PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

**Cite this article:** Leimar O, Dall SRX, Houston AI, McNamara JM. 2022 Behavioural specialization and learning in social networks. *Proc. R. Soc. B* **289**: 20220954. https://doi.org/10.1098/rspb.2022.0954

Received: 17 May 2022
Accepted: 14 July 2022

**Subject Category:**
Behaviour

**Subject Areas:**
behaviour, theoretical biology, cognition

**Keywords:**
behavioural consistency, animal personality, reinforcement learning, game theory

**Authors for correspondence:**
Olof Leimar
e-mail: olof.leimar@zoologi.su.se
John M. McNamara
e-mail: john.mcnamara@bristol.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6124001.

THE ROYAL SOCIETY PUBLISHING

# Behavioural specialization and learning in social networks

Olof Leimar[1], Sasha R. X. Dall[2], Alasdair I. Houston[3] and John M. McNamara[4]

[1]Department of Zoology, Stockholm University, 106 91 Stockholm, Sweden
[2]Centre for Ecology and Conservation, University of Exeter, Penryn TR10 9FE, UK
[3]School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK
[4]School of Mathematics, University of Bristol, Bristol BS8 1UG, UK

OL, 0000-0001-8621-6977; AIH, 0000-0002-5769-7692; JMM, 0000-0002-4235-3045

Interactions in social groups can promote behavioural specialization. One way this can happen is when individuals engage in activities with two behavioural options and learn which option to choose. We analyse interactions in groups where individuals learn from playing games with two actions and negatively frequency-dependent payoffs, such as producer–scrounger, caller–satellite, or hawk–dove games. Group members are placed in social networks, characterized by the group size and the number of neighbours to interact with, ranging from just a few neighbours to interactions between all group members. The networks we analyse include ring lattices and the much-studied small-world networks. By implementing two basic reinforcement-learning approaches, action–value learning and actor–critic learning, in different games, we find that individuals often show behavioural specialization. Specialization develops more rapidly when there are few neighbours in a network and when learning rates are high. There can be learned specialization also with many neighbours, but we show that, for action–value learning, behavioural consistency over time is higher with a smaller number of neighbours. We conclude that frequency-dependent competition for resources is a main driver of specialization. We discuss our theoretical results in relation to experimental and field observations of behavioural specialization in social situations.

## 1. Introduction

The issue of why individuals differ in behavioural tendencies has received much attention in recent years [1–3], with a focus on genetic or other differences emerging early in development. One influential idea is that frequency dependence can promote specialization [4]. Here, we explore the possibility that learning with frequency-dependent rewards, such as rewards from playing games, can give rise to specialization.

Early in the development of game theory in biology it was found that there can be asymmetric evolutionarily stable strategies (ESSs), with the 'bourgeois' ESS for the hawk–dove game as a well-known example [5,6]. In this game two individuals interact and the ESS is polarized, in the sense that one player uses hawk and the other dove. It turns out that there are similar ESSs for group sizes larger than two, such that players polarize into using different behavioural options [7]. The selection favouring polarization is stronger in smaller groups. Here, we extend the idea of behavioural specialization to groups interacting in a social network, where the group size might be large but the number of network neighbours of an individual could be small. We focus on learning leading to specialization, because social interactions are often repeated in a group and persist over times long enough for learning to be important.

The idea that frequency-dependent learning leads to specialization was introduced some time ago [8], with producer–scrounger relations [9] as a possible example. Recent foraging experiments have demonstrated that negatively

frequency-dependent learning can result in behavioural diversity, with preferences becoming established after 25–50 foraging experiences per individual [10], which corresponds to rather fast learning. Producer–scrounger experiments with birds also indicate that behavioural specialization involves learning [11] and that behaviour is consistent over time if the social environment (the flock mates) is constant, but tends to change in new social environments [12]. Stable producer–scrounger relations are also found in bats that live in large groups, but interact when foraging with a small number of other individuals, thus forming a social network [13].

The general idea of frequency-dependent learning in social groups is thus well established and has experimental support, but up to now it is not known how the social environment, in particular, the number of network neighbours, influences the rate of establishment and the temporal stability of behavioural specialization. Our aim here is to examine these questions, using game-theory models of groups of individuals that learn, based on rewards (i.e. payoffs), which actions to prefer when interacting with neighbours in a social network. In addition to the producer–scrounger game [9,14,15], where individuals have the options to produce (i.e. search for a food source) or to scrounge (i.e. attempt to exploit food sources found by producers), we also study a caller–satellite game [16–18] and the hawk–dove game [6,19].

Calling and acting as satellite are male behavioural options in species in which males call to attract females or, alternatively, act as satellites to nearby callers, attempting to intercept approaching females. In anurans, calling involves a form of male–male competition [20], so that males can be seen as interacting with neighbours in a social network [21], and the situation could be similar in other species with calling males.

The hawk–dove game is frequently used to examine contests between individuals, but it gives a highly schematic of such behaviour. Contests in social groups often produce dominance hierarchies with individual recognition, but there may be examples of fights in social groups with limited or no individual recognition, such as in some species of crickets [22,23], where learning to prefer hawk versus dove in aggressive interactions, which corresponds to dominant versus subordinate behaviour, can provide a modelling starting point. Also, for repeated hawk–dove interactions between two individuals, a reinforcement-learning model showed polarization, one individual using hawk and the other dove [24].

In each of the games we study, we idealize the situation by assuming that group members do not differ in traits like learning, foraging or fighting abilities, in order to focus on the particular effects of frequency-dependent learning. For learning, we use reinforcement-learning approaches (action–value learning and actor–critic learning) that encapsulate basic learning concepts from animal psychology [25]. Action–value learning is the simplest of these and is an implementation of the Rescorla–Wagner model for operant conditioning. The learned probabilities of choosing actions are based on differences in estimates of the value (expected reward) from using an action.

## 2. Methods

Our general approach is to study reinforcement learning in games with two actions (behavioural options) for individuals in a group of size $N$ that interact with neighbours in a social network. Figure 1 shows the kind of networks we study, with an illustration of two learning trajectories for a producer–scrounger game (see the electronic supplementary material for detailed descriptions of our methods).

The networks we use are either regular ring lattices (figure 1*a*) or small-world networks (figure 1*b*) [26,27]. The nodes of a network represent a group of $N$ individuals and the network edges represent connections between a group member and the neighbours with which it interacts. For a ring lattice, each group member has $K$ neighbours (figure 1*a*). A small-world network is obtained from a ring lattice by 'rewiring' some connections to a random, previously unconnected group member, with $p_{\mathrm{rew}}$ the probability of rewiring (figure 1*b*).

We use two implementations of reinforcement learning [25]: action–value learning and actor–critic learning. Action–value learning is a simple implementation of the classical Rescorla–Wagner model of conditioning [28], modified for instrumental conditioning. With two actions, for instance produce (P) and scrounge (S), a learning individual maintains and updates two estimates (e.g. $Q_{\mathrm{P}}$ and $Q_{\mathrm{S}}$) of the value (reward) of performing each action. As in the Rescorla–Wagner learning updates, the change in a value is the product of a learning rate ($\alpha$) and the 'surprise', i.e. the difference between the actual perceived reward ($R$) and the currently estimated value (e.g. $Q_{\mathrm{P},t+1} = Q_{\mathrm{P}\ t} + \alpha(R_t - Q_{\mathrm{P}t})$ after performing action P in round $t$). The probability of choosing an action is a sigmoid (logistic) function of the difference in estimated values between that action and the alternative action, multiplied by a parameter $\beta$ giving the sensitivity to differences in estimated values (e.g. a sigmoid function of $\beta(Q_{\mathrm{P}} - Q_{\mathrm{S}})$; figure 1*c* illustrates action–value learning trajectories).

Actor–critic learning is a commonly used but more complex mechanism, which is related to so-called two-factor learning theory [29,30]. In this approach, the learning of values and the the updating of action preferences are coupled but separate psychological mechanisms. The expected value of a round, using the current action preferences, is updated using one learning rate (as in Rescorla–Wagner), and the action preferences, defined as the logit of the probability of choosing an action, are updated using another learning rate, but with the same value difference (the 'surprise'). We show results from using the actor–critic learning rule in the electronic supplementary material, where the details of the rule are also described.

### (a) Games

For greater generality, we study three different two-action games with negative frequency dependence. In a round of the producer–scrounger game (with a total of $T$ rounds), each group member chooses whether to produce or to scrounge. A producer has a probability $\lambda$ of finding food. On finding food, the producer consumes an amount of value $V_1$, after which scroungers can arrive, sharing the remaining amount $V_2$ with the producer. We assume that scroungers come from the producer's neighbours, but that a maximum of $\hat{n}_{\mathrm{S}}$ scroungers can participate (if there are more available, $\hat{n}_{\mathrm{S}}$ are randomly selected).

The caller–satellite game describes a group of males that can either call (C) to attract females, or to act as satellite (S) to neighbouring callers. They choose the action to use in each of a number $T$ of rounds. Each caller has an effective call strength $s$. Because of interference (e.g. aggression) between callers, the call strength decreases with the number of neighbouring callers ($s = 1 - \gamma_0 k_{\mathrm{C}}/k$, where $k$ is the number of neighbours and $k_{\mathrm{C}}$ is the number of these that call). The total number of females that are attracted to a group is proportional to the sum of the call strengths (with $f$ the constant of proportionality). An attracted female approaches one of the callers with probability proportional to his call strength. If there are no satellites, the female mates with the caller, if there is a single satellite they each have a chance of 0.5 of mating, and if there are $k_{\mathrm{S}}$ satellite neighbours of the caller, each satellite has a probability $0.5/k_{\mathrm{S}}$ of mating. This gives the

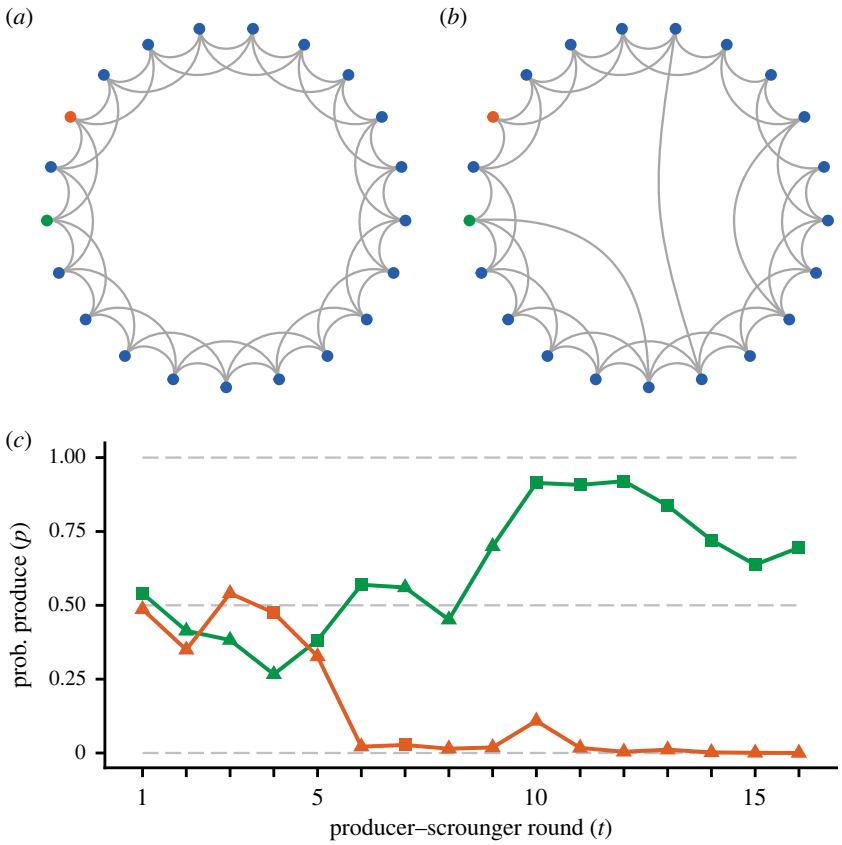


**Figure 1.** Illustration of networks with social interactions. (*a*,*b*) Coloured points represent individuals in a group and grey lines connect neighbours. Neighbours have interactions, implemented as games of a specified kind, such as producer–scrounger, caller–satellite or hawk–dove. Groups consist of 21 individuals ($N = 21$), presented as points along the perimeter of a circle. Each individual in (*a*) is connected to two neighbours in the clockwise and two in the counterclockwise direction, so each has four neighbours ($K = 4$). In graph theory such a network is called a regular ring lattice. (*b*) Shows a so-called small-world network, obtained from the one in (*a*) through a 'rewiring procedure', as described by Watts & Strogatz [26]. The probability of rewiring a connection is $p_{\text{rew}} = 0.1$. (*c*) Illustration of the probabilities to produce and the actions taken (squares denote produce and triangles scrounge) for two individuals, shown colour coded, in the network in (*a*). (Online version in colour.)

caller an advantage in mating with the female. The reason can be that the female is trying to locate the caller and, possibly, that satellites interfere with each other when trying to intercept the female. The reward for mating is $V_1$.

For the hawk–dove game, we assume that each group member has an expected number $T$ of rounds (contests). Contestants are selected by first choosing a random group member and then a random opponent among the neighbours. Each contest is a standard hawk–dove game, with a benefit (reward) $V$ of winning and a cost (penalty) $C$ of losing a hawk–hawk fight. Details of this and the other games are found in the electronic supplementary material.

## (b) Learning simulations

Our results are based on individual-based simulations of learning in groups, typically 500 groups per case. As parameters we used $V_1 = 1$, $V_2 = 3$ and $\hat{n}_S = 2$ for the producer–scrounger game; $\gamma_0 = 0.75$, $f = 2$ and $V_1 = 2$ for the caller–satellite game; and $V = 1$ and $C = 2$ for the hawk–dove game.

For action–value learning, we used $\alpha = 0.1$ and $\alpha = 0.01$ as learning rates for fast and slow learning, and $\beta = 8$ as the sensitivity to the difference in estimated values in the probability of choosing an action.

## (c) Description of polarization

We describe the degree of polarization of the individual learned action probabilities $p$ in a group using a polarization index, $F = (\text{Var}(p)/(\bar{p}(1-\bar{p}))$. The index is a normalized variance of the individual probabilities $p$. It is inspired by Wright's fixation index as used in population genetics [31]. If all group members have the same $p$, $F = 0$, and if the probabilities are either 0 or 1, but vary between individuals, $F = 1$. With several groups, we average the index over groups.

To describe individual consistency over time, we use an autocorrelation, implemented as the correlation between the individual values of logit($p$) at two points in time, as a function of the time difference (i.e. the time lag). This corresponds to the general approach of using a correlation of behaviour at two points in time to measure behavioural consistency [3,4].

# 3. Results

The types of networks and learning processes we model are illustrated in figure 1. With these kinds of social networks, but for a larger group size ($N = 99$), we simulated action–value learning for the producer–scrounger game (figure 2). For fast learning we find that substantial polarization into producers (P) and scroungers (S) emerges fairly rapidly, in particular for a small number of neighbours (figure 2*a*,*b*,*d*). For slow learning it takes longer for polarization to develop, but with a small number of neighbours, effects of frequency-dependence are strong, and polarization eventually reaches approximately the same level as for fast learning (figure 2*d* shows the first 1000 rounds). By contrast, with many neighbours and with all members connected, slow learning leads

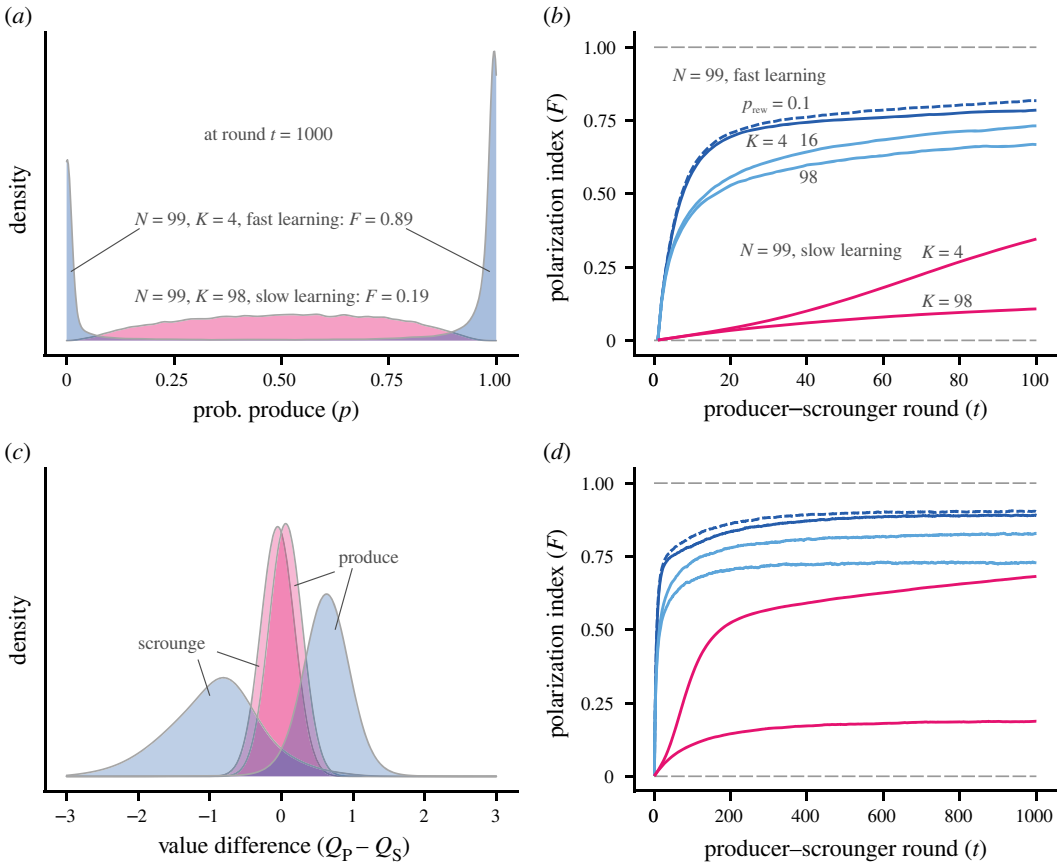


**Figure 2.** Behavioural polarization when there is action–value learning in a producer–scrounger game. Data are from 500 simulated groups per case and each group has $N = 99$ members. (*a*) Distributions of the probability $p$ to act as a producer after $t = 1000$ rounds of learning. Blue indicates a case where learning is fast ($\alpha = 0.10$) and each individual is connected to $K = 4$ neighbours. Red is a case where learning is slow ($\alpha = 0.01$) and all group members are connected ($K = 98$). The values of the group-mean polarization index $F$ at $t = 1000$ for the two cases are indicated. (*b*) Change over time of the group-mean polarization index $F$ for a number of cases. Blue curves show cases with fast learning ($\alpha = 0.10$) and red cases with slow learning ($\alpha = 0.01$), each labelled with the value of $K$. The dashed dark-blue line shows polarization in a small-world network obtained through rewiring ($p_{rew} = 0.1$) from the network illustrated by the dark-blue solid line, with $K = 4$. (*c*) Distributions of the difference between the estimated values of producing ($Q_P$) and scrounging ($Q_S$) after $t = 1000$ rounds of learning, for the two cases in (*a*). The distributions are split according to an individual's most recent action, scrounge or produce. (*d*) Same as (*b*) but over a greater number of rounds of learning. (Online version in colour.)

to a steady-state polarization with rather low value of the index $F$ that we use to measure polarization ($0 \leq F \leq 1$; figure 2*a*,*d*). The explanation is that slow learning and many neighbours give rise to distributions of the difference in estimated values that overlap between group members that used P and S in the final round (reddish distributions in figure 2*c*), because learning averages long histories of nearly identical reward distributions. With fast learning, the estimated values represent learning over a smaller number of previous rounds, giving rise to distinct estimated value distributions between group members that used P and S in a given round.

The distributions of the difference $Q_P - Q_S$ in figure 2*c* are split up according to the current action (P or S) used by an individual, and illustrate polarization. Thus, for $K = 4$ and fast learning (blue), the distributions for current producers and scroungers are separated, corresponding to strong polarization, whereas for $K = 98$ and slow learning (red) they are largely overlapping, corresponding to weak polarization.

Results for the caller–satellite game (figure 3*a*) and the hawk–dove game (figure 3*b*) were qualitatively similar to the producer–scrounger game, with rapid polarization for fast learning and a small number of neighbours. Small-world networks produced similar, and sometimes somewhat higher, polarization compared to the ring lattice they were constructed from ($p_{rew} = 0.1$;figures 2*b*,*d* and 3*a*,*b*).

Even though fast learning can give rise to pronounced polarization with many neighbours, the number of neighbours still has an important effect on individual behavioural consistency, as illustrated in figure 4*a*. We found higher temporal autocorrelation with smaller number of neighbours, for time lags of up to a few hundred rounds, for all three games (figure 4*b*,*c*,*d*). Our understanding is that this is caused by consistent differences between individuals in the expected rewards of actions, because of stronger effects of frequency dependence, in a similar way as was found by McNamara *et al.* [7] for smaller groups.

We repeated the learning simulations shown in figures 2–4 with actor–critic learning instead of action–value learning, and the results are shown in electronic supplementary material, figures S1–S3. Actor–critic learning shows some similarity to action–value learning in producing a somewhat faster build-up of polarization with a smaller number of neighbours in a social network. There is also a qualitative difference in that, after many rounds, actor–critic learning gives rise to extreme polarization, with very high consistency over time (electronic supplementary material, figure S3). Thus, after many rounds of actor–critic learning individuals develop strong action

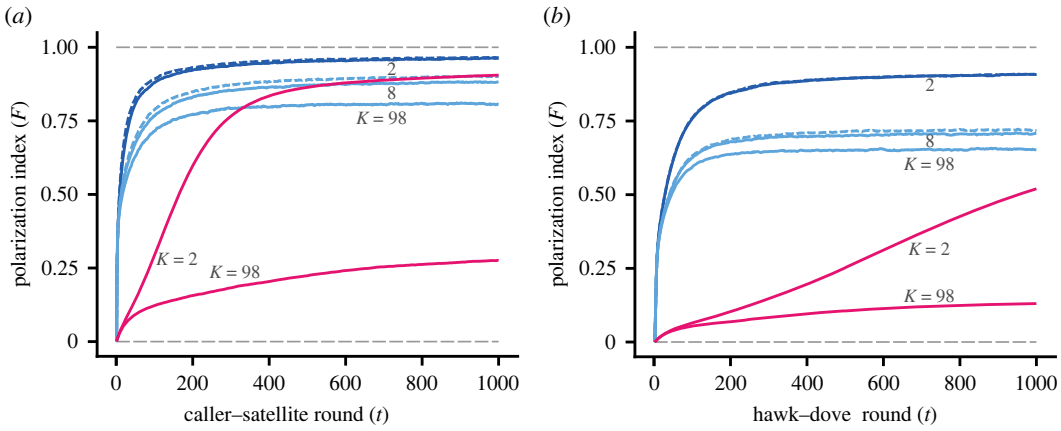


**Figure 3.** Behavioural polarization with action–value learning in caller–satellite and hawk–dove games. Data are from 500 simulated groups per case and each group has $N = 99$ members. (*a*) Change over time of the group-mean polarization index $F$ for a number of cases of a caller–satellite game. Blue curves show cases with fast learning ($\alpha = 0.10$) and red cases with slow learning ($\alpha = 0.01$), each labelled with the number of neighbours $K$. The dashed lines shows polarization in small-world networks obtained through rewiring ($p_{\text{rew}} = 0.1$) from the networks with $K = 2$ and $K = 8$, respectively. (*b*) Same as (*a*) but for a hawk–dove game. (Online version in colour.)

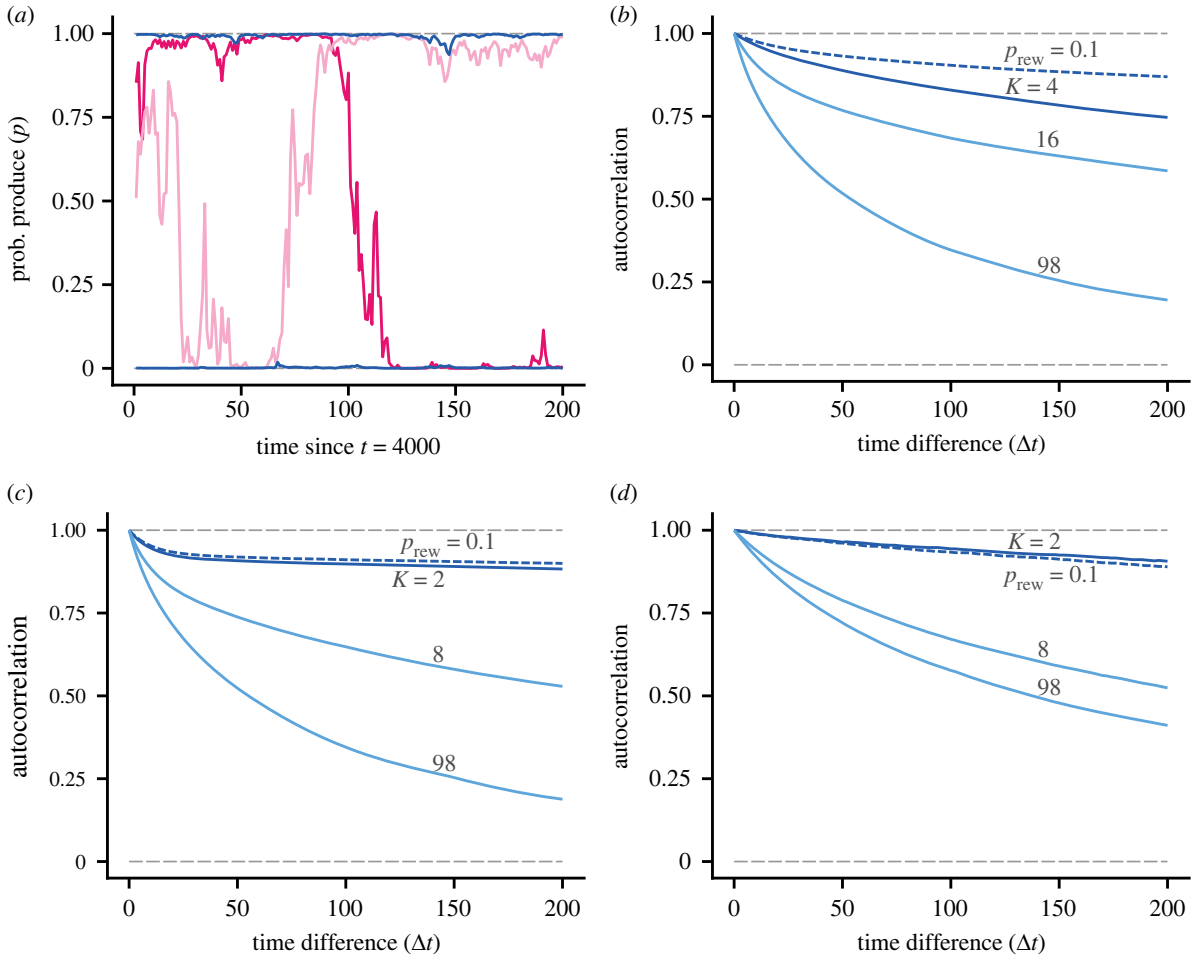


**Figure 4.** Illustration of behavioural consistency for different cases of social networks and games. Consistency tends to be higher in social networks with fewer neighbours. The group size is $N = 99$ for all cases. (*a*) Four examples of the individual probability $p$ to act as a producer. The dark-blue curves show two examples with $K = 4$ neighbours, and the reddish curves show examples with $K = 98$ neighbours. In order to illustrate steady-state situations, the curves start at round $t = 4000$. (*b*) Autocorrelation for the logit of the probability to act as producer, for the fast-learning cases in figure 2*b*,*d* and using the same colour coding. In order to illustrate steady-state situations, the autocorrelations were computed from rounds between $t = 4000$ and $t = 5000$. (*c*) Autocorrelation for the logit of the probability to act as caller, for the fast-learning cases in figure 3*a*. (*d*) Autocorrelation for the logit of the probability to act as hawk, for the fast-learning cases in figure 3*b*. The autocorrelations are estimated from simulated individuals in five groups. (Online version in colour.)

preferences, which limit their exploration of actions. This could be an unrealistic aspect of actor–critic learning, because reversal learning studies indicate that the algorithm takes longer to learn a reversal than is found in experiments [32]. A conclusion from a recent review [33] of the applicability of reinforcement learning algorithms, including action–value

learning and actor–critic learning, is that both these have some support from neuroscience, but that more work is needed to develop a better understanding of reinforcement learning as implemented in real neural systems.

As a check of the robustness of our results, we simulated learning for the producer–scrounger game over a greater number of rounds and for a greater group size (electronic supplementary material, figure S4). Finally, similar distributions as in figure 2*c* but for the caller–satellite and hawk–dove games are shown in electronic supplementary material, figure S5.

## 4. Discussion

The general idea of behavioural specialization from frequency dependence [4,34], and in particular from frequency-dependent learning [8], forms the basis of our modelling approach. Experimental observations are consistent with such specialization through learning [11]. It is also experimentally established that learning can result in behavioural diversity rather than in uniformity and conformity [10]. These studies further show that learning to specialize happens after a fairly limited number of foraging events per individual, roughly corresponding to our model assumptions of fast learning.

The traditional approach in game theory in biology is to examine genetically determined strategies [6]. In small groups, the fact that an individual never encounters itself (in pairwise interactions) can influence whether a mixed ESS or a polymorphism of pure strategies is the expected outcome [35–37]. There are similar effects for learning in small groups. With negative frequency dependence, an individual's preference for an action can cause others to learn to prefer a different action, and vice versa, and this is an explanation for behavioural specialization [7].

Theoretical analyses of learning in games, both in economics [38] and biology [24], tend to focus on the endpoints of learning, reached after many rounds of interaction. This allows investigation of correspondences between learning outcomes and game equilibria, such as ESSs, but it is important to consider possible limitations of the approach. In reality individuals might need to learn rather quickly, so that the consequences of learning after a fairly small number of rounds is the thing that matters. This should favour high rates of learning. As our results here illustrate, the rate of learning can have a qualitative influence on behavioural specialization (see also sections 5.2–5.5 in [19] for a discussion of effects of learning rates and the number of rounds). Recent experimental work in neuroscience further illustrates that learning is a complex process where individuals can adjust their learning rate, depending on how changeable the environment is likely to be [39].

Concerning social networks, there are observations on foraging in wild great tits (*Parus major*) indicating that individuals associate with a limited number of other birds [40]. For bats there are more detailed field observations of the number of producer–scrounger network neighbours [13], with individuals typically having only a handful of other group members that they predominantly interact with. There is also evidence that individuals show consistency over time in producer–scrounger relationships [12,41,42].

Our models assume that individuals do not differ in their inherent tendencies to prefer or learn about behavioural options. The reason for the assumption is to focus specifically on frequency-dependent learning, but it is likely to be an oversimplification of real situations. For instance, producer–scrounger studies have found that producing can correlate with better performance in a learning task [43], or that there are sex-differences in the tendency to produce [13,41,42]. It is even possible that consistency in the order in which individuals engage in an activity can influence which action they specialize on Dubois *et al.* [44]. Still, experiments show that individuals can change specialization in a new social environment [12].

Less is known about frequency-dependent learning of caller–satellite specialization in the field. Observations indicate that males use calls to assess the size or strength of neighbouring males in anurans and that this influences their behaviour [20,21]. There is thus the possibility that learning about the social environment plays a role in behavioural specialization, and it is also likely that variation in individual characteristics has a considerable influence on which behaviour is learnt.

As mentioned, our hawk–dove model could be a simple starting point for modelling of social dominance in small groups of individuals with limited individual recognition. This might be the case for males in some species of crickets [22,23,45] but, again, individual characteristics relating to fighting ability are likely to be important in these situations.

In conclusion, our results show that frequency-dependent learning can give rise to behavioural specialization in a social network. We have identified the number of network neighbours and the rate of learning as potentially important for the speed at which specialization emerges in a group, and possibly also for the strength of polarization and the consistency of behaviour over time. Further experimental work investigating these aspect would improve our understanding of the factors behind behavioural specialization.

Data accessibility. C++ source code for the individual-based simulations is available at GitHub, together with instructions for compilation on a Linux operating system: https://github.com/oleimar/behavspec.

Electronic supplementary material is available online [46].

Authors' contributions. O.L.: conceptualization, formal analysis, funding acquisition, software, writing—original draft, writing—review and editing; S.R.X.D.: conceptualization, writing—review and editing; A.I.H.: conceptualization, writing—review and editing; J.M.M.: conceptualization, writing—review and editing. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. The authors declare no competing interests.

Funding. This work was supported by a grant (2018-03772) from the Swedish Research Council to O.L.

Acknowledgements. We thank Redouan Bshary, Slimane Dridi and Paul Smaldino for helpful comments.

## References

1. Sih A, Bell A, Johnson JC. 2004 Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol.* **19**, 372–378. (doi:10.1016/j.tree.2004.04.009)

2. Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. 2007 Integrating animal

temperament within ecology and evolution. *Biol. Rev.* **82**, 291–318. (doi:10.1111/j.1469-185X.2007.00010.x)

3. Bell AM, Hankison SJ, Laskowski KL. 2009 The repeatability of behaviour: a meta-analysis. *Anim. Behav.* **77**, 771–783. (doi:10.1016/j.anbehav.2008.12.022)
4. Dall SRX, Houston AI, McNamara JM. 2004 The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol. Lett.* **7**, 734–739. (doi:10.1111/j.1461-0248.2004.00618.x)
5. Maynard Smith J, Parker GA. 1976 The logic of asymmetric animal contests. *Anim. Behav.* **24**, 159–175. (doi:10.1016/S0003-3472(76)80110-8)
6. Maynard Smith J 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
7. McNamara JM, Houston AI, Leimar O. 2021 Learning, exploitation and bias in games. *PLoS ONE* **16**, e0246588. (doi:10.1371/journal.pone.0246588)
8. Giraldeau L-A. 1984 Group foraging: the skill pool effect and frequency-dependent learning. *Am. Nat.* **124**, 72–79. (doi:10.1086/284252)
9. Barnard CJ, Sibly RM. 1981 Producers and scroungers: a general model and its application to captive flocks of house sparrows. *Anim. Behav.* **29**, 543–550. (doi:10.1016/S0003-3472(81)80117-0)
10. Aljadeff N, Giraldeau L-A, Lotem A. 2020 Competitive advantage of rare behaviours induces adaptive diversity rather than social conformity in skill learning. *Proc. R. Soc. B* **287**, 20201259. (doi:10.1098/rspb.2020.1259)
11. Morand-Ferron J, Giraldeau L-A. 2010 Learning behaviorally stable solutions to producer–scrounger games. *Behav. Ecol.* **21**, 343–348. (doi:10.1093/beheco/arp195)
12. Morand-Ferron J, Wu G-M, Giraldeau L-A. 2011 Persistent individual differences in tactic use in a producer–scrounger game are group dependent. *Anim. Behav.* **82**, 811–816. (doi:10.1016/j.anbehav.2011.07.014)
13. Harten L, Matalon Y, Galli N, Navon H, Dor R, Yovel Y. 2018 Persistent producer-scrounger relationships in bats. *Sci. Adv.* **4**, e1603293. (doi:10.1126/sciadv.1603293)
14. Giraldeau L-A, Caraco T 2000 *Social foraging theory*. Princeton, NJ: Princeton University Press.
15. Afshar M, Giraldeau L-A. 2014 A unified modelling approach for producer–scrounger games in complex ecological conditions. *Anim. Behav.* **96**, 167–176. (doi:10.1016/j.anbehav.2014.07.022)
16. Lucas JR, Howard RD. 1995 On alternative reproductive tactics in anurans: dynamic games with density and frequency dependence. *Am. Nat.* **146**, 365–397. (doi:10.1086/285805)
17. Lucas JR, Howard RD, Palmer JG. 1996 Callers and satellites: chorus behaviour in anurans as a stochastic dynamic game. *Anim. Behav.* **51**, 501–518. (doi:10.1006/anbe.1996.0056)
18. McCauley SJ, Bouchard SS, Farina BJ, Isvaran K, Quader S, Wood DW, St. Mary CM. 2000 Energetic dynamics and anuran breeding phenology: insights from a dynamic game. *Behav. Ecol.* **11**, 429–436. (doi:10.1093/beheco/11.4.429)
19. McNamara JM, Leimar O 2020 *Game theory in biology: concepts and frontiers*. Oxford, UK: Oxford University Press.
20. Arak A. 1983 Sexual selection by male–male competition in natterjack toad choruses. *Nature* **306**, 261–262. (doi:10.1038/306261a0)
21. Arak A. 1988 Callers and satellites in the natterjack toad: evolutionarily stable decision rules. *Anim. Behav.* **36**, 416–432. (doi:10.1016/S0003-3472(88)80012-5)
22. Alexander RD. 1961 Aggressiveness, territoriality, and sexual behavior in field crickets (Orthoptera: Gryllidae). *Behaviour* **17**, 130–223. (doi:10.1163/156853961X00042)
23. Iwasaki M, Delago A, Nishino H, Aonuma H. 2006 Effects of previous experience on the agonistic behaviour of male crickets, *Gryllus bimaculatus*. *Zool. Sci.* **23**, 863–872. (doi:10.2108/zsj.23.863)
24. Dridi S, Lehmann L. 2014 On learning dynamics underlying the evolution of learning rules. *Theor. Popul. Biol.* **91**, 20–36. (doi:10.1016/j.tpb.2013.09.003)
25. Sutton RS, Barto AG 2018 *Reinforcement learning: an introduction second edition*. Cambridge, MA: MIT Press.
26. Watts DJ, Strogatz SH. 1998 Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)
27. Albert R, Barabási A-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)
28. Rescorla RA, Wagner AR. 1972 A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: current research and theory* (eds AH Black, WF Prokasy), pp. 64–99. New York, NY: Appleton-Century-Crofts.
29. Mowrer OH. 1951 Two-factor learning theory: summary and comment. *Psychol. Rev.* **58**, 350–354. (doi:10.1037/h0058956)
30. Maia TV. 2010 Two-factor theory, the actor-critic model, and conditioned avoidance. *Learn. Behav.* **38**, 50–67. (doi:10.3758/LB.38.1.50)
31. Hartl DL, Clark AG 2007 *Principles of population genetics*, 4th edn. Sunderland, MA: Sinauer Associates.
32. Lloyd K, Becker N, Jones MW, Bogacz R. 2012 Learning to use working memory: a reinforcement learning gating model of rule acquisition in rats. *Front. Comput. Neurosci.* **6**, 87. (doi:10.3389/fncom.2012.00087)
33. Averbeck B, O'Doherty JP. 2022 Reinforcement-learning in fronto-striatal circuits. *Neuropsychopharmacology* **47**, 147–162. (doi:10.1038/s41386-021-01108-0)
34. Bergmüller R, Taborsky M. 2010 Animal personality due to social niche specialisation. *Trends Ecol. Evol.* **25**, 504–511. (doi:10.1016/j.tree.2010.06.012)
35. Vickery WL. 1988 How to cheat against a simple mixed strategy ESS *J. Theor. Biol.* **127**, 133–139. (doi:10.1016/S0022-5193(87)80124-8)
36. Maynard Smith J. 1988 Can a mixed strategy be stable in a finite population?. *J. Theor. Biol.* **130**, 247–251. (doi:10.1016/S0022-5193(88)80100-0)
37. Vickery WL. 1988 Reply to Maynard Smith. *J. Theor. Biol.* **132**, 375–378. (doi:10.1016/S0022-5193(88)80222-4)
38. Fudenberg D, Levine DK 1998 *The theory of learning in games*. Cambridge, MA: MIT Press.
39. Grossman CD, Bari BA, Cohen JY. 2022 Serotonin neurons modulate learning rate through uncertainty. *Curr. Biol.* **32**, 586–599. (doi:10.1016/j.cub.2021.12.006)
40. Aplin LM, Farine DR, Morand-Ferron J, Cole EF, Cockburn A, Sheldon BC. 2013 Individual personalities predict social behaviour in wild networks of great tits (*Parus major*). *Ecol. Lett.* **16**, 1365–1372. (doi:10.1111/ele.12181)
41. Aplin LM, Morand-Ferron J. 2017 Stable producer–scrounger dynamics in wild birds: sociability and learning speed covary with scrounging behaviour. *Proc. R. Soc. B* **284**, 20162872. (doi:10.1098/rspb.2016.2872)
42. Evans AW, Williams DM, Blumstein DT. 2021 Producer–scrounger relationships in yellow-bellied marmots. *Anim. Behav.* **172**, 1–7. (doi:10.1016/j.anbehav.2020.11.018)
43. Katsnelson E, Motro U, Feldman MW, Lotem A. 2011 Individual-learning ability predicts social-foraging strategy in house sparrows. *Proc. R. Soc. B* **278**, 582–589. (doi:10.1098/rspb.2010.1151)
44. Dubois F, Giraldeau L-A, Réale D. 2012 Frequency-dependent payoffs and sequential decision-making favour consistent tactic use. *Proc. R. Soc. B* **279**, 1977–1985. (doi:10.1098/rspb.2011.2342)
45. Ashikaga M, Sakura M, Kikuchi M, Hiraguchi T, Chiba R, Aonuma H, Ota J. 2009 Establishment of social status without individual discrimination in the cricket. *Adv. Rob.* **23**, 563–578. (doi:10.1163/156855309X420066)
46. Leimar O, Dall SRX, Houston AI, McNamara JM. 2022 Behavioural specialization and learning in social networks. Figshare. (doi:10.6084/m9.figshare.c.6124001)