Evolution of learned strategy choice in a frequency-dependent game

Edith Katsnelson[1,*], Uzi Motro[2], Marcus W. Feldman[3] and Arnon Lotem[1]

[1]*Department of Zoology, Tel-Aviv University, Tel-Aviv 69978, Israel*
[2]*Department of Ecology, Evolution and Behavior, Department of Statistics, and The Center for Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel*
[3]*Department of Biology, Stanford University, Stanford, CA 94305, USA*

In frequency-dependent games, strategy choice may be innate or learned. While experimental evidence in the producer–scrounger game suggests that learned strategy choice may be common, a recent theoretical analysis demonstrated that learning by only some individuals prevents learning from evolving in others. Here, however, we model learning explicitly, and demonstrate that learning can easily evolve in the whole population. We used an agent-based evolutionary simulation of the producer–scrounger game to test the success of two general learning rules for strategy choice. We found that learning was eventually acquired by all individuals under a sufficient degree of environmental fluctuation, and when players were phenotypically asymmetric. In the absence of sufficient environmental change or phenotypic asymmetries, the correct target for learning seems to be confounded by game dynamics, and innate strategy choice is likely to be fixed in the population. The results demonstrate that under biologically plausible conditions, learning can easily evolve in the whole population and that phenotypic asymmetry is important for the evolution of learned strategy choice, especially in a stable or mildly changing environment.

Keywords: social foraging; producer–scrounger game; phenotypic asymmetries; evolutionary simulation

1. INTRODUCTION

Group-living animals are often involved in frequency-dependent population games in which the payoff from a behaviour depends on its frequency in the population. Such games underlie the concept of an evolutionarily stable strategy (ESS) [1,2]. While pure or mixed ESSs were initially analysed as genetic, various studies have suggested that strategy choice may also develop through a learning process that can result in the population approaching the theoretical ESS [3–7].

It is well recognized that learning may be beneficial when it allows animals to refine their behaviour in relation to a changing environment [8,9] and ecological or phenotypic conditions that might not have been important in their genetic evolution [10]. In these cases, learning enables prediction of the behaviours that are likely to be associated with higher payoffs. However, learning to choose among strategies in frequency-dependent games is much more complex because the payoff from each alternative is determined not only by current environmental conditions, but also by the choices made by all other members of the population [5,11]. Moreover, dynamic fluctuations in the population's strategy profile are likely to occur in finite populations, causing players to experience variation in payoffs even when the external environment is stable [12–15]. Such stochastic deviations

from the ESS are especially likely in small populations and when populations are subdivided or change in size.

Learned strategy choice in frequency-dependent games has been demonstrated in several systems [16,17], and there is some recent experimental evidence for its existence in the producer–scrounger game [18–20]. In this game, producers actively search for resources, whereas scroungers take advantage of resources discovered by others [21,22]. In a recent theoretical analysis, Dubois *et al.* [23] used an analytical model to test whether a learner mutant that can optimize its producer–scrounger behaviour can invade a population in which two types of non-learners, that play producer or scrounger with fixed probabilities, are maintained in a frequency-dependent equilibrium. They showed that this kind of learning mutant can invade the population of non-learners but it does not evolve to fixation, and the majority of individuals in the group remain non-learners. According to Dubois *et al.* [23], learners can initially invade the population as they perfectly 'buffer' the producer–scrounger bias created between the non-learners. Subsequently, additional learners no longer have an advantage over non-learners, and a polymorphism of learners and non-learners is maintained, in which the proportion of learners may not exceed 40 per cent. The prediction that follows from Dubois *et al.*'s model is that a population in which all individuals are able to learn to choose among strategies is unlikely to evolve. These interesting results emphasize the non-trivial nature of learning within a frequency-dependent game and are consistent with recent models of behavioural polymorphism [24]. It is important to note, however, that in Dubois *et al.* the

learning process was instantaneous. Learners did not have to actually learn from trial and error. Instead, learners' strategy was determined according to the predicted mean gain expected from each strategy under the current conditions in a population of non-learners (i.e. learners immediately acted according to the correct solution for a learner in a population of non-learners, possibly with some error).

Here, we explore the evolution of learning using a different approach. To capture the dynamic learning process within a frequency-dependent game, we use an agent-based evolutionary simulation in which we explicitly model the learning process of each learner. Accordingly, learners have to sample their environment (repeatedly, and simultaneously with other learners) and their behaviour may not necessarily be optimal at any given moment [25–27].

Learned strategy choice in the producer–scrounger game was initially studied in a computer simulation by Beauchamp [4], who showed that several learning rules can generate the expected game equilibrium when played by all individuals in the population (see also Hamblin & Giraldeau [5]). However, the fact that learning rules for strategy choice can generate a game equilibrium does not fully explain their evolution, nor does it indicate how such rules compare to the alternative mechanism of innate strategy choice. We study the conditions under which learned strategy choice can invade a finite population of non-learners (playing either pure or mixed strategies) that is already in equilibrium. The advantage of learning in this case cannot be due to the population not being in equilibrium but must be due to real advantage over innate strategies. We account for three possible sources of variation that may favour learning. First, we consider a stable environment in which, as mentioned already, variation in payoffs resulting from dynamic fluctuations in the population's strategy profile may promote the evolution of learning. The second source of variation is environmental change that may lead to changes in the game's payoff matrix, thus altering the ESS solution [2,5]. Finally, the third source of variation in the payoff to each strategy is phenotypic asymmetry among players. Phenotypic differences between individuals are potentially diverse in their origin; possible examples might include variation in cognitive skills, in physical abilities or in any other trait that causes players to vary in their ability or state. Thus, it is reasonable to expect that different individuals should experience different costs or benefits when selecting each strategy [28–32]. In theory, the adaptive value of learned strategy choice would be derived from responding correctly to one or more of these sources of variation in payoffs. We tested two different learning rules under four conditions that differ in the source of payoff variation, thus illustrating the factors that can limit or promote the evolution of learned strategy choice in frequency-dependent games.

2. THE MODEL
(a) *The basic model*
Learning is a dynamic process involving stochastic sampling errors that influence subsequent sampling steps, and eventually produce a wide distribution of possible outcomes. Using an agent-based simulation programmed with MATLAB (v. R2006a), we explicitly model this process for each individual in the group (see Ruxton & Beauchamp [33] for elaboration on the rational and application of this approach). We first construct a basic evolutionary simulation with learning absent, using a population of haploid organisms with one bi-allelic gene, hereafter called 'the foraging strategy gene'. This gene has alleles F_1, whose carriers use a pure producer–foraging strategy, and F_2, whose carriers' foraging strategy is pure scrounging (mixed strategies and learned strategy choice will be considered later). For the population size, we chose a naturally plausible group of seed-eating birds: $n = 60$. We were also able to reproduce the main results with a population size of $n = 300$, as part of another study (M. Arbilly 2009, unpublished data). The foraging environment in our model is patchy, and includes food patches containing 20 food items, and single indivisible food items. To avoid the complex dynamics of patch depletion, we assume an infinite environment, where there is no depletion and a patch is not visited more than once during players' lifetimes. The individuals perform 100 foraging steps at each of which they act according to the allele at their foraging–strategy locus, i.e. individuals play a pure strategy game (later we also consider a mixed strategy). At each step, an individual that plays 'producer' can find a food patch with probability p, or a single-food item with probability $1 − p$ (these probabilities reflect the relative abundance of each food source in the environment). A scrounger joins a randomly chosen producer that has found a food patch (assumed to be identifiable owing to its feeding behaviour, hence, scroungers only join producers that found a food patch and not those that found a single-food item). A producer that finds a food patch monopolizes a fixed part of the patch, an assumption known as the 'finder's advantage' [22], which we set to four food items. The remaining 16 food items are shared evenly among all individuals at the patch (the single producer that found the patch and the scroungers that joined it; thus, a producer might share a patch with one scrounger, several or none). For a sufficiently small p (but not too small), this modelling framework creates a negatively frequency-dependent effect that is consistent with the basic description of the producer–scrounger model by Barnard & Sibly [21] (see electronic supplementary material, appendix A and figure SA1 therein). Foraging steps in the game are played simultaneously, first by all producers, and then by all scroungers (assigning them to producers that found patches). The total number of food items accumulated by each individual over the 100 steps of the game determines its fitness.

At the end of each generation (100 steps of the game), selection is performed by taking the 50 per cent of the population with the highest fitness score (i.e. truncated selection) and allowing them to reproduce and create the next generation [27,33]. Although individuals are haploid (see earlier text), reproduction is sexual, in the sense that the selected 30 individuals pair randomly (making 15 pairs), and each pair creates four offspring from their combined gene pool; hence, the population size is kept constant with 60 individuals. When more than one gene is considered (i.e. when learning genes are included in pure strategy game, see later text), genes are assumed to be unlinked (as if they are on different chromosomes). Hence, if the parents carry different alleles, then each offspring will randomly carry one of them for each locus.

To allow producer–scrounger equilibrium formation in the population, we ran the simulations for 20 generations before allowing mutation. From this point onwards, bi-directional mutation between F_1 and F_2 was allowed, with a probability of 0.01 at each generation.

For simplicity, and to avoid a possible bias that favoured one of the strategies, the value of p (producer's probability of finding a patch at each foraging step) was set to allow a stable producer–scrounger equilibrium, with similar average frequencies of producers and scroungers. Accordingly, the value $p = 0.298$ was chosen as it generated an equilibrium with the closest to equal average numbers of producers and scroungers (see electronic supplementary material, appendix A for further details).

(b) Introducing learned strategy choice

We tested two general learning rules: the linear operator (LO) rule and the random sampling (RS) rule. The evolution of each learning rule was tested separately in different sets of simulations. Learning in the model refers to an individual's ability to choose between the two optional strategies based on their own experience [4]. A 'learning gene' (alleles $L_1 \ldots L_n$) was added to the model. An individual that carried allele L_1 at the learning gene was a non-learner whose behaviour was determined by the allele carried at the foraging gene (F_1 or F_2). An individual that carried one of the learning alleles ($L_2 \ldots L_n$) was a learner and applied the learning rule with parameters of memory or sampling steps coded by each of the different learning alleles $L_2 \ldots L_n$ (see later text). A learner did not express its social foraging gene (F_1 or F_2); instead, it used the learning rule to choose between social foraging strategies. Simulations were initialized with non-learners only (all individuals begin with allele L_1 at the learning gene) to simulate invasion of learners into a population of non-learners. As in the basic model, to allow establishment of a producer–scrounger equilibrium in a population of non-learners, we ran the simulations for 20 generations before including mutations. From this point onwards, mutation at both genes was allowed in both loci, with a probability of 0.1 at each generation (we use higher mutation rate for these simulations as we have larger number of competing alleles). In the case of mutation, the mutated allele was randomly replaced with one of the other possible alleles.

(i) The conditions under which learning success was tested

To investigate what makes learning adaptive, we explored its evolution under four different conditions, reflecting variations in the environmental stability and in the phenotypic asymmetry:

Stable-symmetric. The environment is stable, i.e. patch sizes are constant (20 food items). Individuals are assumed to be equal in their ability to play each foraging strategy, and consequentially in the payoffs they receive from each tactic, given that conditions are equal (i.e. symmetric game). Note that under these conditions, the only changes that can affect the learning process are the intrinsic changes in the game's dynamics resulting from fluctuation in gene frequencies or from the number of individuals that choose to play each strategy.

Changing-symmetric. Individuals are equal in their ability to play each strategy, as mentioned already (symmetric

game), but the environment changes across generations. In each generation, we randomly increased or decreased the size of the patches by 15 per cent (i.e. from 20 to either 17 or 23 food items), while the finder's advantage remains unchanged (assuming that it represents the producer's ability to monopolize food items prior to scrounger's arrival). Changes in different generations were independent. As an environment in which patches contain 17 food items is expected to promote producing, whereas an environment in which patches contain 23 food items is expected to promote scrounging, the changing environment offers a potential benefit for learning to choose the correct strategy in each generation. A change of 15 per cent was first chosen as it allowed the persistence of the producer–scrounger game, i.e. the two pure strategies coexist (in only one out of 300 simulations of 1000 generations with no mutations did the game not persist, while using a larger environmental change of 20 per cent resulted in 27 collapses of the game out of 300). Later, we also explore the effect of lower and higher levels of environmental change.

Stable-asymmetric. The environment is stable (as in *stable-symmetric* condition), but individuals are asymmetric (i.e. they differ in their ability to succeed with each foraging strategy). This asymmetry is assumed to be non-heritable, and was assigned randomly to each individual at birth. Accordingly, each individual was assumed to have a phenotypic constraint in performing one of the strategies so that its payoff at each foraging step was reduced by 15 per cent if it used that strategy. The value of 15 per cent was chosen to correspond with the value of the environmental change in the *changing-symmetric* condition, but we also explored other levels of asymmetry. Such phenotypic asymmetries may emerge either from environmental or from developmental events, such as laying order or the amount of testosterone or nutrients deposited in a passerine egg [34,35], or as a result of multiple genetic and environmental factors that generate asymmetries (i.e. individual differences) with low genetic heritability. The effect of such asymmetries on the success of the producer and scrounger strategies may be mediated by physical or cognitive traits. For example, a strong, fast individual may perform well enough as a scrounger, but might easily miss food patches when it searches for food independently as a producer.

Changing-asymmetric. In this condition, there are environmental changes in patch size across generations, as in *changing-symmetric* condition, and in addition, individuals are phenotypically asymmetric, as in *stable-asymmetric* condition.

(ii) The learning rules

Two learning rules were tested in the simulations, representing two extreme possibilities in terms of sampling method and dynamical updating of the learning process:

The linear operator rule. This rule has frequently been used in research on foraging dynamics or decision-making [4,6,36]. It includes a mechanism that updates the value of the foraging strategies, and a mechanism to choose between the strategies (a decision rule). Let $V_{i,t}$ denote the value of foraging strategy i ($i = 1$ for producing and 2 for scrounging) at step t ($t = 1, \ldots, 100$) when using the LO rule. $V_{i,t}$ is updated according to the

individual's experience:

$$V_{i,t} = \alpha V_{i,t-1} + (1 - \alpha)P_{i,t} \quad \text{if strategy } i$$
$$\text{was played at step } t$$

and

$$V_{i,t} = V_{i,t-1} \quad \text{if strategy } i \text{ was not played at step } t,$$

where $P_{i,t}$ is the payoff experienced when using strategy i at step t, and α is a memory factor, weighting the value of strategy i at the last step in which this strategy was used with the recent payoff experienced when using this strategy at step t. We define $V_{i,0} = 1$ ($i = 1,2$). Note that the value of a strategy remains unchanged as long as this strategy is not used. As a decision rule for choosing a strategy at the next step, we used simple proportional matching, i.e. the probability of choosing each of the two possible foraging strategies was proportional to their current values. In exploring the LO rule, three possible learning alleles (L_2, L_3 and L_4) that differed in the value of the memory factor α used by the learner ($\alpha = 0.50$, 0.85 and 0.95, respectively) were randomly introduced by mutations to a population of non-learners.

Random sampling rule. This rule resembles imprinting-style learning, or conventional statistics: it involves an initial learning period of k steps during which the learner samples both strategies in a random order. At the end of this learning period, the learner chooses the strategy with the higher average payoff, and will employ this strategy in all the remaining steps of the game. When exploring the RS rule, we considered six possible learning alleles (L_2–L_7) that differed in the length of the learning period used by the learner (with $k = 3$, 5, 20, 50, 80 and 100 learning steps, respectively). As in the *LO rule*, learning alleles were introduced by mutations to the population of non-learners. Note that with $k = 100$, learners do not actually learn, but rather explore randomly throughout their lives, and their behaviour is thus equivalent to the mixed strategy of playing producer and scrounger with equal probabilities (hereafter referred to as a mixed-random strategy).

We defined fixation of alleles according to the last generation in the simulation: if all the individuals in the last generation carried the same allele, then it was referred to as fixation of this allele.

(iii) *The cost of learning*
Although learning may involve energetic or physiological costs [37], we deliberately ignored such costs in the present study to focus on the intrinsic constraints of learned strategy choice that are characteristic of the learning process itself. Our conclusions may therefore set the minimal requirements for the evolution of learned strategy choice.

3. RESULTS
(a) *The evolution of learned strategy choice in a mixed population of pure players*
We first tested the evolution of learning in a population, where non-learners (allele L_1) play according to their pure strategy allele (F_1 or F_2), and learning was introduced through mutations. We ran 100 simulations of 10 000 generations each; each simulation was initialized with equal frequencies of F_1 and F_2. The results (figure 1) show that learning prevailed under all tested conditions,
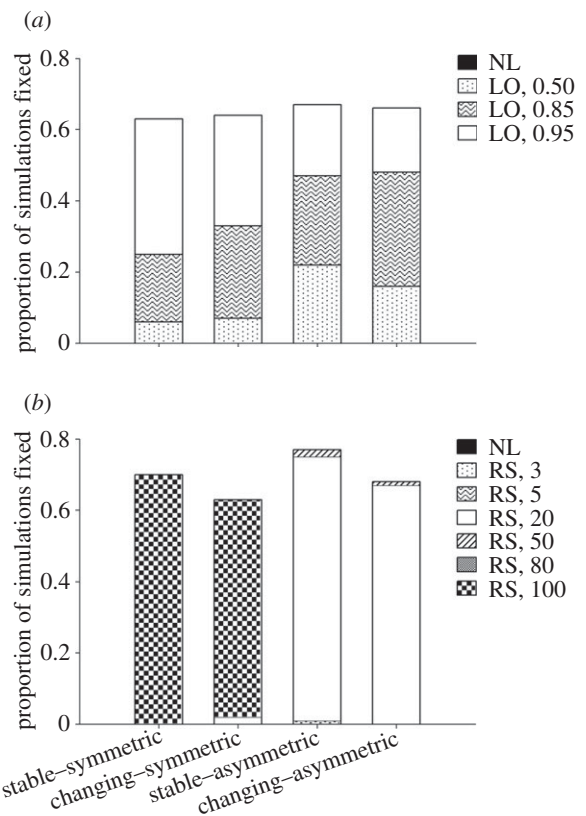

Figure 1. Proportion of runs where the different alleles reached fixation out of 100 simulations of 10 000 generations under the conditions shown, if non-learners are pure producers or scroungers. Black shade (NL) represents the non-learning allele L_1, whereas all other shades represent the different learning alleles (see legend). (*a*) Learning is based on the linear operator (LO) rule, values in legends indicating the memory factor α. (*b*) Learning is based on a random sampling (RS) rule, values in legends indicating the number of learning steps k. In both cases, the population started with pure producers and scroungers carrying the non-learning allele (i.e. genotypes F_1L_1 and F_2L_1), and the learning alleles were introduced through mutation.

with some learning alleles (with certain learning parameters) being more likely to reach fixation than others. As a result of the high mutation rate (0.1 per gene per generation), fixation did not occur in all populations. However, in this case (as well in figures 2–4), very similar results are obtained when all 100 runs are analysed based on the average frequencies of the different learning alleles during the last 1000 generations (available upon request).

A closer examination of the results suggests that the sweeping success of learning may not represent the advantage of learning *per se*, but rather the advantage of a mixed strategy over a pure one. This advantage is clearly indicated by the success of the random sampling learning rule with 100 learning steps under the *stable-symmetric* and *changing-symmetric* conditions (figure 1*b*). In this case, random exploration continues to the end of the game (with no remaining steps to apply what has been learned) and thus learning is equivalent to a mixed-random strategy (playing producer and scrounger with equal probabilities). An advantage for the mixed strategy over a pure one has been proposed in the past for finite populations [38]. The reasons for this will be explained in §4. Most critically, however, because learned strategy choice is also an individually
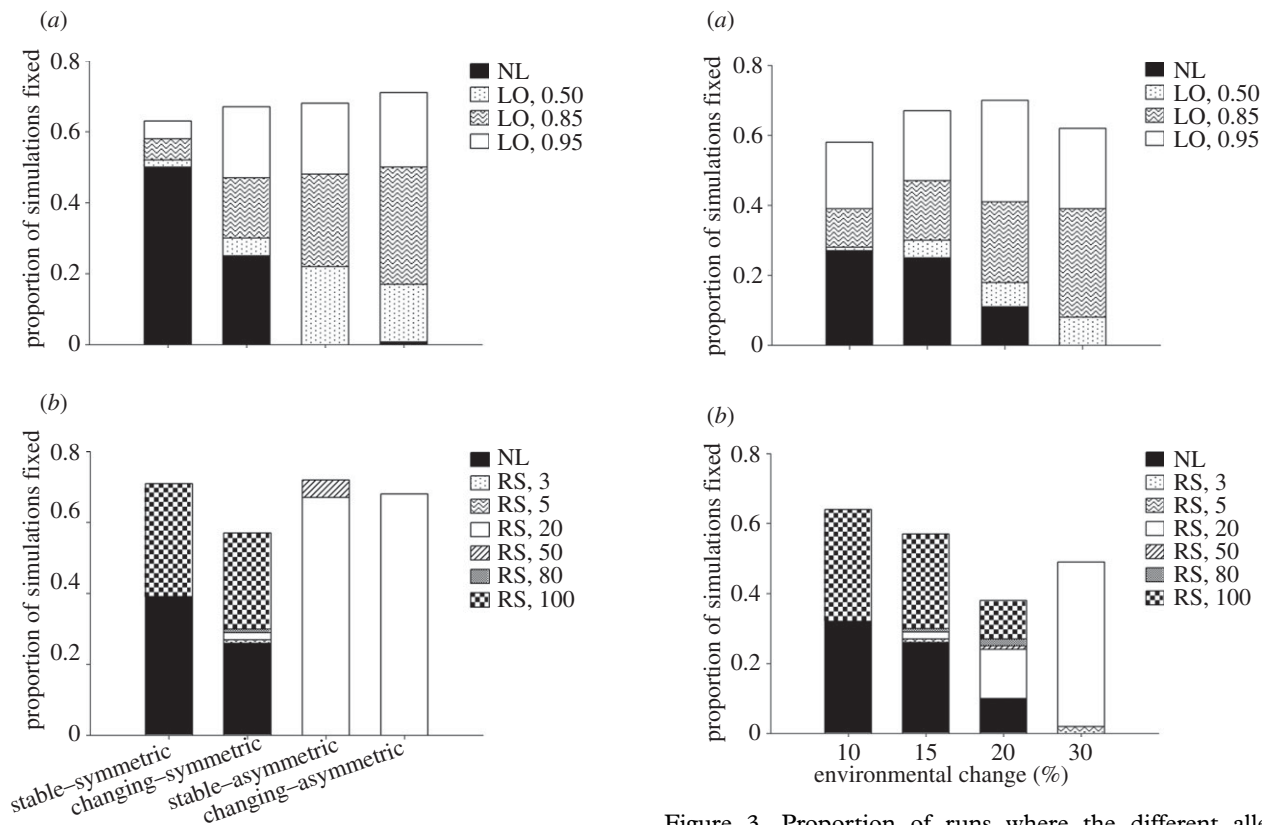
Figure 2. Proportion of runs where the different alleles reached fixation out of 100 simulations of 10 000 generations under different conditions when non-learners are playing a mixed-random producer–scrounger strategy. Black shade (NL) represents the non-learning allele L_1, whereas all other shades represent the different learning alleles (see legend). (*a*) Learning is based on the linear operator (LO) rule, values in legends indicating the memory factor α. (*b*) Learning is based on a random sampling rule (RS), values in legends indicating the number of learning steps k. In both cases, the population started with non-learners playing the mixed-random strategy (i.e. learning allele L_1), and the learning alleles were introduced through mutations.

mixed strategy (albeit with probabilities that are adjusted by experience), the advantage of learning illustrated by figure 1 may be derived in whole or in part from being mixed rather than from choosing the correct strategy. To examine whether the learning process itself has any advantage, we must therefore test the success of learning in the presence of a competing innate mixed strategy. This we do in the subsequent section.

(b) Learned strategy choice in a population of mixed-strategy players
To test whether the advantage of learning is derived from it being a better strategy choice rather than from being merely an individually mixed strategy, we tested the evolution of learning in a monomorphic population in which non-learners (that use innate strategy choice) play a mixed-strategy game. This was performed by re-running the earlier-mentioned simulations with one crucial change: this time, at each foraging step, an individual that carries the L_1 allele (i.e. a non-learner) plays either the producer or the scrounger strategy with equal probabilities (mixed-random strategy). The mixed strategy corresponds to the equilibrium values

Figure 3. Proportion of runs where the different alleles reached fixation out of 100 simulations of 10 000 generations for different values of environmental change under changing-symmetric conditions. Black shade (NL) represents the non-learning allele L_1, whereas all other shades represent the different learning alleles (see legend). (*a*) Learning is based on the linear operator (LO) rule, values in legends indicating the memory factor α. (*b*) Learning is based on the random sampling (RS) rule, values in legends indicating the number of learning steps k. In both cases, the population started with non-learners playing the mixed-random strategy (i.e. learning allele L_1), and the learning alleles were introduced through mutations.

chosen for the pure strategy game (see §2*a*). That is, at each step, it applies a 'coin flipping' strategy choice, and subsequent choices are independent. This resulted in a one-locus model in which non-learners play a mixed strategy and learners are introduced by mutations.

The results (figure 2) show that under the tested parameters, learned strategy choice was consistently successful in invading a population of mixed-strategy players only when players were phenotypically asymmetric. Under these conditions, the LO learning rule evolved successfully with all three memory factors, and the RS rule exhibited a clear advantage for a learning period of $k = 20$ steps. Note that learning reached fixation in most of these cases. These results are different from those of Dubois *et al.* [23], where learning never reached fixation and hardly exceeded 40 per cent. Learning could also evolve under the *changing-symmetric* condition when learners use the LO rule, but it was not more successful than the innate strategy (figure 2*a*). Later, we explore whether learning can become more successful under the changing-symmetric condition if the level of environmental change is greater than the 15 per cent tested here.

It is relatively easy to understand why learning failed to evolve under the *stable-symmetric* condition. In this case,
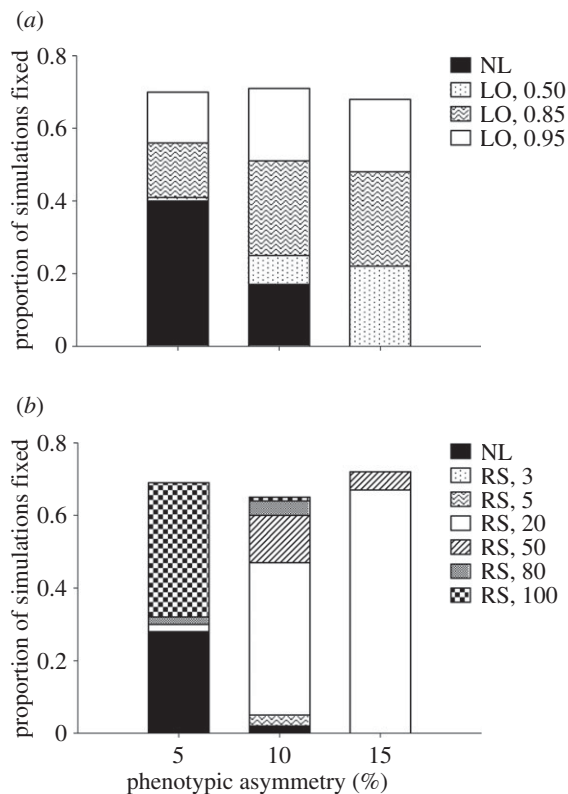
Figure 4. Proportion of runs where the different alleles reached fixation out of 100 simulations of 10 000 generations for different values of phenotypic asymmetry. Black shade (NL) represents the non-learning allele L_1, whereas all other shades represent the different learning alleles (see legend). (*a*) Learning is based on the linear operator (LO) rule, values in legends indicating the memory factor α. (*b*) Learning is based on the random sampling (RS) rule, values in legends indicating the number of learning steps k. In both cases, the population started with non-learners playing the mixed-random strategy (i.e. learning allele L_1), and the learning alleles were introduced through mutations.

the ESS is a one-to-one ratio between strategies, and mixed players apply this solution as precisely as possible. Learners, on the other hand, must learn this solution from experience, and as both learning rules involve sampling errors, they cannot converge to the mixed solution more often than the innate mixed players. Furthermore, with the RS rule, learners also deviate from the ESS when shifting from mixed exploration to a single choice. The only case under the *stable-symmetric* condition where the learning allele seems to be as successful as the non-learning allele is when it uses the RS rule with 100 steps. This is expected, however, because it is equivalent to a mixed strategy with no learning at all (see earlier text). A detailed analysis of the learning process with both learning rules is provided in the electronic supplementary material, appendix B.

(i) *Exploring higher levels of environmental change and lower levels of phenotypic asymmetries*
As mentioned earlier, we initially examined environmental change of magnitude 15 per cent, as this allows the persistence of a stable game when players use a pure strategy. Under this level of environmental change, learning had limited success and was not better than innate

strategy choice. It is expected, however, that for a sufficiently high level of environmental change, learning may become better than innate strategy choice. Therefore, we also explored learning success under the *changing-symmetric* condition where the magnitude of change was 10, 15, 20 and 30 per cent. The results (figure 3) confirm that learning is increasingly successful as the magnitude of environmental change increases.

We mentioned earlier that phenotypic asymmetry of 15 per cent was sufficient to guarantee the success of learning (figure 2). It is quite intuitive that increasing the level of asymmetry even further will improve the efficiency of learning. We were able to confirm this by running simulations with phenotypic asymmetries of 20 and 30 per cent (available upon request). It is perhaps more interesting to explore the effect of reducing phenotypic asymmetry, or in other words, to find out what is the minimal level of phenotypic asymmetry that would be sufficient to promote learning. We examined this by running evolutionary simulations under the *stable-asymmetric* condition with asymmetry levels of 5 and 10 per cent. The results (figure 4) show that reducing the asymmetry causes a decline in the evolutionary success of learning, with success in most runs under asymmetry of 10 per cent and in only a few or none under asymmetry of 5 per cent.

4. DISCUSSION

Although several studies have suggested that individuals can learn to choose between strategies in a game [3,4–6], only recently has the evolution of learned strategy choice been explored, but without modelling the learning process explicitly [23]. Here, we used an agent-based evolutionary simulation to capture the dynamic process of learning to choose between strategies in the producer–scrounger frequency-dependent game. We found that learning was unlikely to evolve in a stable or mildly changing environment. Yet, learning was superior to innate strategy choice and evolved successfully when players were phenotypically asymmetric and were able to learn which strategy they could play best, or when environmental change was sufficiently high. Moreover, under these conditions, learning usually spread in the population and reached fixation. In the following, we discuss the possible implications of our results in relation to several aspects of learning and strategy choice in games.

(a) *Learning strategy may succeed because it has the benefit of a mixed strategy*
In the first part of our analysis, learned strategy choice prevailed when competing with a mixture of pure genetic strategies (figure 1). However, this success was also achieved when learners were programed to use 100 random sampling steps, with no remaining steps to apply what has been learned. In this case, learning is precisely equivalent to an innate mixed strategy with $p = 0.5$. This advantage of mixed over pure strategies may appear to be inconsistent with the notion that the two ESS solutions are mathematically equivalent [2,39]. However, several studies have shown that in finite populations, a mixed strategy can take over a comparable mixture of pure strategies (see [12,38] for further elaboration). This invasion of mixed players can be initiated by genetic

drift. Then, as soon as mixed players spread, pure producers or scroungers will be selected against when they increase in frequency unless a coordinated invasion of both types of pure strategies occur simultaneously, which is unlikely. For the evolution of learned strategy choice, this process is crucially important because learned strategy choice is also a mixed strategy (albeit with probabilities adjusted by experience). The apparent advantage of learning over pure strategies may therefore be derived from being mixed rather than from learning. Indeed, in our subsequent analysis, we showed that the advantage of learning is much smaller, and that learned strategy choice may actually fail to evolve when it competes against an innate mixed strategy coding for the game's equilibrium (compare figures 1 and 2). Accordingly, one should always consider the possible evolution of competing individuals playing the comparable innate mixed strategy, or alternatively, why they cannot evolve. By contrasting the two strategies, we tested the benefit of learning that is solely derived from making better choices based on recent experience (i.e. owing to learning *per se*).

(b) *The constraints on learning to choose among strategies in a game*

Learning is often considered an adaptation to environmental change, but it also requires some level of environmental predictability [8,9,40]. Without it, recent experience cannot reinforce the behaviours that are likely to be adaptive in the future. When it comes to strategy choice in games, where game dynamics cannot be predicted genetically, we might intuitively expect that learned strategy choice would be useful. But our results illustrate that this may not be the case: under the *stable-symmetric* condition, the stochastic fluctuations in the number of producers and scroungers between generations (electronic supplementary material, figure SA1 in appendix A) disappear when pure players are replaced by innate mixed players. Random deviations from the equilibrium can still occur, but these are small and not consistent throughout the generations. Under these conditions, learning that inevitably involves sampling errors is less efficient than the innate mixed strategy that already plays the ESS. Thus, it seems that the mere existence of fluctuations in the population's strategy profile may not be sufficient to promote the evolution of learned strategy choice in frequency-dependent games.

Under the *changing-symmetric* condition, a change in the 'external' environment was introduced, and this allowed learning to evolve; however, its clear-cut advantage over an innate mixed strategy was not achieved without a great deal of change (30%, figure 3). When learning takes place within a dynamic game, the effect of environmental change may be overshadowed by the change in the frequency with which each strategy is played by other foragers. As a result, slight changes in the environment may not be enough to promote the evolution of learning within games, and a relatively strong environmental change may be needed.

Under both the *stable-asymmetric* and the *changing-asymmetric* conditions, individual asymmetry of 15 per cent provided learning with the degree of predictability necessary for learning to be clearly advantageous (figure 2). Decreasing the level of asymmetry to 10 per cent or below makes the

evolution of learned strategy choice less probable (figure 4). In reality, the degree of individual asymmetry may be greater than 15 per cent, and this may explain the emergence of strong individual differences in strategy choice [41].

In contrast to the earlier-mentioned results, Dubois *et al.* [23] demonstrated with an analytical model that learners can invade a population of non-learners under both stable or changing environments, but can never increase to high frequency or reach fixation (see §1). They suggested that learning in some individuals prevents learning from evolving in others (see also [42]). The apparent inconsistency between the results of our analysis and those of Dubois *et al.* [23] suggests that the evolution of learned strategy choice in frequency-dependent games is a non-trivial process that may be sensitive to different conditions or model assumptions.

The first difference between the two approaches is that in Dubois *et al.*'s model, learners are assumed to have the correct solution instantaneously (accurately or with some error) without the need to learn from trial and error, but based on calculating the optimal behaviour for a learner in a population of non-learners (see p. 3611 in Dubois *et al.* [23]). As a result, even a small fraction of learners in the population can immediately act as a 'buffer' that brings the population to equilibrium and allows the persistence of non-learners that become equally successful due to this 'buffer'. In our model, on the other hand, the learning process is simulated explicitly and learners have to sample the environment and choose their strategy based on experience. Such a learning process is noisy, takes time and can be far from optimal. This is illustrated both by our own analysis (electronic supplementary material, appendix B) as well as by previous work on learning dynamics [6,27,36,43,44]. Without sufficient predictability, as in the symmetric and changing environment conditions, recent experience cannot reinforce behaviours that are likely to be adaptive in the future.

The second difference between our model and that of Dubois *et al.* is that in their model they usually considered a population of two types of non-learners that use mixed strategy with fixed probabilities. The two types of fixed non-learners maintained frequency-dependent equilibrium while neither played the game solution (the mixed ESS). Hence, learners could act as a 'buffer', at any given time doing better than fixed producers or fixed scroungers but not better than both [23]. Our aim, on the other hand, was to test the evolution of learning when competing against non-learners that play the ESS solution. Interestingly, when Dubois *et al.* [23] tested two types of fixed individuals that played the producer–scrounger with equal probabilities, learners could not invade the population at all. This corresponds with our results under the *symmetric* conditions when learning could not invade a population of non-learners playing the mixed ESS.

Thus, the two models are actually compatible, and both illustrate that learning may not evolve easily in a population where non-learners already express the mixed ESS.

(c) *Empirical predictions and further implications*

Our model predicts that with a sufficient degree of environmental change or consistent asymmetries in rewards, learning to choose between strategies will spread in the population and reach high frequencies. In the light of

recent empirical evidence for learned strategy choice in the producer–scrounger game [19,20], it will be interesting to examine whether learning is applied by all individuals or only by some of them, as predicted by Dubois *et al.* [23]. It will also be interesting to examine whether learned strategy choice is more prevalent in a changing environment or in heterogeneous populations where individuals are phenotypically asymmetric in their ability to succeed with each strategy. Phenotypic asymmetry does not necessarily mean that individual characteristics are fixed for life (although for simplicity we modelled it this way). Individuals may also change their characteristics (such as relative strength or status compared with other group members) or states [45]. Yet, for learning to be beneficial, these characteristics need to be consistent over a sufficient length of time that first enables the individual to learn which is the more rewarding strategy, and second, would leave enough time to benefit from using it. Note that learning provides a mechanism of conditional expression that is especially advantageous when it is difficult for the animal to know *a priori* its state or characteristics relative to others (i.e. to determine its individual asymmetry). For example, a rule of the kind 'play scrounger if physically strong' may not be useful if an individual cannot tell its strength relative to other group members. Moreover, as 'strength' itself may be determined by multiple genetic and environmental factors, and because it may not be the only determinant of scrounging success (e.g. various factors such as experience may also be important), learning may be the only feasible way to choose a strategy in relation to phenotypic asymmetries. This may be consistent with the observed correlations between being fast, aggressive or an efficient forager, and the tendency to scrounger [46–48]. The result may appear to be a consistent correlation between a group of physical and behavioural traits known as animal personality (see [45,49–51], for review). Interestingly, this association may emerge from initial genetic or phenotypic asymmetry and a cascade of conditional expressions that are based on learned strategy choice.

REFERENCES

1 Maynard-Smith, J. & Price, G. R. 1973 The logic of animal conflict. *Nature* **246**, 15–18. (doi:10.1038/246015a0)

2 Maynard-Smith, J. 1982 *Evolution and the theory of games.* Cambridge, UK: Cambridge University Press.

3 Harley, C. B. 1981 Learning the evolutionarily stable strategy. *J. Theor. Biol.* **89**, 611–633. (doi:10.1016/0022-5193(81)90032-1)

4 Beauchamp, G. 2000 Learning rules for social foragers: implications for the producer–scrounger game and ideal free distribution theory. *J. Theor. Biol.* **207**, 21–35. (doi:10.1006/jtbi.2000.2153)

5 Hamblin, S. & Giraldeau, L. A. 2009 Finding the evolutionarily stable learning rule for frequency-dependent foraging. *Anim. Behav.* **78**, 1343–1350. (doi:10.1016/j.anbehav.2009.09.001)

6 Bernstein, C., Kacelnik, A. & Krebs, J. R. 1988 Individual decisions and the distribution of predators in a patchy environment. *J. Anim. Ecol.* **57**, 1007–1026. (doi:10.2307/5108)

7 Borgers, T. & Sarin, R. 1997 Learning through reinforcement and replicator dynamics. *J. Econ. Theor.* **77**, 1–14. (doi:10.1006/jeth.1997.2319)

8 Stephens, D. W. 1991 Change, regularity, and value in the evolution of animal learning. *Behav. Ecol.* **2**, 77–89. (doi:10.1093/beheco/2.1.77)

9 Dunlap, A. S. & Stephens, D. W. 2009 Components of change in the evolution of learning and unlearned preference. *Proc. R. Soc. B* **276**, 3201–3208. (doi:10.1098/rspb.2009.0602)

10 Bateson, P. 1979 How do sensitive periods arise and what are they for? *Anim. Behav.* **27**, 470–486. (doi:10.1016/0003-3472(79)90184-2)

11 Houston, A. I. & Sumida, B. H. 1987 Learning rules, matching and frequency-dependence. *J. Theor. Biol.* **126**, 289–308. (doi:10.1016/S0022-5193(87)80236-9)

12 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)

13 Crawford, V. P. 1989 Learning and mixed-strategy equilibria in evolutionary games. *J. Theor. Biol.* **140**, 537–550. (doi:10.1016/S0022-5193(89)80113-4)

14 Fogel, D. B., Fogel, G. B. & Andrews, P. C. 1997 On the instability of evolutionary stable strategies. *Biosystems* **44**, 135–152. (doi:10.1016/S0303-2647(97)00050-6)

15 Fogel, G. B., Andrews, P. C. & Fogel, D. B. 1998 On the instability of evolutionary stable strategies in small populations. *Ecol. Model* **109**, 283–294. (doi:10.1016/S0304-3800(98)00068-4)

16 Kedar, H., Rodriguez-Girones, M. A., Yedvab, S., Winkler, D. W. & Lotem, A. 2000 Experimental evidence for offspring learning in parent–offspring communication. *Proc. R. Soc. Lond. B* **267**, 1723–1727. (doi:10.1098/rspb.2000.1201)

17 Bell, K. E. & Baum, W. M. 2002 Group foraging sensitivity to predictable and unpredictable changes in food distribution: past experience or present circumstances? *J. Exp. Anal. Behav.* **78**, 179–194. (doi:10.1901/jeab.2002.78-179)

18 Belmaker, A. 2007 Learning to choose between foraging strategies in adult house sparrows. MS thesis, Tel-Aviv University, Tel-Aviv, Israel.

19 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)

20 Katsnelson, E., Motro, U., Feldman, M. W. & Lotem, A. 2008 Early experience affects producer–scrounger foraging tendencies in the house sparrow. *Anim. Behav.* **75**, 1465–1472. (doi:10.1016/j.anbehav.2007.09.020)

21 Barnard, C. J. & Sibly, R. M. 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)

22 Giraldeau, L. A. & Caraco, T. 2000 *Social foraging theory.* Princeton, NJ: Princeton University Press.

23 Dubois, F., Morand-Ferron, J. & Giraldeau, L. A. 2010 Learning in a game context: strategy choice by some keeps learning from evolving in others. *Proc. R. Soc. B* **277**, 3609–3616. (doi:10.1098/rspb.2010.0857)

24 Wolf, M., van Doorn, G. S. & Weissing, F. J. 2008 Evolutionary emergence of responsive and unresponsive personalities. *Proc. Natl Acad. Sci. USA* **105**, 15 825–15 830. (doi:10.1073/pnas.0805473105)

25 Real, L. A. 1991 Animal choice behavior and the evolution of cognitive architecture. *Science* **253**, 980–986. (doi:10.1126/science.1887231)

26 Niv, Y., Joel, D., Meilijson, I. & Ruppin, E. 2002 Evolution of reinforcement learning in foraging bees: a simple explanation for risk averse behavior. *Neurocomputing* **44**, 951–956. (doi:10.1016/S0925-2312(02)00496-4)

27 Arbilly, M., Motro, U., Feldman, M. W. & Lotem, A. 2010 Co-evolution of learning complexity and social foraging strategies. *J. Theor. Biol.* **267**, 573–581. (doi:10.1016/j.jtbi.2010.09.026)

28 Houston, A. I. & McNamara, J. M. 1988 The ideal free distribution when competitive abilities differ: an approach based on statistical-mechanics. *Anim. Behav.* **36**, 166–174. (doi:10.1016/S0003-3472(88)80260-4)

29 Barta, Z. & Giraldeau, L. A. 1998 The effect of dominance hierarchy on the use of alternative foraging tactics: a phenotype-limited producing–scrounging game. *Behav. Ecol. Sociobiol.* **42**, 217–223. (doi:10.1007/s002650050433)

30 Fishman, M. A., Lotem, A. & Stone, L. 2001 Heterogeneity stabilizes reciprocal altruism interactions. *J. Theor. Biol.* **209**, 87–95. (doi:10.1006/jtbi.2000.2248)

31 Dubois, F., Giraldeau, L. A. & Grant, J. W. A. 2003 Resource defense in a group-foraging context. *Behav. Ecol.* **14**, 2–9. (doi:10.1093/beheco/14.1.2)

32 McElreath, R. & Strimling, P. 2006 How noisy information and individual asymmetries can make 'personality' an adaptation: a simple model. *Anim. Behav.* **72**, 1135–1139. (doi:10.1016/j.anbehav.2006.04.001)

33 Ruxton, G. D. & Beauchamp, G. 2008 The application of genetic algorithms in behavioural ecology, illustrated with a model of anti-predator vigilance. *J. Theor. Biol.* **250**, 435–448. (doi:10.1016/j.jtbi.2007.10.022)

34 Schwabl, H. 1993 Yolk is a source of maternal testosterone for developing birds. *Proc. Natl Acad. Sci. USA* **90**, 11 446–11 450. (doi:10.1073/pnas.90.24.11446)

35 Groothuis, T. G. G. & Carere, C. 2005 Avian personalities: characterization and epigenesis. *Neurosci. Biobehav. Rev.* **29**, 137–150. (doi:10.1016/j.neubiorev.2004.06.010)

36 McNamara, J. M. & Houston, A. I. 1987 Memory and the efficient use of information. *J. Theor. Biol.* **125**, 385–395. (doi:10.1016/S0022-5193(87)80209-6)

37 Mery, F. & Kawecki, T. J. 2004 An operating cost of learning in *Drosophila melanogaster*. *Anim. Behav.* **68**, 589–598. (doi:10.1016/j.anbehav.2003.12.005)

38 Bergstrom, C. T. & Godfrey-Smith, P. 1998 On the evolution of behavioral heterogeneity in individuals and populations. *Biol. Philos.* **13**, 205–231. (doi:10.1023/A:1006588918909)

39 Krebs, J. R. & Davies, N. B. 1993 *An introduction to behavioural ecology.* Oxford, UK: Blackwell Scientific Publications.

40 Feldman, M. W., Aoki, K. & Kumm, J. 1996 Individual versus social learning: evolutionary analysis in a fluctuating environment. *Anthropol. Sci.* **104**, 209–231.

41 Beauchamp, G. 2001 Consistency and flexibility in the scrounging behaviour of zebra finches. *Can. J. Zool.* **79**, 540–544. (doi:10.1139/z01-008)

42 Giraldeau, L. A., Caraco, T. & Valone, T. J. 1994 Social foraging: individual learning and cultural transmission of innovations. *Behav. Ecol.* **5**, 35–43. (doi:10.1093/beheco/5.1.35)

43 March, J. G. 1996 Learning to be risk averse. *Psychol. Rev.* **103**, 309–319. (doi:10.1037/0033-295X.103.2.309)

44 Gross, R., Houston, A. I., Collins, E. J., McNamara, J. M., Dechaume-Moncharmont, F. X. & Franks, N. R. 2008 Simple learning rules to cope with changing environments. *J. R. Soc. Interface* **5**, 1193–1202. (doi:10.1098/rsif.2007.1348)

45 Dall, S. R. X., Houston, A. I. & McNamara, J. M. 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)

46 Liker, A. & Barta, Z. 2002 The effects of dominance on social foraging tactic use in house sparrows. *Behaviour* **139**, 1061–1076. (doi:10.1163/15685390260337903)

47 Beauchamp, G. 2006 Phenotypic correlates of scrounging behavior in zebra finches: role of foraging efficiency and dominance. *Ethology* **112**, 873–878. (doi:10.1111/j.1439-0310.2006.01241.x)

48 Marchetti, C. & Drent, P. J. 2000 Individual differences in the use of social information in foraging by captive great tits. *Anim. Behav.* **60**, 131–140. (doi:10.1006/anbe.2000.1443)

49 Bell, A. M. 2007 Evolutionary biology: animal personalities. *Nature* **447**, 539–540. (doi:10.1038/447539a)

50 Sih, A. & Bell, A. M. 2008 Insights for behavioral ecology from behavioral syndromes. *Adv. Stud. Behav.* **38**, 227–281. (doi:10.1016/S0065-3454(08)00005-3)

51 Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. 2004 Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79**, 241–277. (doi:10.1086/422893)