We are specifically interested in cases where the benefit expectation in equilibrium under individual rational choice is less than the pro-social optimum, i.e.

$$\mathbb{E}_{\Pi^\star} F(Z) < \max_Z F(Z) \tag{2}$$

Congestion games, a well-established model of $n$-player social dilemmas [55] in which Nash equilibria are not pro-social, are a suitable model of this. We thus define our environment first as a single congestion game and later as a network of congestion games (see Section 2.2). A congestion game generally models a situation in which players make use of a set of capacity-limited resources. Each action that a player can choose requires the use of one or more resources and there is a penalty for resource use. This penalty increases with the number of players utilising the same resource. In other words, the payoff for a player is inverse in the number of players utilising the same resource(s). This captures, for example, some foraging scenarios where overcrowding a food source can reduce the foraging efficiency of individuals [56, 57].[5] etc. More generally and more importantly, we can also take this as an abstract model of a task

[4] Any other evolvable representation of a universal function approximator would be possible but a neural network seems both a technically sensible and a biologically more relevant choice.

[5] Note, however, that crowding does no universally decrease foraging efficiency [58]

choice in which two tasks have to be balanced rather than individually optimised, for example different food sources due to nutritional geometry [59].

In our setting, each individual congestion game represents a binary choice between two tasks (for example foraging at one of two sources, $N_T = 2$). We call this an *atomic* game. The individual reward for each task is frequency-dependent and monotonically decreasing in the number of agents engaging in this task. Despite of this, the collective reward for all agents in a single task is increasing in the number of agents. The specific choice of the reward function is not crucial. For the purpose of our experiments and motivated by the fact that the logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$ is a suitable fundamental choice for a capacity limited resources, such as the yield of a food patch [60], we define

$$f_k(\beta_k) = M_k - \frac{\sigma(10\beta - b) - \sigma(-b)}{1 - \sigma(-b)} \tag{3}$$

where $\beta_k$ is the proportion of the agent population engaging in task $T_k$, $M_k$ is the maximum individual payoff that can be obtained from task $T_k$ and $b$ is a parameter determining the inflection point of $f$. With this, the payoff for the entire group of agents becomes

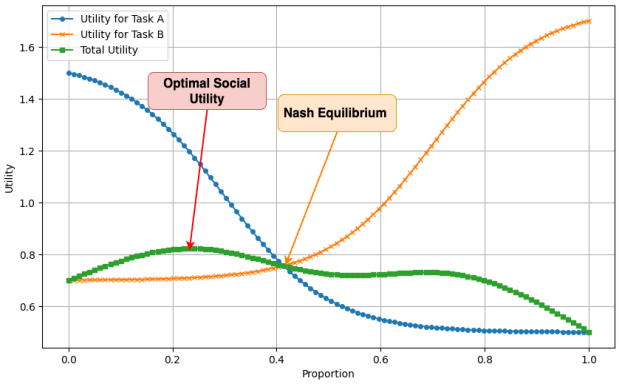


Figure 2: Individual and collective payoff for a binary task choice: The maximum of the collective payoff (green curve) does not coincide with the Nash equilibrium (intersection of the individual payoffs). Thus, the Nash equilibrium is not pro-social.

$$F(Z) = \sum_j \beta_j f_j(\beta_j) \tag{4}$$

where $\beta_j$ is the population fraction engaging in task $T_j$ under the joint task profile $Z$. See Figure 2 for the relationship between these functions: the Nash equilibrium (on which cross-learning converges) is generally not the social optimum. Note that we have restricted atomic games to binary choices only for simplicity and to save notation. The extension to $N_T > 2$ is straight forward.

### 2.2 Model for task networks

In the more general case, we consider a network of tasks that induces a Markov game [55].

We assume there is a reachability relation between tasks (i.e. agents can migrate between some tasks but not others) modelled as an undirected graph $\mathcal{G} = (T, E)$ in which tasks are the vertices and edges connect the tasks between which the agents can migrate. The entire set of tasks $T$ is partitioned into $N_G$ disjoint subsets, each of which constitutes one atomic game $G_i$ as defined in Section 2.1: $G = \bigcup_i G_i$.

At each point in time an agent participates in one particular game $G_i$. Agents can migrate through the task network by moving between games according to the reachability graph $\mathcal{G}$. An agent can migrate from $G_i$ to any game $G_j$ for which $\exists (T_k, T_l) \in E : T_k \in G_i \wedge T_l \in G_j$. We assume that $\mathcal{G}$ is connected, so that agents can ultimately reach any task in the network.

$\mathcal{G}$ induces a Markov game $M = (G, A, C, P, R)$, in which $G$ is the set of states (atomic games), $A$ the set of agents, $C$ the joint action space (restricted in such a form that only tasks in $G_j$ are accessible to an agent in $G_j$), $P : G \times T \times G \to [0, 1]$ the transition kernel, and $R = (r_1, \ldots, r_{N_A})$ with $r_i : G \times C \to \mathbb{R}$ the payoff function for all agents. Note that the joint action profile enters into $R$ as individual game payoffs are frequency dependent. $\mathcal{G}$ induces the following structure for $P$:

$$P(G_i, T_j, G_k) = \begin{cases} p_{ik} > 0 & \text{if } \exists (T_m, T_n) \in E : T_m \in G_i \wedge T_n \in G_k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

We initially assume that the payoffs of all atomic games are additive, i.e. the one-step payoff of the entire network game is the sum of the payoff of all games in the network.[6]

[6] Recall that this is very different from the payoff of all tasks being additive, which would render the Nash equilibrium pro-social.

However, other forms of combining individual game payoffs are possible. Of particular biological relevance is a payoff combination in which some tasks are limiting the overall benefit. For example, we may want to capture the situation that no task can ultimately generate a fitness benefit if brood care is entirely neglected. Such situations can be modelled by a *minimum* payoff instead of an additive payoff, which leads to a *maximin* optimisation problem instead of a standard maximisation problem. This can be handled in similar forms. We leave the details for future work.

Finally, we assume that none of the games in $G$ is redundant, i.e. the population fraction allocated to each game in the Nash equilibrium of $M$ lies in the interior of the simplex $\Delta N_G$.

## 3 Simulation Results

We first simulate binary choice between two congestive tasks in an atomic game as described in Section 2.1.3 and then proceed to simulate complete task networks.

### 3.1 Binary choices (atomic games)

For binary choice scenarios, we use the binary congestion problem ($N_T = 2$) defined in Section 2.1.3 with ten randomly randomly drawn parameter sets. These parameter sets are given in Table 1. We use multiple problems to show that a single unified perception function can be evolved that applies to different settings rather than just one particular scenario.[7] We then apply the evolutionary procedure defined in Section 2.1.2 to colonies in which individuals behave as defined in Section 2.1.1.

| Task | Benefit ($B_A$, $B_B$) | Cost for A ($C_A$) | Cost for B ($C_B$) |
|---|---|---|---|
| 1 | 2 | 0.3 | 0.5 |
| 2 | 2 | 0.5 | 0.3 |
| 3 | 2 | 0.6 | 0.2 |
| 4 | 2 | 0.4 | 0.2 |
| 5 | 2 | 0.8 | 0.6 |
| 6 | 2 | 0.1 | 0.3 |
| 7 | 2 | 0.9 | 0.7 |
| 8 | 2 | 0.5 | 0.6 |
| 9 | 2 | 0.7 | 0.9 |
| 10 | 2 | 0.2 | 0.6 |

Table 1: Parameter list for benefits and costs for A and B

We evolve a population of perception functions, the fitness of which is defined as the difference between the maximally achievable collective payoff and the collective payoff achieved in the steady state by a population of cross learners based on this perception function.

Note that the cross learning procedure is only used to find the steady-state population distribution and that the actual payoffs entering into the fitness calculation are derived from the true collective payoff as follows: Let $\beta^\star = \operatorname{argmax}_\beta F(\beta)$ be the population fraction for the first task that maximises the collective fitness. Let $x^\star_\theta$ be the steady-state proportion of the population engaging in the first task as it results from cross-learning according to the perception function $N_\theta$. The fitness $H$ of this perception function is then defined as:[8]

$$H(N_\theta) = F(x^\star_\theta) - F(\beta^\star) \tag{6}$$

The resulting perception functions for the parameters in Table 1 are compared with the true objectives in Figure 3. We observe that in all cases, a (almost) unimodal perception function evolves with an *argmax* that is is very close to the *argmax* of the true objective, inline with our hypothesis.

Note that these simulation should really only be understood as a sanity check. It is easy to see abstractly that the outcomes cannot be different. There is selective pressure on the global *argmax* of the perception function to be the same as that of the true objective. This is because the true collective payoff is optimal under this condition, implying

[7] We acknowledge that this is a relatively small spread of problems. The reason is that the simulation involves the artificial evolution of a large population of agent-based simulations, which is computationally a very costly process.

[8] Note that we have trivially overloaded the fitness function $F$ at this point, since it was originally defined on an action profile. However, as can easily be seen from Eq. 4, the function essentially works by computing the population fractions from the joint action profile, so the meaning of this should be straightforward.

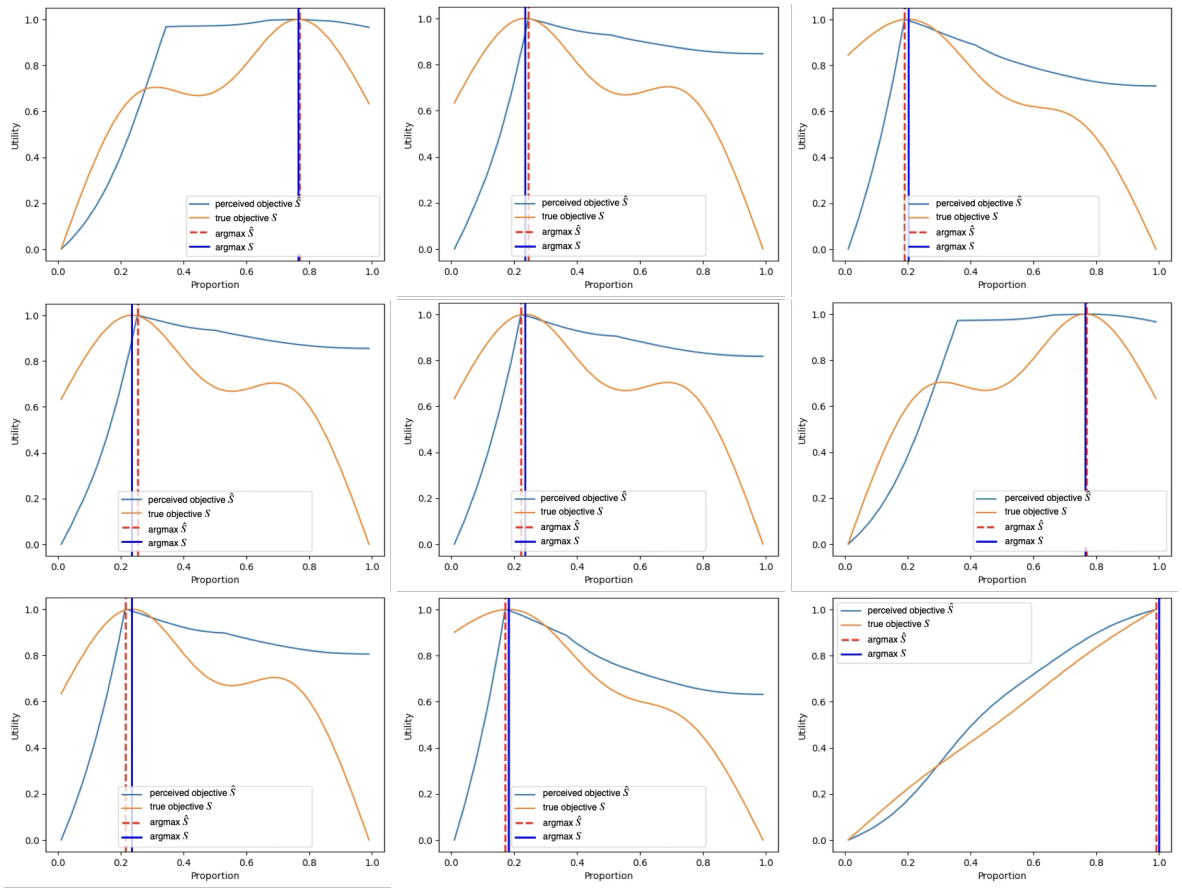


Figure 3: Comparison between the true collective objective and the perceived objective for different objective parameter settings: The perceived objective (blue line) generally evolves to be a unimodal function with the global optimum at a point that is close to the global optimum of the true objective (orange line).

maximum fitness. Furthermore, there is selective pressure on the perception function being unimodal as the fitness is highest if the *argmax* is reached by cross learning from every possible starting point. Cross learning ultimately performs a stochastic gradient ascent, so a unimodal perception function is required to be able to reach the optimum from any starting point. Thus, a unimodal perception function that has its optimum at the same point as the true objective has the optimal fitness and should normally evolve. The values of the perception function will normally be different from the true objective. There is no selective pressure on the absolute values, since only the derivative is relevant for the direction of the dynamics.[9] Of course, there is no guarantee that an evolutionary process will find the global optimum. In our case, we might expect this to fail if the true objective has multiple local optima that are close in their objective values. In such cases, the stochastic nature of the perception function evolution might develop towards a local optimum.

It is important to contrast these outcomes with those that would result if we eliminated the perception function and let the agents perform cross learning using the immediate task reward $[f(Z)]_j$ instead to adapt their behaviour when performing task $T_j$. We know that the steady state would be a Nash equilibrium. Independently of the specific parameters, all our reward functions have the shape shown in Figure 2. The behaviour adaptation would thus equilibrate where the payoff for both tasks is equal $[f(Z)]_1 = [f(Z)]_2$. As is evident in the figure, this is generally not the point of optimal collective payoff.

In summary, the collective of agents has evolved a perception function that, independently of the particular parameters of the environment, allows them to optimise the collective reward (and fitness) for the colony without requiring explicit coordination or even explicit awareness of the social context. This optimum would not be reached if the agents exhibited the same behaviour but used the immediately sensed the reward of the task executed.

[9]Note that of ease of handling and presentation, we have rescaled the perception functions such that the global maximum of the perception function and the true objective are the same.

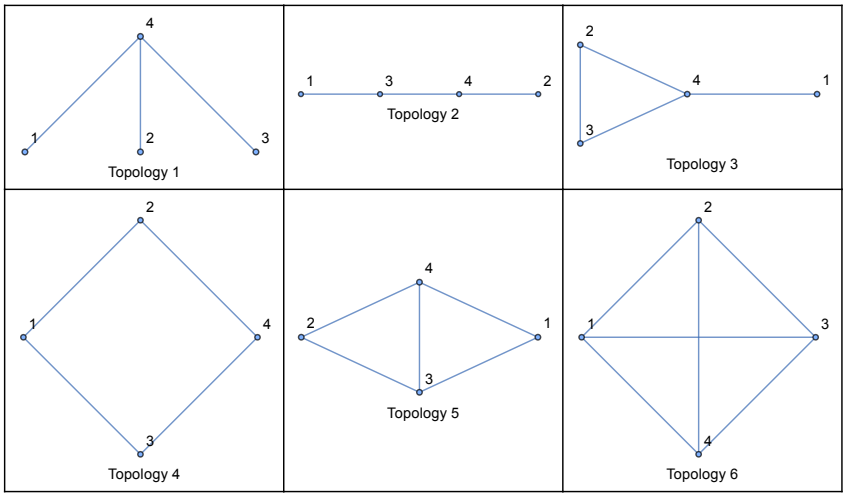


Figure 4: Game network topologies. Vertex labels correspond to the indexes of atomic games in the example runs detailed in Figs. 7, 8.

Thus, simply introducing an evolvable perception function allows the collective to evolve pro-social behaviour without requiring any additional social coordination mechanisms beyond those already used for the control of solitary behaviour.

### 3.2 Task networks

We now show that we can extend the approach from interactions that are limited to a single (fully connected) task group to a network of tasks as defined in Section 2.2.

Our simulation network comprises $N_T = 8$ tasks in $N_G = 4$ atomic games. This is because four atomic games is the smallest configuration in which an investigation can be performed across different topologies. We investigated 600 scenarios, corresponding to 100 random parameter sets for each of the possible six topologies (see Fig. 4). We use $\alpha e^{-\beta x}$ with $0 < \alpha, \beta \leq 1$ as the congestion function. All simulations run for 25,000 update steps. We established a performance baseline by determining the network payoff for 10,000 random parameter sets for each configuration.

Optimality in the table and all figures is measured as the fraction achieved of the theoretically possible payoff improvement over the minimal payoff for these parameters:

$$o = \frac{F(X^\star) - min_X\, F(X)}{max_X\, F(X) - min_X\, F(X)} \tag{7}$$

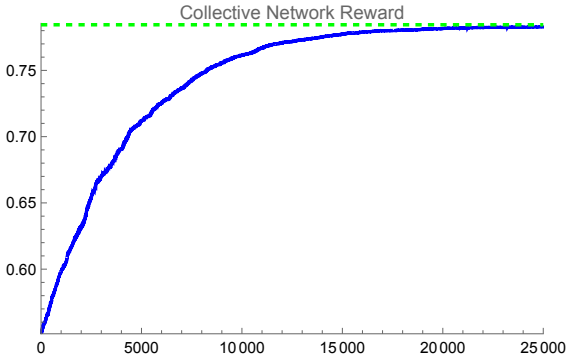


Figure 5: Network payoff over 25,000 update steps. The vertical scale is the absolute network payoff (not the optimality level) and the green line indicates the theoretically achievable maximal payoff.

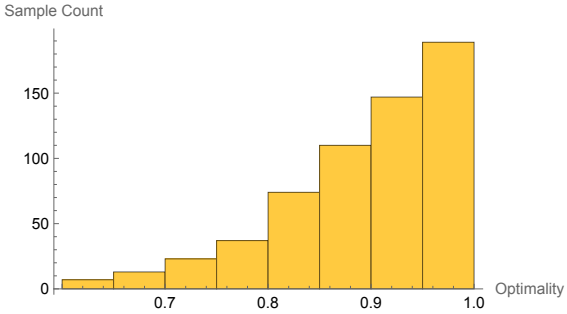


Figure 6: Simulation outcomes for 600 runs with 25,000 update steps each across all 6 topologies.

## Objective Evolution

All runs exhibit the same network payoff dynamics that is increasing over time and concave but they converge at different levels of optimality (Figure 5). Obviously, not every simulation reaches the optimum in the fully stochastic setting. However, the final payoff is significantly above the baseline in the vast majority of cases (Table 2) and we rarely see a process stuck at low payoffs (Figure 6). No significant differences between different topologies can be observed.

We give examples of two typical runs for each topology, one converging close to the optimal distribution (Figure 7) and one converging to a different distribution (Figure 8).

| Topology | Optimality Mean (n=100) | Optimality Std Dev | Baseline Mean (n=10000) | Baseline Std Dev |
|---|---|---|---|---|
| 1 | 0.875 | 0.093 | 0.599 | 0.118 |
| 2 | 0.897 | 0.086 | 0.598 | 0.118 |
| 3 | 0.902 | 0.082 | 0.600 | 0.118 |
| 4 | 0.887 | 0.090 | 0.599 | 0.118 |
| 5 | 0.909 | 0.078 | 0.599 | 0.119 |
| 6 | 0.901 | 0.078 | 0.598 | 0.118 |
| **All** | **0.895** | **0.085** | **0.599** | **0.118** |

Table 2: Comparison of optimality to baseline across all topologies

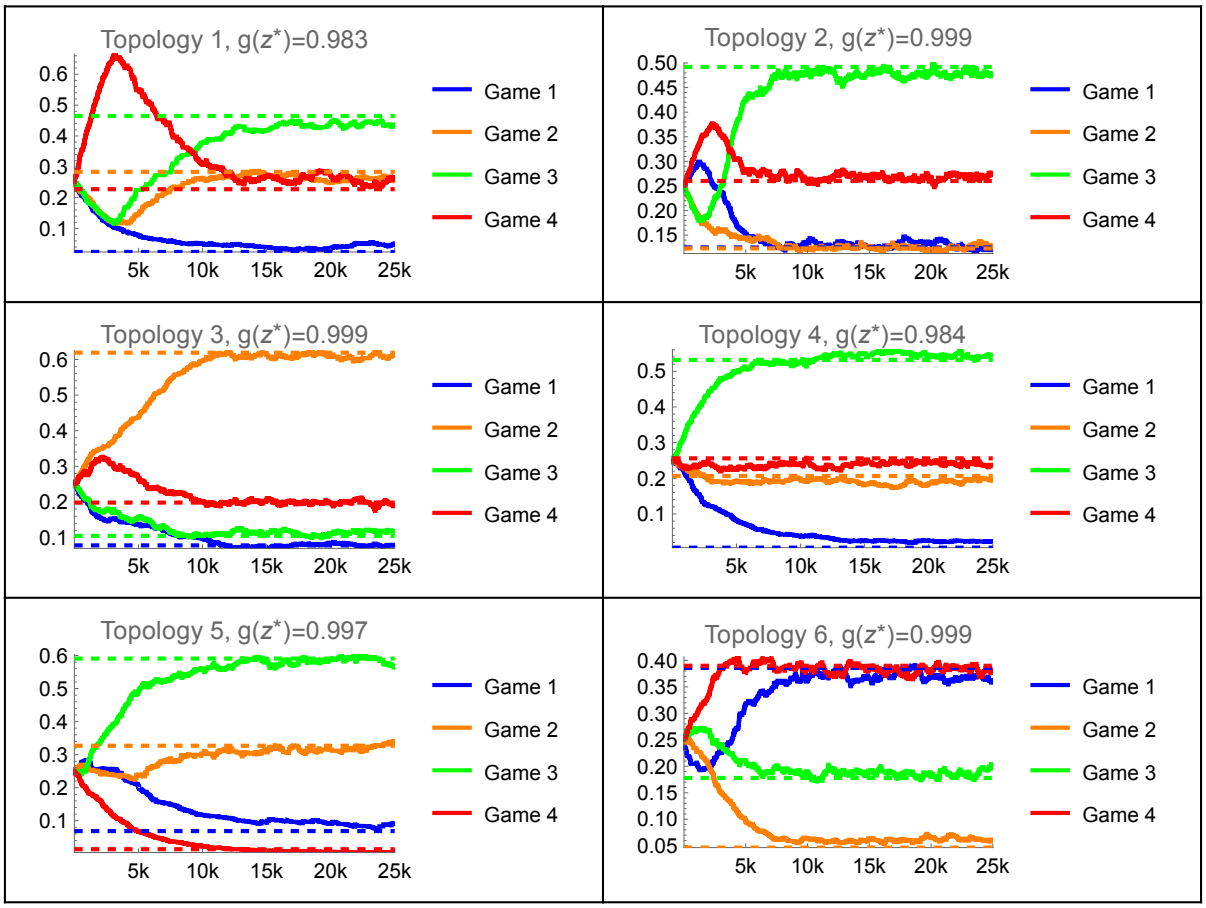


Figure 7: Distribution of the agent population across games over time for sample runs that converge (close) to the optimal distribution. Solid lines indicate the fraction of the population in the respective game, dashed lines indicate the fraction at the theoretical optimum of the problem for the given parameters. Game indices correspond to the vertex labels in Fig. 4.

Upon closer examination, at least one of the following two causes seems to be involved in almost all cases where the network converges to a distribution that is significantly different from the optimal one:

- *Node crash*: When the occupancy at a node becomes very low, the improvement in objective value that agents moving from this node to neighbouring nodes can achieve may not be large enough to overcome the random fluctuations of the stochastic process on the timescale of the simulation. This can be particularly pronounced at articulation points that can disconnect the graph (Examples: Topology 1–3).
- *Lack of pressure*: When the potential landscape is relatively flat, the gradient may not be large enough to overcome the stochastic fluctuations on the timescale of the simulation. In these cases the objective value is often still close to the optimum but the distribution of agents across games is different. Examples: Topologies 1, 2, 4–6.

### 3.3 Convergence in Task Networks

Any convergence argument must rest on the fact that the introduction of the perception function reshapes all atomic games into potential games. This also implies that the whole network game has been reshaped into a state-based potential game [61], a simpler form of a Markov potential game [62].

Convergence in Markov games is generally a very challenging area of current research in multi-agent learning. Here, we sketch how a convergence proof for our model can build upon the existing convergence proof for state-based potential games.

To show that our Markov game $M$ is a state-based potential game, we need to ensure that (Def. 3.2 [61])

(i) For any agent $i$, joint action $a$ and action $a_i'$: $U_i(a_i', a_{-i}, x) - U_i(a, x) = \Phi(a_i', a_{-i}, x) - \Phi(a, x)$

(ii) For any action $a$, state $x$ and state $x'$ reachable via from $x$ via $a$: $\Phi(a, x') \geq \Phi(a, x)$

The first condition is the usual condition of potential games adapted for state-based games: it demands the existence of a global potential function $\Phi$ such that the change in utility for an individual agent that unilaterally changes its action is exactly reflected in the global potential function $\Phi$.[10] The second condition is specific to state-based potential games and demands that any action-invariant trajectory must proceed along strictly non-decreasing paths in the global potential $\Phi$.

The first condition is trivially fulfilled if we take $\phi_i$ as the utility and define the global potential as

$$\Phi(X) = \sum_{i=1}^{N_G} \phi_i(x_i) \tag{8}$$

The second condition can be satisfied with a slight modification of the game's action set and transition kernel. We designate a specific *move* action for leaving an atomic game. The agent now remains in the same game until they explicitly decide to leave. This makes the modified transition kernel

$$P(G_i, A, G_k) = \begin{cases} p_{ik} > 0 & \text{if } \exists (T_m, T_n) \in E : T_m \in G_i \wedge T_n \in G_k \wedge A = move \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

We define the utility of moving as constant, regardless of state. Since only the leave action has a non-empty set of reachable states, this satisfies condition (ii).

$M$ is thus a state-based potential game.

Theorem 5.1 of [61] then guarantees that the game will converge to an action profile or, more precisely, a limiting distribution of action profiles $A^\star$ that maximise the global potential, $A^\star \in \operatorname{argmax}_A \Phi(A)$, under the assumption that the agents perform state-based log-linear learning and Theorem 6.1 of [63] extends this to payoff-based log-linear learning. By construction of our potential functions, this maximises the true objective.

Payoff-based log-linear learning and cross learning are both forms of stochastic hill-climbing that only assess the current action payoff with some subtle differences in their dynamics. Convergence in potential games tends to be quite robust against changes in the finer details of the behaviour revision protocol in general [64] and for log-linear learning in particular [63]. In combination with the simulation results, this gives us some confidence that these results carry over.[11] However, a formal proof still remains to be completed.

[10] Note that this can be relaxed to *ordinal* state-based potential games in which the requirement only is that $U_i(a_i', a_{-i}, x) - U(a, x) > 0 \Rightarrow \Phi(a_i', a_{-i}, x) - \Phi(a, x) > 0$.

[11] At least for potential functions with a unique local optimum as cross-learning may otherwise converge to a local optimum.

While such a formal proof would be valuable and interesting, it is questionable whether it is of cardinal importance for our main argument. A proof will necessarily always be bound to a precisely defined behaviour revision protocol. We will, however, never know the intricate details of the behaviour revision in any natural system or society in full. The aforementioned robustness of (possibly approximate) results thus is of much greater importance and at the current state of multi-agent research this is arguably best established via simulations.

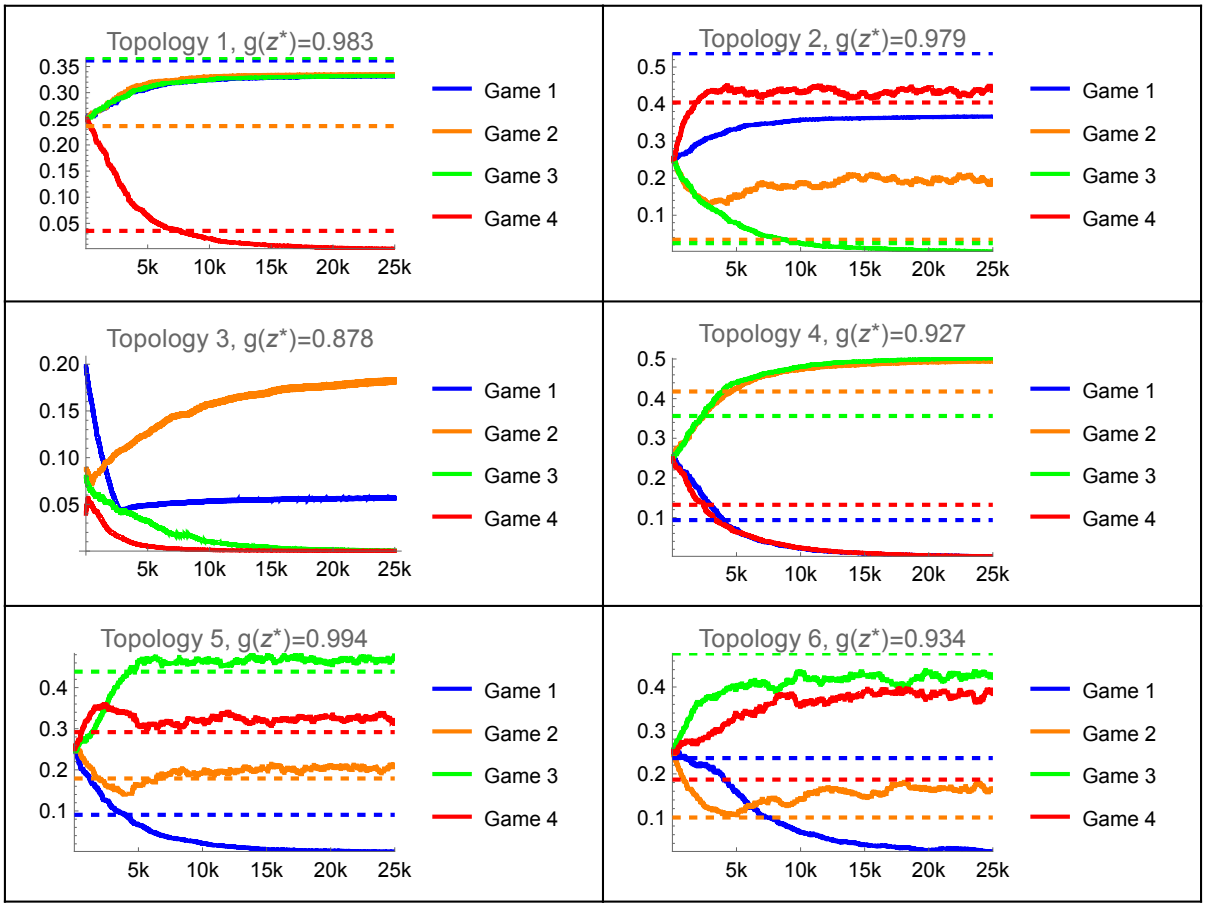


Figure 8: Distribution of the agent population across games over time for sample runs that converge to a distribution different from the optimal distribution. Solid lines indicate the fraction of the population in the respective game, dashed lines indicate the fraction at the theoretical optimum of the problem for the given parameters. Game indices correspond to the vertex labels in Fig. 4.

Figure 9: A "Living Bridge" formed by *Eciton* (from [65])

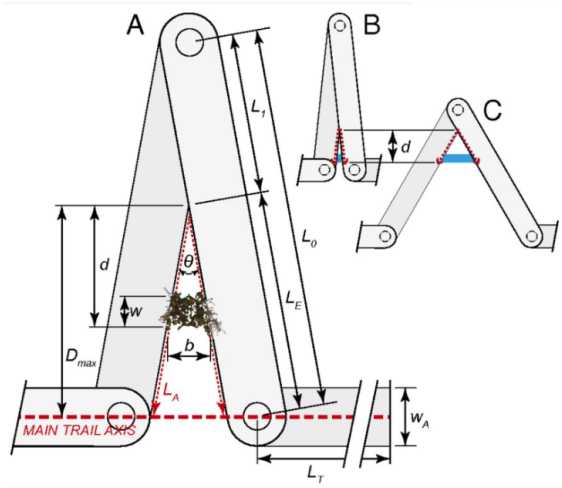


Figure 10: Experimental apparatus (from [65])

## 4 Application to *Eciton* bridge building

To demonstrate that our model, beyond being conceptually plausible, is consistent with real biological scenarios we apply it to the case of Army ant (*Eciton*) bridge building. Army ants form assemblages of their own bodies for a variety of purposes, including to form "bridges" that span gaps in the colony's foraging trail (Fig. 9). Based on data from field experiments, Reid at al. [65] presented a mathematical model of this behaviour revealing that it is consistent with the assumption that these bridges are built to optimise foraging traffic and transport. The experiments inserted an adjustable angular device (Fig. 10) into natural foraging trails, prompting the animals to bridge across this forced "detour". Initially, a very small bridge forms at the point where the two arms of the apparatus meet. Subsequently, increasingly many ants join the bridge, allowing it to lengthen and move further down the arms of the apparatus. However, at some point this process stops and the bridge stays approximately in place.

The work by Reid at al. explains that the final position of the bridge depends on the angle of the apparatus and is optimal in terms of balancing the benefit of a shorter trail with the cost of sequestering ants from the foraging process to bridge building. While their work elegantly explains why these bridges are built and how they should be positioned, [65] does not offer any mechanistic explanation of how individual workers can control their behaviour to achieve this collective benefit. Indeed, this is not easy to see, since no worker can directly sense the quantities to be optimised. Our model provides an explanation of how this could happen.

| Symbol | Definition |
|---|---|
| N | Total number of ants available on the foraging trail. |
| $L_T$ | Length of the main trail (not including the experimental gap). |
| $L_E$ | The length of the inner edge of one arm. |
| $L_1$ | The Length of apparatus arm that overlaps with the other arm. |
| $L_0$ | The length of the arm between the two hinge points. |
| $L_A$ | Effective inner length of the apparatus (the "gap"), which depends on the angle $\theta$. |
| $d$ | Distance the living bridge has moved from the junction of the bridge toward the main trail axis. |
| $D_{\max}$ | Maximum possible distance $d$ can reach (i.e., the full shortcut). |
| $b = 2\,d\,\tan\!\left(\frac{\theta}{2}\right)$ | The physical length of the bridge once it has extended a distance $d$. |
| $w_\theta$ | Empirically measured ratio of bridge width to bridge length (depends on $\theta$). |
| $\ell_n,\, w_n$ | Length and width of an individual ant (used to estimate how many ants form the bridge). |

Table 3: Key measurements

## Objective Evolution

The hypothesis of [65] is that the ants optimise an (empirical) traffic density that is estimated as $\rho = \frac{N - \frac{n_b}{\alpha}}{f}$ where $N$ is the total number of foragers, $n_b$ the number of foragers sequestered into a bridge, and $f$ the total travel distance.[12] $\alpha$ is a free parameter that varies between experiments and is supposed to capture various influences on the global density as well as other factors, such as differing body sizes.

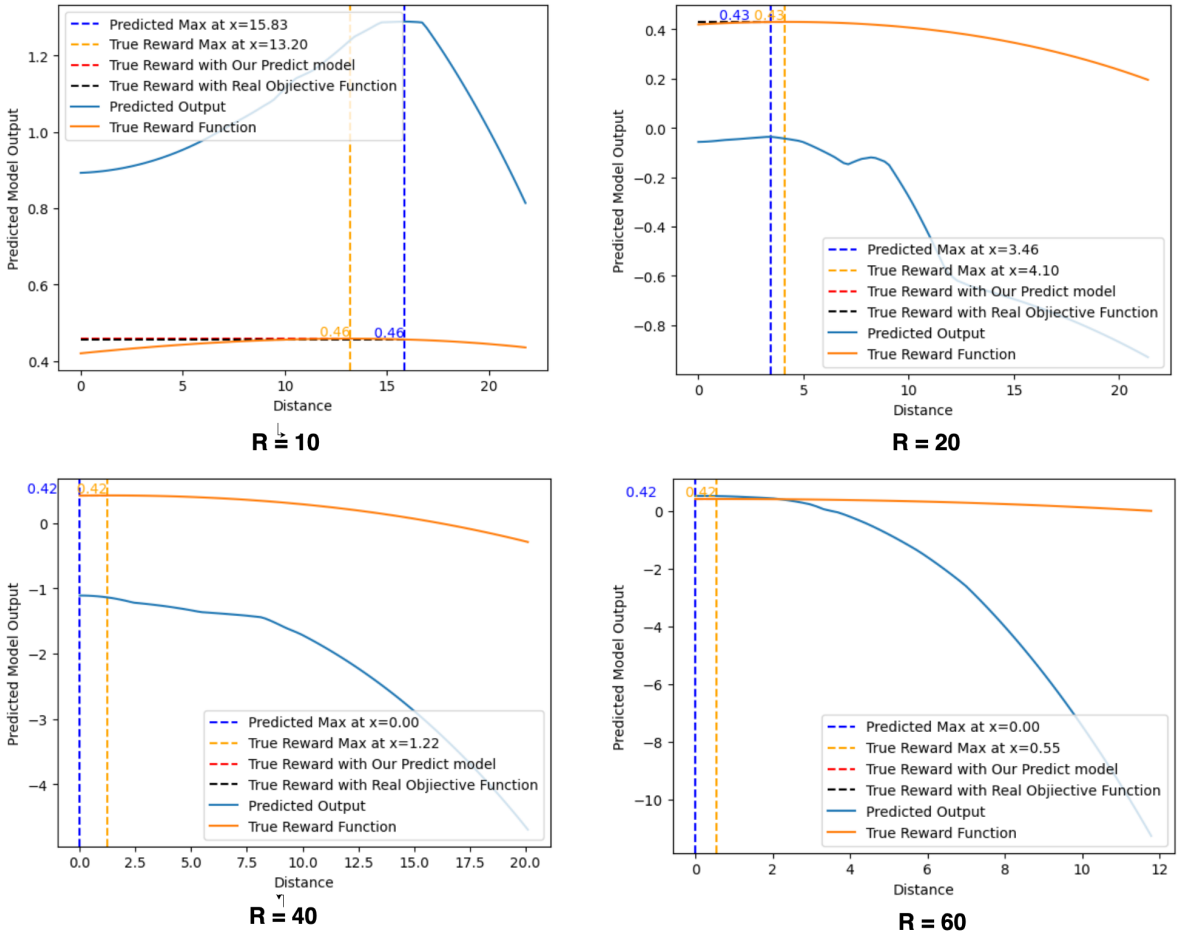


Figure 11: True and evolved perceived objective for *Eciton* bridge building

From the perspective of an individual worker, an ant that encounters a bridge has to make a decision whether to continue foraging and use the bridge as it finds it or to stop foraging and join the bridge, thereby implicitly extending it. It appears that no individual worker can acquire sufficient information from direct sensing to make this decision such that $\rho$ is collectively optimised.

We show in a simulation that objective evolution produces a perception function $N_\theta$ that, when used as the input for cross-learning, allows the workers to collectively optimise the transport. It is perfectly plausible to assume that a worker on the bridge can sense the local bridge density $s_1 = n_b/b$, where $b$ is the length of the bridge, as well as the average encounter rate on the trail $s_2 = (N - n_b)/f$. Our simulation uses these two quantities as the only direct sensory input.

[12] We note that for $\rho$ to be a density, $f$ would have to be an area, not a length. While units would obviously differ, this is numerically equivalent up to a constant under the assumption of a constant trail width. For consistency with [65], we keep $f$ a length to enable easier comparison with this work.

Figure 11 shows the evolved perception functions including the optimal bridge positions (measured in distance from the junction point) according to Reid et al's mathematical optimisation model (dashed orange vertical line) and as resulting from our simulation model (dashed blue vertical line).

Figure 12 gives a comparison of the optimal distance in dependence on the angle of the apparatus as predicted by Reid's optimisation model (orange line) and as reached by objective evolution (blue line). We see that the data is in reasonable agreement.

In summary, we see that objective evolution can offer a possible explanation for the proximate mechanisms individuals use in this concrete biological example to adapt their behaviour so that the collective benefit is optimised even without being able to sense the relevant quantities directly and in the absence of any social coordination mechanisms.

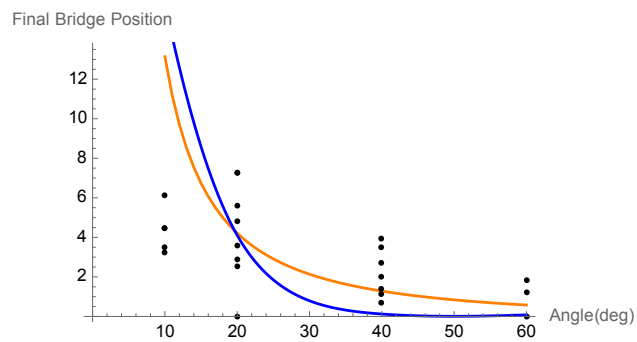


Figure 12: Comparison between the final position of the bridge position predicted by Reid's model (orange) and our model (blue)in the context of the experimental observations (dots) as a function of the angle $\theta$ [65].

## 5 Discussion and conclusions

We have outlined a possible explanation of how pro-social behaviour can be achieved in a collective in the absence of central control by individuals that only make simple decisions based on locally perceived information. Surprisingly, this does not require the individuals to explicitly exchange information regarding the collective benefit. All that is required is the possibility of an evolving perception function that modulates direct sensory input. The only additional assumption that needs to be made for pro-social behaviour to emerge is that the individuals adjust their behaviour via simple reinforcement mechanisms. The latter is firmly inline with the existing literature.

Interestingly, the pro-social decision-making can even be maintained in a complex network of decisions. This is because pro-social atomic decisions allow the relevant information to propagate throughout the decision network.

We have illustrated how this can work with the example of task allocation, where a colony needs to regulate the allocation of workers to a variety of tasks with competing task demands.

Figure 13 summarises the overall mechanism. Individuals control a decision variable $x$ by applying simple reinforcement learning according to an evolving perception function $\phi$ and using their directly sensed local information $\vec{y}$. The joint outcome of this regulation determines the global colony fitness $F$. Since the reinforcement learning mechanisms effectively result in stochastic hill-climbing on $\phi$, the evolutionary pressure causes $\phi$ to assume the shape of a unimodal function that has its peak at the $x$ position that optimises the regulation outcome. Since the evolutionary pressure on $\phi$ originates from the collective benefit the maximum of the regulation outcome has to be pro-social. $\phi$ thus establishes an information flow between the individual behaviour and the collective benefit, which cannot be perceived by the individual and is effectively only "visible" at the evolutionary level.

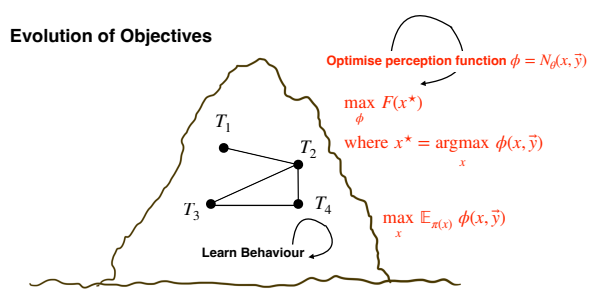


Figure 13: The evolving perception function $\phi$ establishes an information channel from the colony level (fitness) to the decision-making process of the individual.

From the perspective of game theory, $\phi$ is a potential function and reshapes the binary decisions in atomic games as potential games. In this way, it establishes a shared objective so that the pro-social optimum becomes a Nash equilibrium. Since the reinforcement process will reach Nash, rational players, determining their actions according to $\phi$ will behave pro-socially. This reshaping of the individual games turns the whole decision network into a state-based potential game, which enables players to optimise their behaviour in a decentralised fashion.

While our question is related to the emergence of altruism, a core question in evolutionary biology, it is subtly different. We are primarily interested in a question of information flow, namely how an individual can adjust its behaviour to achieve improved collective benefit if it has no direct information on how its behaviour impacts this. The answers

to both questions, however, appear to be intimately connected; maybe not surprisingly so. The missing link in the information flow can only be established by evolution. Arguably the most surprising aspect is that pro-social behaviour must almost necessarily evolve in a collective in which individuals regulate their behaviour with simple reinforcement learning under the assumption of an evolving perception function.

There are, of course, many limitations to this model. It certainly does not cover the complexity of real task allocation in social insects in which not all individuals behave identically [31]. We are not attempting to build a concrete qualitative model of task allocation but just to elucidate a new hypothesis of fundamental principles of collective decision making using task allocation as an example, somewhat in the tradition of [37, 66]

Our model rests on the assumption that locally sensed information can be "sufficiently informative" regarding the collective benefit. What exactly "sufficiently informative" means is somewhat open to interpretation. In the strict mathematical optimisation model detailed above we have taken this to mean that the local information can be used to reliably recover the sign of the gradient of the collective benefit in respect to the decision variable. This is a relatively strong assumption but only necessary in this form if we want to show that the true exact optimum will be reached. True optimisation is generally not to be expected and rarely seen in biological systems. Thus, it should be possible to relax this assumption and still have a system that develops good albeit not optimal pro-social behaviour based on the same principles.

Could this theory be tested experimentally? It should in principle be possible to pick a species with the potential for fast evolution and adaptation, such as clonal raider ants [67] or potentially Pharaoh ants, and to artificially manipulate the reproductive success of the colony according to controlled task profiles. Modifying nutritional geometry might offer an approach to manipulate task profiles in a meaningful way [68]. In this way it could be shown whether the collective allocation process can be evolutionarily manipulated at the colony level. While this would be extremely interesting, it would also appear to be a formidable experimental undertaking. However, experimental evidence that collective behaviours are shaped by colony-level selection has already been presented for Pharaoh ants [69] and breeding programs, at least in honey bees, are not uncommon [70] and can be successful in relatively few generations [71] so that such experiments do not have to be impossible.

For evaluating this hypothesis in the context of a concrete system, whether ants or any other collectively living organism, a crucial question obviously is whether the required neural circuitry could evolve. While this seems possible in general [72] we are probably quite far from answering this in the affirmative, specifically for ants, since their neuroethology is still in its relatively early stages [73]. Within insects, it might be more promising to progress such a line of investigation with honey bees, whose brain is much better understood.

Another, not quite as ambitious, experimental approach would proceed along the lines of Section 4 by conducting further manipulative experiments in which (a proxy of) collective benefit is clearly quantifiable but the individuals cannot sense this benefit directly. Even though few such experiments seem to exist in the literature, they are clearly feasible as [65] and Burd&Howard's work on *Atta* [74, 75] have shown. This would provide a more immediate way to gather further experimental support for our hypothesis.

## Acknowledgments

This work was supported in part by ARC Discovery grant DP200100036 to BM. Some parts of this research have previously been published in "Task Allocation and Information Sharing Strategies: A Study of Collective Decision-Making in Ant Colonies", Yue Yang, PhD thesis Monash University, submitted April 15, 2025.

## References

[1] Simon KA Robson and James FA Traniello. Division of labor in complex societies: a new age of conceptual expansion and integrative analysis, 2016.

[2] Samuel N Beshers and Jennifer H Fewell. Models of division of labor in social insects. *Annual review of entomology*, 46(1):413–440, 2001.

[3] Raphaël Jeanson and Anja Weidenmüller. Interindividual variability in social insects–proximate causes and ultimate consequences. *Biological Reviews*, 89(3):671–687, 2014.

[4] Robert Wilson. Game-theoretic analysis of trading processes. Technical report, Stanford Univ. Inst. for Mathematical Studies in the Social Sciences, 1985.

[5] James FA Traniello and Rebeca B Rosengaus. Ecology, evolution and division of labour in social insects. *Animal behaviour*, 53(1):209–213, 1997.

[6] Bert Hölldobler, Edward O Wilson, et al. *The ants*. Harvard University Press, 1990.

[7] Jennifer H Fewell and Susan M Bertram. Division of labor in a dynamic environment: response by honeybees (apis mellifera) to graded changes in colony pollen stores. *Behavioral ecology and sociobiology*, 46:171–179, 1999.

[8] Stefanie K Gazda, Richard C Connor, Robert K Edgar, and Frank Cox. A division of labour with role specialization in group–hunting bottlenose dolphins (tursiops truncatus) off cedar key, florida. *Proceedings of the Royal Society B: Biological Sciences*, 272(1559):135–140, 2005.

[9] Nathaniel S Rieger, Evan H Stanton, and Catherine A Marler. Division of labour in territorial defence and pup retrieval by pair-bonded california mice, peromyscus californicus. *Animal Behaviour*, 156:67–78, 2019.

[10] Patrick Siefert, Nina Buling, and Bernd Grünewald. Honey bee behaviours within the hive: Insights from long-term video analysis. *PLOS ONE*, 16(3):e0247323, March 2021.

[11] Daniel Charbonneau, Benjamin Blonder, and Anna Dornhaus. Social insects: a model system for network dynamics. In *Temporal networks*, pages 217–244. Springer, 2013.

[12] David Grimaldi, Michael S Engel, and Michael S Engel. *Evolution of the Insects*. Cambridge University Press, 2005.

[13] Daniel Charbonneau and Anna Dornhaus. When doing nothing is something. how task allocation strategies compromise between flexibility, efficiency, and inactive agents. *Journal of Bioeconomics*, 17(3):217–242, 2015.

[14] Deborah M Gordon. From division of labor to the collective behavior of social insects. *Behavioral ecology and sociobiology*, 70:1101–1108, 2016.

[15] Greg J Hunt, Gro V Amdam, David Schlipalius, Christine Emore, Nagesh Sardesai, Christie E Williams, Olav Rueppell, Ernesto Guzmán-Novoa, Miguel Arechavaleta-Velasco, Sathees Chandra, et al. Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften*, 94(4):247–267, 2007.

[16] Ricarda Scheiner, Robert E Page, and Joachim Erber. Sucrose responsiveness and behavioral plasticity in honey bees (apis mellifera). *Apidologie*, 35(2):133–142, 2004.

[17] Ch Brandes. Genetic differences in learning behavior in honeybees (apis mellifera capensis). *Behavior genetics*, 21(3):271–294, 1991.

[18] Deborah M. Gordon. The organization of work in social insect colonies. *Nature*, 380(6570):121–124, 1996.

[19] Deborah M Gordon. The organization of work in social insect colonies. *Complexity*, 8(1):43–46, 2002.

[20] Deborah M. Gordon. From division of labor to the collective behavior of social insects. *Behavioral Ecology and Sociobiology*, 70(7):1101–1108, 2016.

[21] Ana Duarte, Franz J Weissing, Ido Pen, and Laurent Keller. An evolutionary perspective on self-organized division of labor in social insects. *Annual Reviews of Ecology, Evolution, and Systematics*, 42:91–110, 2011.

[22] Deborah M Gordon and Natasha J Mehdiabadi. Encounter rate and task allocation in harvester ants. *Behavioral Ecology and Sociobiology*, 45(5):370–377, 1999.

[23] Daniel Charbonneau and Anna Dornhaus. When doing nothing is something. how task allocation strategies compromise between flexibility, efficiency, and inactive agents. *Journal of Bioeconomics*, 17(3):217–242, 2015.

[24] Chelsea N Cook and Michael D Breed. Social context influences the initiation and threshold of thermoregulatory behaviour in honeybees. *Animal Behaviour*, 86(2):323–329, 2013.

[25] Michael J Greene and Deborah M Gordon. Interaction rate informs harvester ant task decisions. *Behavioral Ecology*, 18(2):451–455, 2007.

[26] Jennifer H Fewell and Jon F Harrison. Scaling of work and energy use in social insect colonies. *Behavioral Ecology and Sociobiology*, pages 1047–1061, 2016.

[27] Danielle P. Mersch. The social mirror for division of labor: what network topology and dynamics can teach us about organization of work in insect societies. *Behavioral Ecology and Sociobiology*, 70(7):1087–1099, 2016.

[28] Michael J Greene and Deborah M Gordon. Interaction rate informs harvester ant task decisions. *Behavioral Ecology*, 18(2):451–455, 2007.

[29] Eric Bonabeau, Guy Theraulaz, and Jean-Louis Deneubourg. Quantitative study of the fixed threshold model for the regulation of division of labour in insect societies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1376):1565–1569, 1996.

[30] Yun Kang and Guy Theraulaz. Dynamical Models of Task Organization in Social Insect Colonies. *Bulletin of Mathematical Biology*, 78(5):879–915, 2016.

[31] Elva J. H. Robinson, Ofer Feinerman, and Nigel R. Franks. Flexible task allocation and the organization of work in ants. *Proceedings of the Royal Society B: Biological Sciences*, 276(1677):4373–4380, 2009.

[32] Deborah M. Gordon. The dynamics of the daily round of the harvester ant colony. *Animal Behaviour*, 34:1402–1419, 1986.

[33] Deborah M. Gordon. Group-level dynamics in harvester ants: young colonies and the role of patrolling. *Animal Behaviour*, 35:833–834, 1987.

[34] Deborah M. Gordon. Dynamics of task switching in harvester ants. *Animal Behaviour*, 38(2):194–204, 1989.

[35] Deborah M Gordon, Brian C Goodwin, and Lynn EH Trainor. A parallel distributed model of the behaviour of ant colonies. *Journal of theoretical Biology*, 156(3):293–307, 1992.

[36] Thomas Schmickl and Istvan Karsai. Integral feedback control is at the core of task allocation and resilience of insect societies. *Proceedings of the National Academy of Sciences of the United States of America*, 115(52):13180–13185, 2018.

[37] Ana Duarte, Franz J Weissing, Ido Pen, and Laurent Keller. An evolutionary perspective on self-organized division of labor in social insects. *Annual Reviews of Ecology, Evolution, and Systematics*, 42:91–110, 2011.

[38] Iain D Couzin. Collective cognition in animal groups. *Trends in cognitive sciences*, 13(1):36–43, 2009.

[39] Ellouise Leadbeater and Lars Chittka. Social learning in insects—from miniature brains to consensus building. *Current biology*, 17(16):R703–R713, 2007.

[40] Scott Camazine, Karl Crailsheim, Norbert Hrassnigg, Gene E Robinson, Bernhard Leonhard, and Helga Kropiunigg. Protein trophallaxis and the regulation of pollen foraging by honey bees (apis mellifera l.). *Apidologie*, 29(1-2):113–126, 1998.

[41] Mario L Muscedere, Tara A Willey, and James FA Traniello. Age and task efficiency in the ant pheidole dentata: young minor workers are not specialist nurses. *Animal Behaviour*, 77(4):911–918, 2009.

[42] Rui Chen, Bernd Meyer, and Julian Garcia. A computational model of task allocation in social insects: ecology and interactions alone can drive specialisation. *Swarm Intelligence*, 14(2):143–170, 2020.

[43] Robert E. Page and Sandra D. Mitchell. Self organization and adaptation in insect societies. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 289–298. The University of Chicago Press on behalf of the Philosophy of Science Association, 1990.

[44] G. Theraulaz, E. Bonabeau, and J. L. Deneubourg. Response threshold reinforcement and division of labour in insect societies. *Proceedings of the Royal Society B: Biological Sciences*, 265(1393):327–332, 1998.

[45] Gene E. Robinson. Regulation of division of labor in insect societies. *Annual Review of Entomology*, 37:637–665, 1992.

[46] Samuel N. Beshers and Jennifer H. Fewell. Models of division of labor in social insects. *Annual Review of Entomology*, 46:413–440, 2001.

[47] Jacques Gautrais, Guy Theraulaz, Jean-Louis Deneubourg, and Carl Anderson. Emergent polyethism as a consequence of increased colony size in insect societies. *Journal of theoretical biology*, 215(3):363–373, 2002.

[48] Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of economic theory*, 77(1):1–14, 1997.

[49] Mikhail Tsetlin. *Automaton Theory and Modeling of Biological Systems*, volume 102 of *Mathematics in Science and Engineering*. Academic Press, New York, 1973.

[50] Moein Khajehnejad, Julian García, and Bernd Meyer. Social learning versus individual learning in the division of labour. *Biology*, 12(5):740, 2023.

[51] Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.

[52] Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1–2):145–156, 1978.

[53] Evripidis Gkanias. *Insect neuroethology of reinforcement learning*. Phd thesis, University of Edinburgh, 2023.

[54] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[55] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, New York, NY, USA, 1st edition, 2008.

[56] Laura-Alexandra Poissonnier, Raphaël Jeanson, and Audrey Dussutour. Competition regulates ant traffic on crowded trails. *eLife*, 8:e48945, 2019.

[57] Tomer J. Czaczkes, Christoph Grüter, and Francis L. W. Ratnieks. Trail pheromone strength in lasius niger: a new assay, population variation and effects of crowding. *Journal of the Royal Society Interface*, 10(86):20120439, 2013.

[58] Audrey Dussutour, Samuel Beshers, Jean-Louis Deneubourg, and Vincent Fourcassié. Crowding increases foraging efficiency in the leaf-cutting ant atta colombica. *Insectes Sociaux*, 54(2):158–165, 2007.

[59] David Raubenheimer and Stephen J Simpson. Nutritional ecology and foraging theory. *Current Opinion in Insect Science*, 27:38–45, 2018. Pests and resistance * Behavioural ecology.

[60] David W. Stephens and John R. Krebs. *Foraging Theory*. Princeton University Press, Princeton, NJ, 1986.

[61] Jason R. Marden. State based potential games. *Automatica*, 48(12):3075–3088, 2012.

[62] Stefanos Leonardos, Will Overman, Ioannis Panageas, and Georgios Piliouras. Global convergence of multi-agent policy gradient in markov potential games, 2021.

[63] Jason R. Marden and Jeff S. Shamma. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior*, 75(2):788–808, July 2012.

[64] William H Sandholm. *Population games and evolutionary dynamics*. MIT press, 2010.

[65] Chris R Reid, Matthew J Lutz, Scott Powell, Albert B Kao, Iain D Couzin, and Simon Garnier. Army ants dynamically adjust living bridges in response to a cost–benefit trade-off. *Proceedings of the National Academy of Sciences*, 112(49):15113–15118, 2015.

[66] Ana Duarte, Ido Pen, Laurent Keller, and Franz J. Weissing. Evolution of self-organized division of labor in a response threshold model. *Behavioral Ecology and Sociobiology*, 66(6):947–957, 2012.

[67] Waring Trible, Vikram Chandra, Kip D. Lacy, Gina Limón, Sean K. McKenzie, Leonora Olivos-Cisneros, and Daniel J. C. Kronauer. A caste differentiation mutant elucidates the evolution of socially parasitic ants. *Current Biology*, 33(6):1047–1058, 2023.

[68] Enikő Csata and Audrey Dussutour. Nutrient regulation in ants (hymenoptera: Formicidae): A review. *Myrmecological News*, 29:111–124, August 2019.

[69] Justin T. Walsh, Simon Garnier, and Timothy A. Linksvayer. Ant collective behavior is heritable and shaped by selection. *The American Naturalist*, 196(5):541–554, 2020.

[70] J. A. Pérez-Sato, N. Châline, S. J. Martin, W. O. H. Hughes, and F. L. W. Ratnieks. Multi-level selection for hygienic behaviour in honeybees. *Heredity*, 102(6):609–615, 2009.

[71] Robert E. Jr. Page and M. Kim Fondrk. The effects of colony-level selection on the social organization of honey bee (*Apis mellifera* l.) colonies: colony-level components of pollen hoarding. *Behavioral Ecology and Sociobiology*, 36(2):135–144, 1995.

[72] Emre O. Neftci and Bruno B. Averbeck. Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence*, 1(3):133–143, 2019.

[73] Dominic D. Frank and Daniel J.C. Kronauer. The budding neuroscience of ant social behavior. *Annual Review of Neuroscience*, 47(1):167–185, 2024.

[74] Martin Burd and Jerome J. Howard. Global optimization from suboptimal parts: foraging sensu lato by leaf-cutting ants. *Behavioral Ecology and Sociobiology*, 59(2):234–242, 2005.

[75] Martin Christopher Burd and Jerome J. Howard. Optimality in a partitioned task performed by social insects. *Biology Letters*, 4(6):627–629, 2008.