

## SOCIAL LEARNING IN A MULTI-AGENT SYSTEM

Jason NOBLE, Daniel W. FRANKS

*School of Computing*

*University of Leeds*

*Leeds LS2 9JT, United Kingdom*

*e-mail: {jasonn, dwfranks}@comp.leeds.ac.uk*

Revised manuscript received 10 November 2003

**Abstract.** In a persistent multi-agent system, it should be possible for new agents to benefit from the accumulated learning of more experienced agents. Parallel reasoning can be applied to the case of newborn animals, and thus the biological literature on social learning may aid in the construction of effective multi-agent systems. Biologists have looked at both the functions of social learning and the mechanisms that enable it. Many researchers have focused on the cognitively complex mechanism of imitation; we will also consider a range of simpler mechanisms that could more easily be implemented in robotic or software agents. Research in artificial life shows that complex global phenomena can arise from simple local rules. Similarly, complex information sharing at the system level may result from quite simple individual learning rules. We demonstrate in simulation that simple mechanisms can outperform imitation in a multi-agent system, and that the effectiveness of any social learning strategy will depend on the agents' environment. Our simple mechanisms have obvious advantages in terms of robustness and design costs.

**Keywords:** Multi-agent systems, social learning, imitation, artificial life, biology

### 1 INTRODUCTION

In the field of artificial intelligence, learning is usually a process through which a solitary agent acquires information about regularities in its environment, and then uses that information to guide its behaviour. In a multi-agent system, it is certainly possible to equip many agents with mechanisms allowing each one to learn individually. However, it has recently been argued [1] that employing learning in the

context of a multi-agent system may actually change the nature of the learning task, and make novel forms of learning possible. Consider a persistent multi-agent system in which new agents enter a world already populated with experienced agents. In one sense, a new agent begins with a blank slate, as it has not yet had an opportunity to learn about its environment. (Of course the agent will likely be hard-wired with an initial set of behaviours to allow it to get started.) However, a new agent may not need to find out everything about the environment for itself: it may well be possible to benefit from the accumulated learning of the population of more experienced agents.

This situation could characterize highly autonomous software agents operating on the internet, for example. A new agent has not yet learned which search engine to try first, or which auction site offers the best bargains. The agent will probably do best by starting with the sites used most frequently by its peers. Note, however, that the situation described also matches the learning problem facing a newborn animal. For all but the most solitary of species, there is valuable information to be gleaned from the behaviour of others. For example, why should a rat risk poisoning by sampling potential foodstuffs at random when it can simply eat the same things that other rats in the same nest are eating? In biology, learning in multi-agent systems has been studied for some time under the heading *social learning*. There may be lessons in the biological literature for those who are interested in engineering effective multi-agent systems.

Researchers in animal behaviour are increasingly in agreement that social learning is an important element in the overall adaptive strategy of many species (see the book by Heyes and Galef [2] for a general review, or the book by Box and Gibson [3] for case studies on social learning in mammals). From lion cubs learning about how to bring down prey, to rats learning through smell about which foods are safe to eat, it is clear that animals use the behaviour of their more experienced peers as a source of valuable information. However, in reviewing the literature on social learning, we often find that although investigators believe social learning to be occurring in a particular species, there is very little known about the specific mechanisms that underlie it. For example, in Box and Gibson's edited collection [3], many contributors discuss the ecology of their chosen species in order to illustrate the likely roles that social learning may play, but can only speculate about exactly *how* the animals in question actually learn through the observation of another's behaviour.

Another curious fact about the social learning literature is the emphasis on imitation as a potential mechanism. Imitation – or “imitative learning” [4, p. 324] – is a cognitively complex process that requires an animal to not only perceive and reproduce the bodily movements of another, but to understand the changes in the environment caused by the other animal's behaviour, and finally to be able to grasp the “intentional relations” between these, i.e., to know how and why the behaviour is supposed to bring about the goal. This emphasis on imitation is particularly strong in the case of work on primates [5]. A focus on imitation is not surprising: it is likely to be implicated in the explosive cultural evolution of our own species, and so

we have a natural curiosity about the extent to which other animals in general, and our primate cousins in particular, might share this ability. Nevertheless, looking at imitation alone may blind us to the potential of much simpler mechanisms through which the behaviour of one animal can influence the learning experience of another. Clearly, simpler mechanisms have the advantage of being easier to implement in a human-engineered multi-agent system.

Although biologists sometimes do not understand *how* social learning occurs, there is a body of work on *why* it might occur, i.e., on the conditions under which it would offer a selective advantage. Models of cultural transmission [6] and related phenomena such as “highly horizontal transmission” [7] help delineate the conditions under which it will be advantageous for individuals to learn from others rather than finding things out for themselves. However, these models tend to result in rather general conclusions. For example, Laland and co-authors [7] have summarized work on cultural transmission by relating it to rates of change in the environment: in static environments, genetic transmission of hard-wired behaviour patterns can do the job, and in rapidly changing environments, only individual learning can keep up, whereas at intermediate rates of change, social learning will offer an advantage. Results like these are useful, but to get a complete picture of social learning we also need to understand the mechanisms involved.

Individual-based simulations, as used in artificial life research, represent an excellent way of exploring possible social learning mechanisms. ALife models of this type have certainly been explored in related domains such as communication, dominance and territorial behaviour, and flocking or schooling. There are also a significant number of robotics researchers who share the primatologists’ interest in imitation and see it as a way of improving human–robot or robot–robot communication (see [8] for a review of this work).

However, as yet relatively few ALife models have looked at social learning using mechanisms simpler than imitation. One example is work by Toquenaga et al. [9] who constructed a simulation of foraging and nesting behaviour in egrets. They used their model to demonstrate that learning socially about the location of food, through simply approaching other birds, is more likely to evolve when food resources are patchy rather than evenly distributed. Toquenaga et al. suggest that flock foraging and colonial roosting may be based on this principle. In the same vein, Noble, Todd, and Tuci [10] considered the selective pressures affecting social learning about food in Norway rats. They found that a surprisingly simple strategy, which involved ignoring signs of food poisoning in others, was favoured as long as toxins were very dangerous. This finding was used to explain apparently paradoxical aspects of rat behaviour.

Noble and Todd [11] have previously argued that a focus on imitation has led to the power of simpler social learning processes being underestimated; they describe a number of candidate simple mechanisms, and encourage the use of simulation models as a way of demonstrating the potential of such mechanisms. The purpose of the current paper is to follow up on this work by instantiating some of the mechanisms discussed by Noble and Todd [11], in a simple environment that reflects

the learning challenges faced by a hypothetical animal or agent. The central aim is to show that simple mechanisms for social learning can perform well under the right environmental circumstances. We hope to thereby demonstrate that imitative learning has received a disproportionate amount of attention. We also hope to show by analogy that designers of multi-agent systems could improve the performance of their agents by moving beyond individual learning and incorporating social learning.

We must stress that even though our simulated agents are inspired by several different primates, our goal is not to model learning in any particular species, but to set up a specific, controlled environment in which the performance of different social learning strategies can be compared.

In looking at the data that is output by our simulation, we adopted the perspective of an observer who records agent behaviour and tries to come up with conclusions about the underlying learning mechanisms. Human observers of animal and robot behaviour have a propensity to invoke mechanisms that are more complex than those strictly needed to explain that behaviour [11, 12]. For example, the presence of a distinctive behaviour pattern across all members of a group could be taken as evidence for sophisticated social learning, but might result simply from similar individual learning experiences. We hope to show that much of the data a field observer might plausibly collect about a population could easily lead them astray in theorizing about what the agents were really doing.

## **2 THE SIMULATION**

The environment consists of a  $10 \times 10$  toroidal grid world, in which each square is either empty (20%) or contains a resource (80%). In addition, a grid square may contain a tool that agents are free to pick up and take away; agents may also drop tools. Agents occupy one square at a time, and more than one agent can occupy the same square. At each time step an agent chooses one of 12 distinct actions, and performing an action may lead to positive or negative payoffs depending on the context. An agent's choice of action is based on feeding its perceptual state into a reinforcement learning algorithm and choosing the action with the highest expected payoff (with a 20% chance of choosing a random action, in order to encourage exploration). Agents can perceive their local resource and the available tools – one on the ground and one in their grasp – as well as the presence of other agents (see Figure 1). They can also perceive the resource type in a randomly selected square drawn from the eight-square neighbourhood around their current location. If they choose to move, they will travel to this square.

The environment is deliberately simple, even playful: it is certainly not meant to be realistic. The guiding principles were, first, not to include an excessive number of distinct states, as this would make things difficult for the agents' reinforcement learning algorithm, and second, to ensure that some potential rewards were easily discovered and that others were harder to find (although possibly made easier with the right kind of social learning). The names we have given to the various resources,

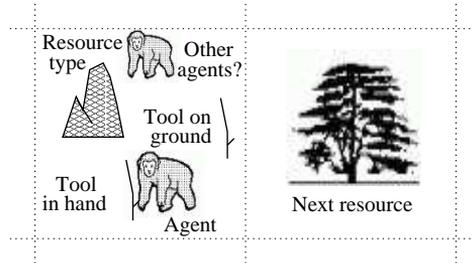


Fig. 1. Aspects of the environment that can be perceived by each agent: the focal agent is in the left-hand grid square, and can perceive its local resource type – in this case a termite nest – the fact that it is grasping a stick, the presence of another stick on the ground, the presence of a second agent, and the proximity of a tree in the next grid square. If the agent chose the *travel* action, it would move to the tree

tools, and actions clearly suggest primate environments and behaviours, but we trust that the reader can appreciate that the same simulation might equally be used to model robots or software agents in an appropriate environment.

Generally, different resources are associated with distinct payoffs, and require the right action or sequence of actions to obtain them. The different resource types are listed below. Actions and tools are shown in italics, means and standard deviations for payoffs are shown as  $\mu$  and  $\sigma$  respectively,  $P$  denotes the prevalence of that resource in the environment, and  $n$  refers to the number of times the resource can be used before it is exhausted. An exhausted resource reverts to empty, and a replacement appears in a randomly chosen empty square elsewhere in the environment.

**Tree:** being in the shade of a tree is pleasant, with a small constant reward ( $\mu = 1$ ,  $\sigma = 0$ ) regardless of the chosen action.  $P = 10\%$ ,  $n = \infty$ .

**Berries:** an easy resource to figure out, berries are assumed to be automatically eaten ( $\mu = 10$ ,  $\sigma = 5$ ) regardless of action. If the agent *shakes* the berry bush, twice the payoff is obtained ( $\mu = 20$ ,  $\sigma = 5$ ).  $P = 10\%$ ,  $n = 20$ .

**Nettles:** normally cause a nasty sting ( $\mu = -10$ ,  $\sigma = 5$ ) regardless of action but if an agent *folds* the nettle leaves, they can be safely eaten ( $\mu = 30$ ,  $\sigma = 5$ ).  $P = 10\%$ ,  $n = 20$ .

**Termite nest:** the best way to get a meal of termites is to *poke* the nest with a *peeled stick* ( $\mu = 100$ ,  $\sigma = 20$ ) – the termites swarm on the stick and the agent can lick it clean. Also a reward for *poking* with a *stick* ( $\mu = 50$ ,  $\sigma = 10$ ) or without any tool ( $\mu = 20$ ,  $\sigma = 5$ ). *Bashing* the nest with a *stone* or a *big stone* will get some termites ( $\mu = 35$ ,  $\sigma = 10$ ).  $P = 10\%$ ,  $n = 10$ .

**Coconuts:** need to be *bashed* to have a chance of opening them and getting the milk ( $\mu = 150$ ,  $\sigma = 20$ ).  $P = 10\%$ ,  $n = 5$ . Chance of successfully opening

a coconut depends on whether *stones* or *big stones* are available in hand and on the ground.

	<i>On ground</i>		
	Nothing	Stone	Big stone
<i>Holding</i>			
None	0.01	0.01	0.01
Stone	0.2	0.6	1.0
Big stone	0.4	0.6	0.8

Table 1

**Monkeys:** can be caught and eaten ( $\mu = 200$ ,  $\sigma = 20$ ) if an agent *chases* them, but only if a second agent is present to block their escape route. Alternatively an agent can *throw* a *stone* for a 10% chance of success, or *throw* a *stick*, a *peeled stick*, or a *big stone* for a 2% chance of success.  $P = 10\%$ ,  $n = 2$ .

**Thorns:** are painful ( $\mu = -20$ ,  $\sigma = 5$ ) regardless of action.  $P = 10\%$ ,  $n = \infty$ .

**Beehive:** there is a 20% chance that the bees will fly out and sting the agent ( $\mu = -200$ ,  $\sigma = 50$ ) regardless of action.  $P = 5\%$ ,  $n = \infty$ . If the agent *shakes*, *pokes*, or *bashes* the hive the chance of stinging is 100%. The hardest trick to learn is getting honey from the hive ( $\mu = 500$ ,  $\sigma = 50$ ,  $n = 1$ ): the agent must either *bash* the nest with a *stone* or a *big stone* for a 50% chance of success (and a 100% chance of getting stung) or smoke the bees out by starting a fire, i.e., *rubbing* two *sticks* of either type together (one on the ground and one in the hand).

**Wasp nest:** the same negative properties as the beehive but with no chance of reward.  $P = 5\%$ ,  $n = \infty$ .

An important action not covered above is *scratching*, which leads to a modest reward ( $\mu = 1$ ,  $\sigma = 0$ ;  $\mu = 2$  if the agent is carrying either kind of *stick*) regardless of the local resource. Scratching provides a kind of local optimum, in that it is easy for an agent to get locked into pursuing the modest rewards of this activity, perhaps under a tree, rather than moving around looking for more profitable resources.

Other actions include *peeling*, which converts a *stick* in the agent's hand into a *peeled stick*; *swapping*, which exchanges the tool on the ground, if any, for the one in the agent's hand, if any; and *travelling*, which moves the agent to the nearby resource they have most recently seen. Note that when an agent attempts an irrelevant action, such as bashing a tree, the payoff is always zero.

Tools are distributed generously, with 40 sticks, 20 stones, and 20 large stones present in the environment. Peeled sticks only appear if an agent peels a stick. Stones do not wear out, but sticks and peeled sticks break after 1000 or 500 time steps respectively; a replacement stick appears at a random location.

## 2.1 Reinforcement Learning Agents

The agent population is 25, but initially agents enter the world one at a time. Every 400 timesteps a new agent is born. When 25 agents are present, the oldest agent is always killed off to make room for the youngest; an agent's lifespan is therefore 10,000 timesteps. The middle-aged agent born 12 places ahead of a newborn is nominally regarded as its parent, and the new agent appears in the same square. The simulation is run for 100,000 time steps, or 10 agent lifetimes. Even though the simulation has "generations" of a sort, there is no genetic aspect to agent behaviour: each individual comes into the world as a blank slate and must learn for itself how to maximize payoffs.

Each agent makes its decisions on the basis of a history of reinforcement: it has learned to associate some states and actions with a reward and others with punishment. The different social learning mechanisms implemented all assume this underlying capacity for reinforcement learning. The specific learning algorithm used was Q-learning [13] which allows an agent to take into account not just immediate but also delayed payoffs. This would be important, for example, in realizing that a higher long-term payoff could be gained by picking up and peeling a stick rather than just poking the termite nest immediately. The learning rate was 0.1, the temporal discounting factor was 0.5, and the proportion of random actions was 0.2.

Q-learning in its simplest form does not include any generalization: if an agent learns that it is good to shake a berry bush when there is a stone on the ground, it will not automatically transfer that knowledge to the case where there is no stone on the ground. Given 60,000 state-action pairs<sup>1</sup> and a lifetime of only 10,000 timesteps it is clear that an agent could not exhaustively sample its space of perceived situations. We therefore included some basic generalization: after updating a state-action pair through Q-learning, all other state-action pairs that differed only in terms of the tools available, the presence of others, and the resource seen in the distance (exception: the resource being left behind for the travel action) also had their expected returns updated by a small amount in the same direction.

## 2.2 Implementing Social Learning

To find baseline levels for payoffs in this environment, we looked at the performance of agents with completely random behaviour, and at that of agents capable only of individual learning. We then implemented three of the simple mechanisms described by Noble and Todd [11]: following or stimulus enhancement, contagious behaviour, and emulation. In addition, we implemented a version of imitative learning. Finally, we looked at the performance obtained when following was combined with each of the other mechanisms.

---

<sup>1</sup> 10 resources  $\times$  5 tool-on-ground options  $\times$  5 tool-in-hand options  $\times$  2 others-present options  $\times$  10 next-resource options  $\times$  12 actions = 60,000.

**Following:** (otherwise known as stimulus enhancement) is implemented simply by having each agent occupy the same square as its parent for the first 25% of its lifetime. Beyond that, agents learn for themselves.

**Contagious behaviour:** if another agent is present, there is a 10% probability that the first agent will do whatever the second just did. An instinctive behavioural contagion is assumed, as when one person's yawning stimulates another to yawn; actions are equally likely to be copied regardless of their payoff.

**Emulation:** if another agent is present and obtains a non-zero payoff, the first agent adjusts its estimated return for *all* state-action pairs related to its current state. The idea is that the agent has become aware of an affordance in the local environment, but has no idea about what specific action is needed to exploit it.

**Imitation:** if another agent is present and obtains a non-zero payoff, the first agent is able to take the perspective of the second. The first agent notes the perceptual state of the other, as well as the action the other chose, and updates its own relevant state-action pair accordingly. The advance on emulation is that the first agent not only knows which action achieved the reward, but also can perceive the other agent's state as opposed to its own (e.g., it might be holding a stick).

## 2.3 Data Collection

How can we collect data from this simulation that parallels the data that field observers would collect when watching real animals? We assumed that observers watched the population for the final 1000 timesteps; i.e., they had a snapshot of the group's behaviour rather than being able to observe the rise and fall of socially transmitted behaviours in the long term. We then looked at the average payoff per agent during these 1000 timesteps. We can imagine that observers look at distinct groups in similar environments and compare the average energy intake across groups. It is natural to assume that a population equipped with a richer social learning ability will be able to extract more energy from its environment.

We also wanted to measure behavioural convergence, i.e., the degree to which the various members of the population tend to choose the same action when faced with the same situation. This was done by compiling a list of 1000 states perceived by members of the population (the state of a randomly chosen individual was noted at each timestep). Then all 25 members of the population were faced with each of the 1000 states in turn, and the proportion of individuals who chose the modal response for each state was recorded and averaged. We can imagine that our observers have an outstanding ability to set up naturalistic experiments in the field, and are able to engineer things so that each member of the population can be placed into the same series of states and allowed to choose an action.

We looked at behavioural convergence because uniformity of behaviour within a population, especially when combined with behavioural variation *across* populations, is usually taken as a sign that social learning is going on. For example, if one group is observed to hunt monkeys by chasing them down, while another group

tends to throw stones at them, then it seems to follow that each animal is not arriving at its own solution to the problem, but rather is being influenced by its fellow group members.

Finally, we looked at simple counts of the occurrence of various behaviours: the performance of more complex and hard-to-learn behaviours would probably be taken as a sign of sophisticated social learning. So, for example, if one species has mastered the trick of peeling sticks before poking them into termite nests, and another related species has not, one might assume that the former species has a better social learning faculty.

### 3 RESULTS AND DISCUSSION

All of the learning mechanisms, both individual and social, ensured that the agents did far better than they would by choosing a random action. Figure 2 shows that the mean payoff per agent per timestep was at least 6 units; the mean payoff for random behaviour was  $-8.997$  units. Behaviours such as scratching, eating berries, folding nettle leaves, and poking termite nests (although not often with tools) were all quite common and allowed the agents to extract a good return from the environment.

Omitting for a moment the cases in which following is combined with another mechanism, we have the surprising result that the best mean payoff is for emulation, not imitative learning. Why does this occur? Emulation means that an agent increases its estimated return for *all* state-action pairs associated with any state in which it witnesses another agent receiving a positive payoff. Effectively this means that emulation encourages exploration around beneficial resources, and results in more inquisitive agents who are persistent enough to crack a tricky resource like nettle leaves. Imitators, on the other hand, are quite conservative. Imitative learning is sophisticated enough to allow an agent to pick up any clever trick it might witness, but imitators appear to be caught in a vicious circle in which the only role models they observe are other conservative imitators (see the behavioural profiles in Figure 4).

Figure 2 also shows that the simple matter of following your parent around for the first quarter of your life is enough to increase your mean payoff significantly. This is more impressive than it seems, as the imposition of following behaviour threatens to *decrease* payoffs by increasing feeding competition: a resource will run out twice as fast if it is being consumed by both parent and child. Contagious behaviour, on the other hand, did not do any better than individual learning. It may be that occasional instinctive copying is unhelpful because agents already try random actions 20% of the time and will thus reach the same result in the end.

When following behaviour is combined with other mechanisms, the story becomes more complex. It is obviously not the case that following and the other three mechanisms mesh together in a simple additive fashion: the efficiency of imitation and contagious behaviour is increased, but the efficiency of emulation is decreased.

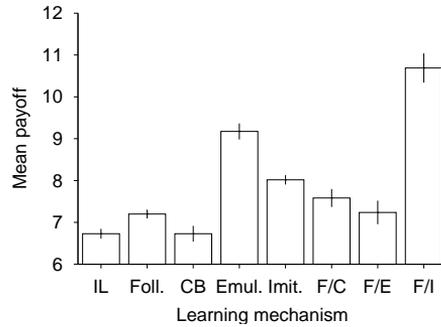


Fig. 2. Mean payoff per agent per timestep across the different learning conditions: individual learning, following, contagious behaviour, emulation, imitation, following with contagious behaviour, following with emulation, and following with imitation. Each data point is averaged across 10 runs, and  $\pm 1$  standard error is shown.

Clearly it is important to imitate or emulate the right people: an experienced parent in one case, and a random cross-section of society in the other.

Behavioural convergence (Figure 3) was 0.55 even in the individual learning case. The convergence score for random behaviour was inevitably much lower at 0.18, and thus we are reminded that similar behaviour may be brought about by similar learning experiences rather than any form of social learning. Convergence was highest in the two imitation conditions, which makes sense. However, the low convergence for emulation belies the fact that these agents were doing well in terms of mean payoff, and convergence rates are the same across the two imitation conditions, despite the fact that it is only in the following/imitation condition that high payoffs are achieved. An observation of high or low behavioural convergence, which might lead one to suspect the presence or absence of effective social learning, is meaningless considered alone.

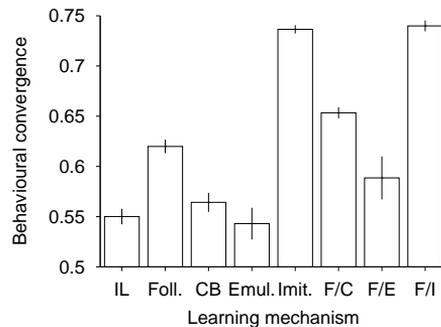


Fig. 3. Behavioural convergence across the different learning conditions. Each data point is averaged across 10 runs, and  $\pm 1$  standard error is shown.

Figure 4 shows frequency counts for several behaviours across the different conditions. It shows that different learning mechanisms lead to quite different behavioural profiles, e.g., emulators spend a lot of time folding nettle leaves, whereas imitators have mastered the art of shaking berry bushes. It is not simply the case that the more “advanced” the social learning mechanism, the more behaviours are added to the repertoire.

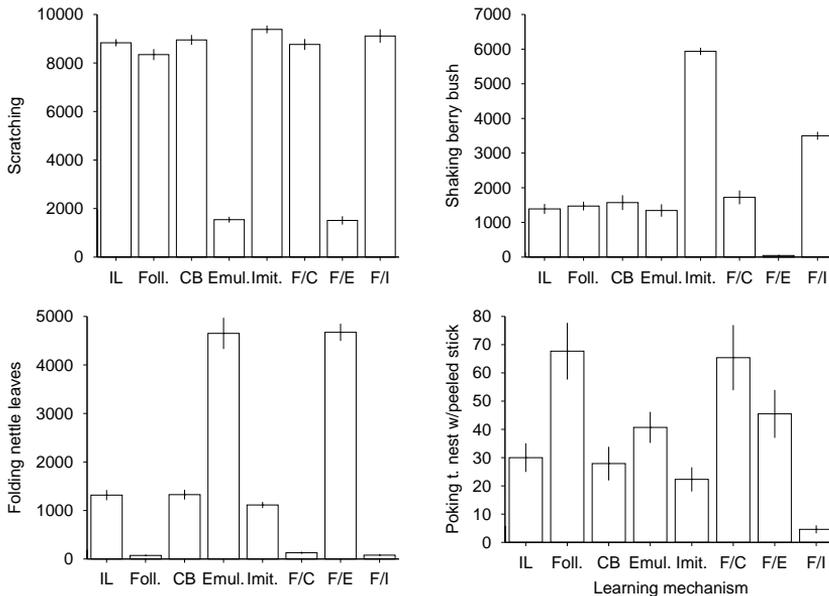


Fig. 4. Frequencies with which various behaviours were observed across the different learning conditions, clockwise from top left: scratching, shaking berry bushes, poking termite nests with peeled sticks, and folding nettle leaves. Each data point is averaged across 10 runs, and  $\pm 1$  standard error is shown.

Another surprising finding is that some behaviours that would seem to suit imitative learning, such as learning to peel a stick and then poke it into a termite nest, are in fact more often performed by agents with simpler mechanisms (Figure 4, bottom right). Agents who follow their parents are the ones who do this most often. Even so, it seems that this behaviour was largely too hard: note the low frequency counts overall. Imitators could certainly have picked up this trick, but again it seems they did not have access to the right role models.

## 4 CONCLUSIONS

We hope to have demonstrated that complex imitative learning is not always going to be the best strategy in a multi-agent system, and that the effectiveness of any

social learning strategy will depend on details of the environment and additional facts about the agents' lifestyles. We readily admit that the impressive performance of emulation in our model, for example, is not likely to be a general result, but is tied to the particulars of the simulated environment and the underlying reinforcement learning system. But that is exactly the point.

The argument which inspired this paper [11] had two aspects: first, simpler social learning mechanisms are inherently more likely to evolve; one reason for this is that complex mechanisms such as imitation will involve significant costs. It follows that we should look for simple mechanisms in nature. Second, in cases of doubt about which mechanism is at work, from an epistemological point of view it is good practice to start by proposing simpler mechanisms. The current simulation results show that an even stronger argument is possible: even if imitation was available via a cost-free, one-step mutation, it still might not be the best learning mechanism for a given species.

We believe our model also shows that observers comparing how well two different groups of animals get on in similar environments, or looking at data on "traditions" or "proto-cultural behaviour" (i.e., behavioural convergence) may find that data to be quite misleading as to the underlying social learning mechanisms at work. It is not the case, for example, that low behavioural convergence necessarily means an absence of social learning. One has to be careful in looking at exactly how behaviours are transmitted from one individual to another. Theories about the adaptive value of social learning, such as [7], are useful, but ideally we want to move beyond such theories and look at hypotheses about the utility of a particular mechanism in a particular environment.

We deliberately omitted an evolutionary dimension from the current model as we wanted to look at the effects of learning alone. However, future work could incorporate an evolutionary aspect, in which different social learning mechanisms were selected for over generational time. The mean payoff results certainly suggest that this would not be a straightforward story of hill-climbing with imitation at the peak. Let us assume that a single-step mutation  $F$  controls following behaviour, and a two-step mutation leads from individual learning to emulation ( $S_1$ ) and thence to imitative learning ( $S_2$ ). This latter assumption is reasonable as imitative learning can be seen as a "sharpened" version of emulation. For an evolving population of individual learners, the final strategy will be emulation if  $S_1$  occurs first, as  $F$  and  $S_2$  will both be associated with a loss in fitness. On the other hand, if  $F$  occurs first, both  $S_1$  and then  $S_2$  will mean increases in fitness, and the population will end up as followers/imitators.

It seems, then, that a focus on imitation as the only kind of social learning worth having is misplaced. Researchers interested in imitation as a potential tool for instructing robots or software agents would do well to consider giving their creations more spartan mechanisms, which have the benefit of being far easier to design and implement, and are inherently more likely to prove robust due to their very simplicity. Assuming that an agent is equipped with a general reinforcement learning system, our results suggest that it would be a good start

to simply have it follow around a more experienced agent, such as a human being.

More generally, if we aspire to build systems that include multiple agents that need to learn about their environment, we could do worse than to start by looking at how social learning allows information to be shared in a nest of rats, a flock of birds, or a troop of monkeys. The evidence from the biological literature and from the simulation presented here is that, with respect to multi-agent learning, much can be achieved with simple mechanisms.

## REFERENCES

- [1] ALONSO, E.—D'INVERNO, M.—KUDENKO, D.—LUCK, M.—NOBLE, J.: Learning in Multi-Agent Systems. *Knowledge Engineering Review*, Vol. 16, 2001, pp. 277–284.
- [2] HEYES, C. M.—GALEF, B. G., JR. (Eds.): *Social Learning in Animals: The Roots of Culture*. Academic Press, San Diego, CA, 1996.
- [3] BOX, H. O.—GIBSON, K. R. (Eds.): *Mammalian Social Learning: Comparative and Ecological Perspectives*. Volume 72 of the Symposia of the Zoological Society of London. Cambridge University Press, 1999.
- [4] TOMASELLO, M.: Do apes ape? In [2], pp. 319–346.
- [5] WHITEN, A.: Imitation of the Sequential Structure of Actions by Chimpanzees *Pan troglodytes*. *Journal of Comparative Psychology*, Vol. 112, 1998, No. 3, pp. 270–281.
- [6] BOYD, R.—RICHERSON, P. J.: *Culture and the Evolutionary Process*. The University of Chicago Press, 1985.
- [7] LALAND, K. N.—RICHERSON, P. J.—BOYD, R.: Developing a theory of animal social learning. In [2], pp. 129–154.
- [8] SCHAAL, S.: Is Imitation Learning the Route to Humanoid Robots? *Trends in Cognitive Sciences*, Vol. 3, 1999, No. 6, pp. 233–242.
- [9] TOQUENAGA, Y.—KAJITANI, I.—HOSHINO, T.: Egrets of a Feather Flock Together. *Artificial Life*, Vol. 1, 1995, No. 4, pp. 391–411.
- [10] NOBLE, J.—TODD, P. M.—TUCI, E.: Explaining Social Learning of Food Preferences without Aversions: An Evolutionary Simulation Model of Norway rats. *Proceedings of the Royal Society of London: Biological Sciences*, Vol. 268, 2001, No. 1463, pp. 141–149.
- [11] NOBLE, J.—TODD, P. M.: Imitation or Something Simpler? Modelling Simple Mechanisms for Social Information Processing. In: K. Dautenhahn and C. Nehaniv (Eds.): *Imitation in Animals and Artifacts*, pp. 423–439. MIT Press, Cambridge, MA, 2002.
- [12] BRAITENBERG, V.: *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1984.
- [13] WATKINS, C. J. C. H.—DAYAN, P.: Q-Learning. *Machine Learning*, Vol. 8, 1992, pp. 279–292.



**Jason NOBLE** did his doctoral thesis on the evolution of communication while working in the School of Cognitive and Computing Sciences at the University of Sussex. After graduating in 1999, he spent two years in a postdoctoral position at the Center for Adaptive Behavior and Cognition, part of the Max Planck Institut für Bildungsforschung in Berlin, where he worked on simulation models for evolutionary psychology. In 2001 he took up his current position, a departmental research fellowship in the School of Computing at the University of Leeds. He has helped establish a biosystems research group in the School and is active in the fields of biological simulation and bio-inspired optimization.



**Daniel W. FRANKS** graduated with a degree in computer science from Sheffield Hallam University in 2001. He is currently a Ph.D. student in the biosystems research group, part of the School of Computing at the University of Leeds, where he is supervised by Dr. Jason Noble. His main area of research involves using simulation models to look at the evolution of aposematism and mimicry. He also has an interest in the evolution of social learning and culture.