Learning in a game context: strategy choice by some keeps learning from evolving in others

Frédérique Dubois^{1,*}, Julie Morand-Ferron²
and Luc-Alain Giraldeau²

¹Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, Canada H3C 3J7

²Département des Sciences Biologiques, Université du Québec à Montréal, C.P. 8888, Succ. Centre-ville, Montréal, Québec, Canada H3C 3P8

Behavioural decisions in a social context commonly have frequency-dependent outcomes and so require analysis using evolutionary game theory. Learning provides a mechanism for tracking changing conditions and it has frequently been predicted to supplant fixed behaviour in shifting environments; yet few studies have examined the evolution of learning specifically in a game-theoretic context. We present a model that examines the evolution of learning in a frequency-dependent context created by a producer–scrounger game, where producers search for their own resources and scroungers usurp the discoveries of producers. We ask whether a learning mutant that can optimize its use of producer and scrounger to local conditions can invade a population of non-learning individuals that play producer and scrounger with fixed probabilities. We find that learning provides an initial advantage but never evolves to fixation. Once a stable equilibrium is attained, the population is always made up of a majority of fixed players and a minority of learning individuals. This result is robust to variation in the initial proportion of fixed individuals, the rate of within- and between-generation environmental change, and population size. Such learning polymorphisms will manifest themselves in a wide range of contexts, providing an important element leading to behavioural syndromes.

Keywords: learning; polymorphism; frequency dependence; producer–scrounger game; personality; behavioural syndrome

1. INTRODUCTION

Social behaviours commonly have frequency-dependent outcomes that require analysis using a game-theoretic approach (Maynard-Smith 1982; Dugatkin & Reeve 1998). Prominent examples are the use of escalated versus ritualized fighting strategies (Maynard-Smith & Price 1973), engaging in reciprocity cooperation or defection (Axelrod & Hamilton 1981; Dugatkin 1997), and choosing habitats or group sizes to settle in (Fretwell & Lucas 1969; Giraldeau & Gillis 1988). The game-theoretic approach applied to animal behaviour predicts that evolution will lead to game solutions that are evolutionarily stable strategies (ESS): a strategy (or mixture of tactics) that, once adopted by a population, cannot be invaded by any other likely alternative (Maynard-Smith 1982). These solutions are generally assumed to have evolved over ecological time by natural selection acting on genetic alternatives, where the power of selection acting over generations replaces the cognitive ability of players in choosing the best possible strategy (Hammerstein 1998). Many of these behavioural games are repeated several times in the course of an individual's lifetime: for instance, an individual may need to decide whether to defend a given resource clump at escalated levels several times a day (Grant 1993). Each time the

animal may be faced with resources that differ in their size, amount, quality or density, and may face an opponent that is sometimes bold, strong or weak. A strategy decision based on the average interaction over the course of an animal's lifetime may do the job, but clearly in many circumstances adjusting one's strategy decision to local conditions may provide a superior return (Clements & Stephens 1995). It is more likely that behavioural mechanisms that allow animals to adopt the best possible behaviour given the local conditions would outperform fixed behaviour and this behavioural mechanism is likely to be learning (Harley 1981; Milinski 1984; Houston & Sumida 1987; Stephens & Clements 1998; Giraldeau & Dubois 2008).

Few studies of learning have considered the consequences of animals living in groups where the learning by some individuals will probably exert effects on the learning of others. When learning arises in a group, it can often reduce the opportunity for learning in naive individuals (Giraldeau 1984; Beauchamp & Kacelnik 1991; Giraldeau *et al.* 1994; Giraldeau & Caraco 2000) and so has the potential to lead to a stable polymorphism of knowledgeable and naive individuals. In this study, we explore the consequences of such frequency-dependent effects of learning on the evolution of learning in a common frequency-dependent context. We study a situation where animals are engaged in a producer–scrounger (PS) game (Barnard & Sibly 1981; Giraldeau & Dubois 2008). The game is

* Author for correspondence (frederique.dubois@umontreal.ca).

Table 1. Definition of the parameters used in the model. For all parameters that could vary, the range of tested values is specified.

symbol	meaning
G	number of foragers within the group, including learning and fixed individuals (range of tested values: 10–100)
λ	encounter rate with food patches (range of tested values: 1–10)
F_M	mean patch quality (range of tested values: 10–50)
a	finder's advantage, that is independent on patch quality (range of tested values: $1 - (F_M - 1)$)
D_W	measure of deviation around the mean patch quality that can be observed within a generation (range of tested values: $0 - (F_M - a)$)
N	number of changes in patch richness during a given generation (range of tested values: 0–20)
D_B	measure of deviation around the mean patch quality that can be observed between generations (range of tested values: $0 - (F_M - a)$)
x_i	proportion of learning individuals at the i th generation
$1 - x_i$	proportion of fixed individuals at the i th generation
$W_{P(1)}$	mean gain expected by a fixed or learning individual, when it plays producer
$W_{S(1)}$	mean gain expected by a fixed or learning individual, when it plays scrounger
α_1	mean proportion of time fixed individuals of type 1 play producer
α_2	mean proportion of time fixed individuals of type 2 play producer
$p_{1(i)}$	proportion of type 1 individuals among fixed foragers at the i th generation
$p_{2(i)}$	proportion of type 2 individuals among fixed foragers at the i th generation, with $p_{2(i)} = 1 - p_{1(i)}$
$\alpha_{F(i)}$	mean proportion of time fixed individuals play producer at the i th generation
$W_{1(i)}$	mean gain expected by a fixed individual of type 1 at the i th generation
$W_{2(i)}$	mean gain expected by a fixed individual of type 2 at the i th generation
$N_{P(i)}$	mean number of individuals playing producer, including learning and fixed foragers, at the i th generation
$N_{S(i)}$	mean number of individuals playing scrounger, including learning and fixed foragers, at the i th generation
ϵ_S	mean bias when a learning individual estimates the scrounger's payoff
ϵ_P	mean bias when a learning individual estimates the producer's payoff, with $\epsilon_P = k\epsilon_S$ and $k > 1$
α_L^*	mean proportion of time learning individuals should play producer in order to maximize their expected payoff, while conditions remain constant
$1 - \alpha_L^*$	mean proportion of time learning individuals should play scrounger in order to maximize their expected payoff, while conditions remain constant

quite general and applies whenever individuals are collectively engaged in searching for hidden resources. In this game all individuals can use one of two mutually exclusive search strategies: producer (looking for one's own resources) and scrounger (looking for producers that have found resources and exploiting those). The payoffs to scrounger are highly negatively frequency-dependent, such that this strategy does better than producer when rare and worse than producer when common (Barnard & Sibly 1981). In most cases, the game predicts a stable ESS mixture of producer and scrounger (Giraldeau & Caraco 2000; Giraldeau & Dubois 2008). We start off with a mixed population of fixed individuals that has already reached an evolutionary equilibrium via the action of natural selection acting on genetic alternatives. We examine the fate of a mutant learning individual that can optimize its use of producer and scrounger, and so take advantage of any sudden change of conditions that call for a different use of producer and scrounger.

2. METHODS

(a) *Calculating the fitness of learning and fixed individuals in a PS game*

As is common in foraging theory (Giraldeau & Caraco 2000; Stephens *et al.* 2007), we assume that food intake is an acceptable currency of fitness, and so develop equations to estimate the intake of fixed and learning individuals. We study the outcome of learning, without considering the actual details of the learning process. We assume, therefore, that whatever learning rule is actually used by the animal,

it succeeds without cost at informing the animal's decision with some known error. A population of G animals (all parameters are listed in table 1) forages for clumped resources. In this population some individuals (producers) obtain all their food by searching for the resource, while others (scroungers) obtain all their food by exploiting a fraction of the food uncovered by producers. In the present case, we assume that the scrounger exerts negative frequency dependence on both itself and the producers, but results remain unaltered when producer payoff is either unaffected or increases with scrounger frequency so long as scroungers do better than producers when rare and worse than producers when common (Giraldeau & Caraco 2000). Scroungers do better than producers when rare because they can exploit the discoveries of several producers while each producer is limited to its own discoveries. When common, however, scrounger does worse than producer because it has fewer producers to exploit and must share each with a larger number of scroungers. In this condition, frequency-dependent selection will maintain a mixed ESS combination of producers and scroungers at a frequency for which the payoffs to each are equal. A food clump contains F_M food items and the finder obtains a items (the finder's advantage) for its exclusive use. A learning individual can choose to use either producer or scrounger, or a given combination by alternating from one to the other at some rate. The learner always chooses the most profitable of these three possibilities. When the learner uses a strategy that consists of alternating between producer and scrounger, we say it is using a mixed strategy with probabilities (α_L) and $(1 - \alpha_L)$ for producer and scrounger, respectively.

We consider an initial population of $(G - 1)$ fixed individuals and one learning individual. The proportion of learning individuals at the i th generation (where i is an index varying from 1 to n) is x_i , and so we set $x_1 = 1/G$. For simplicity, we consider only two types of fixed individuals (i.e. type 1 and type 2 individuals), characterized by the frequency at which they use the producer and scrounger tactics. We denote by α_1 and α_2 , respectively, the frequency at which both types of individuals play producer. Thus, if the parameters α_1 and α_2 are equal to 1 and 0, respectively, fixed individuals of type 1 only play producer, whereas those of type 2 only play scrounger. On the other hand, when both parameters vary between 0 and 1 (with $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$), fixed individuals then use a mixed strategy.

The proportion of fixed individuals of types 1 and 2 among all fixed foragers in the starting generation is $p_{1(1)}$ and $p_{2(1)}$ (with $p_{2(1)} = 1 - p_{1(1)}$), respectively, and so the mean proportion of time-fixed individuals playing producer is $\alpha_{F(1)}$:

$$\alpha_{F(1)} = p_{1(1)} \times \alpha_1 + [1 - p_{1(1)}] \times \alpha_2. \quad (2.1)$$

We assume that patch discovery occurs sequentially and the producer's encounter rate with food patches is λ . The mean gain $W_{j(1)}$ expected by a fixed individual of type j (where $j = 1$ or 2) after T time units of foraging is

$$W_{j(1)} = \alpha_j \times W_{P(1)} + (1 - \alpha_j) \times W_{S(1)}. \quad (2.2)$$

$W_{P(1)}$ corresponds to an individual's expected intake when it plays producer, while $W_{S(1)}$ denotes its average expected gain when it plays scrounger:

$$W_{P(1)} = \lambda T \times \left(a + \frac{F_M - a}{1 + N_{S(1)}} \right) \quad (2.3a)$$

and

$$W_{S(1)} = \lambda T \times N_{P(1)} \times \left(\frac{F_M - a}{1 + N_{S(1)}} \right). \quad (2.3b)$$

In both equations (2.3a) and (2.3b), $N_{P(1)}$ and $N_{S(1)}$ correspond to the mean number of individuals playing producer and scrounger in the first generation, including both fixed and learning individuals. Thus, both values vary according to the proportion of learning (x_1) and fixed ($1 - x_1$) individuals, as well as to the probability that learning individuals play a mixed strategy of α_L (producer) and $1 - \alpha_L$ (scrounger):

$$N_{P(1)} = (1 - x_1)G \times \alpha_{F(1)} + x_1G \times \alpha_L \quad (2.4a)$$

and

$$N_{S(1)} = (1 - x_1)G \times (1 - \alpha_{F(1)}) + x_1G \times (1 - \alpha_L). \quad (2.4b)$$

Similarly, we can calculate the mean gain expected by the learning individual $\omega_{(1)}$ after T time units, if it plays a mixed strategy using probabilities of the producer and scrounger of α_L and $(1 - \alpha_L)$, respectively:

$$\omega_{(1)} = \alpha_L \times W_{P(1)} + (1 - \alpha_L) \times W_{S(1)}. \quad (2.5)$$

Although the above equations provide inaccurate estimates of the expected gains, especially in small populations, because a learning individual can play against itself, we used this common simplification given that it has

Figure 1. The gains expected by a fixed individual of type 1 (W_1), a fixed individual of type 2 (W_2) and the learning mutant (ω) for the first generation, in relation to the frequency at which the mutant uses the producer tactic (α_L). In all three panels, $T = 1$, $\lambda = 2$, $F_M = 12$, $a = 3$, $G = 10$, $D_B = 0$, $D_W = 0$ and $\alpha_{F(1)} = 0.35$. In (a) $\alpha_1 = 1$, $\alpha_2 = 0$ and $p_1 = 0.35$. In (b) $\alpha_1 = 0.5$, $\alpha_2 = 0.2$ and $p_1 = 0.5$. In (c) $\alpha_1 = 0.35$, $\alpha_2 = 0.35$ and $p_1 = 0.5$.

no qualitative effect on the equilibrium states on the model, and hence does not affect our conclusions.

To find the solution to the game, we seek the probability α_L^* , at which the learning individual plays producer in order to maximize its expected gain, $\omega_{(1)}$. To do that, we first estimate the gain expected by the learning individual $\omega_{(1)}$ for any α_L between 0 and 1, and then retain the value for which it is maximal. The gain expected by the learning mutant, and hence its optimal use of the producer and scrounger tactics, depends only on the mean frequency at which fixed individuals produce and not on the rate at which both types of fixed foragers employ this tactic (figure 1). Furthermore, in the case illustrated in figure 1, the learning individual's mean expected gain increases non-linearly as it increases its use of the producer tactic within

a given generation, and the optimal solution is therefore a mixed strategy. This arises because the payoff received by a scrounger is highly negatively frequency-dependent on the frequency of scrounging. Indeed, as the payoffs associated with each tactic increase with the number of producers, the mean gain expected by the learning mutant first increases as it increasingly uses the producer tactic. However, if the scrounger payoff is highly negatively frequency-dependent on the frequency of scrounging, this implies that the payoff difference between the scrounger and the producer becomes increasingly large as the learning individual increases its uses of the producer tactic. As a consequence, it has no interest in playing producer too frequently, because the payoffs it gets when playing scrounger then become very much larger. However, if most of the time the learning individual plays producer, its mean payoff is then almost equal to that of a pure producer (i.e. W_1 in figure 1a), and hence may be smaller than the gain it would expect if it were to play scrounger at a higher frequency.

(b) Adding estimation error

We do not specify the actual learning mechanism adopted by the animal, but allow for this mechanism to either over- or underestimate the value of both producer and scrounger. We assume that the magnitude of this bias increases with the size of what must be estimated (Bateson & Kacelnik 1995). Because producers get a larger payoff from each patch than do scroungers (Giraldeau *et al.* 1990), the bias for producer (ε_P), is larger than the bias for scrounger (ε_S), with $\varepsilon_S = k\varepsilon_P$ and $k < 1$. A learning forager has the same probability q of having a biased estimate when assessing the producer's or the scrounger's payoffs, and so the estimate of its expected gain $\hat{w}_{(1)}$ is equal to

$$\hat{w}_{(1)} = \alpha_L [q(W_{P(1)} + \varepsilon_P) + (1 - q)W_{P(1)}] + (1 - \alpha_L) [q(W_{S(1)} + \varepsilon_S) + (1 - q)W_{S(1)}]. \quad (2.6a)$$

Assuming $q = 1$ and simplifying gives

$$\hat{w}_{(1)} = \alpha_L \times W_{P(1)} + (1 - \alpha_L) \times W_{S(1)} + \alpha_L \times \varepsilon_P(1 - k) + k\varepsilon_P. \quad (2.6b)$$

As the last term of this equation ($k\varepsilon_P$) is independent of α_L , it does not influence the solution α_L^* and so it is ignored. By setting $\varepsilon = \varepsilon_P(1 - k)$ we obtain

$$\hat{w}_{(1)} = \alpha_L \times (W_{P(1)} + \varepsilon) + (1 - \alpha_L) \times W_{S(1)}. \quad (2.6c)$$

(c) The evolution of learning mutants

To estimate the fitness of learning, we assume that learning players play α_L^* , the optimal frequency of producer that maximizes $\hat{w}_{(1)}$, and replace α_L with α_L^* in equations (2.2)–(2.5). Population size is kept constant from one generation to the next. At each generation (i), we set the proportion of fixed foragers of types 1 and 2 among all fixed individuals ($p_{1(i)}$ and $p_{2(i)}$) as well as the proportion of learning foragers (x_i) according to their relative success in the previous generation ($i - 1$). The relative proportion of learning individuals in the next generation (x_2) is thus given by

$$x_2 = \frac{x_1 \times \omega_1}{x_1 \times \omega_1 + (1 - x_1) \times (p_{1(1)}W_{1(1)} + p_{2(1)}W_{2(1)})}, \quad (2.7)$$

while the mean proportion of fixed individuals of type 1 ($p_{1(2)}$) among fixed individuals in the population will be

given by

$$p_{1(2)} = \frac{p_{1(1)} \times W_{1(1)}}{p_{1(1)} \times W_{1(1)} + p_{2(1)} \times W_{2(1)}}. \quad (2.8)$$

The same procedure is repeated from generation to generation until the frequency of learning individuals, as well as the rate at which they use both producer and scrounger alternatives, is such that their fitnesses reach stable equilibrium values.

(d) Adding environmental variability

Learning is predicted to be most useful when conditions vary. We explore environmental variation at two scales: between- and within-generation changes. Inter-generational variability is achieved by randomly drawing the value of food patches at the onset of each generation. We allow the number of food items per patch to take any value between $F_M - D_B$ and $F_M + D_B$, where F_M is the mean patch quality and D_B is a measure of the deviation around the mean that can be observed between generations. The simulation is run 20 times for a given set of parameter values, and the results from all 20 simulations are used to obtain the mean values and standard deviations.

Intra-generational variation is introduced by allowing the number of food items per patch to take N ($N \geq 1$) different values during a generation, with each value lasting the same amount of time. As learning individuals change their use of the producer and scrounger tactics whenever patch value changes, we find the best responses α_L^* for each of the N patch values adopted. At the beginning of each period, the value of the food patches is pulled at random from a set of values ranging from $F_M - D_W$ to $F_M + D_W$, where F_M is the mean patch quality and D_W is a measure of the deviation around the mean that can be observed within a generation. After the N consecutive periods, the mean gains expected by a fixed individual of type 1, a fixed individual of type 2 and a learning forager are estimated for the entire generation. These means are then used to deduce the relative proportion of learning individuals that will be present within the population during the next generation, and the procedure is repeated over consecutive generations until all the parameters reach a plateau. Each simulation is run 20 times for a given set of parameters, and the mean values and standard deviations are estimated from the 20 simulations.

3. RESULTS

(a) Unchanging environments

When there is no variation within or between generations, the model predicts that the mean proportion of producers and scroungers estimated for the population of G foragers rapidly reaches the mixed evolutionarily stable equilibrium predicted by the standard rate-maximizing PS game (Vickery *et al.* 1991; Giraldeau & Beauchamp 1999). In addition, the model predicts that neither fixed nor learning individuals ever reach fixation, except when all fixed individuals use the producer and scrounger tactics with the same frequencies (i.e. $\alpha_1 = \alpha_2$). In this particular case, fixed individuals cannot be invaded by a mutant learning individual and the expected proportion of learning individuals therefore rapidly tends to zero. In all other situations (i.e. when $\alpha_1 \neq \alpha_2$), learning individuals are always maintained within the

Figure 2. The magnitude of estimation bias affects (a) the frequency at which learning individuals should play producer (α_L^*) and (b) the proportion of learning foragers at equilibrium. $F_M = 12$, $a = 3$, $G = 10$, $N = 1$, $T = 1$, $\lambda = 2$, $D_B = 0$ and $D_W = 0$. Filled circles: $\alpha_1 = 1$, $\alpha_2 = 0$, $p_1 = 0.35$; filled squares: $\alpha_1 = 0.8$, $\alpha_2 = 0.2$, $p_1 = 0.25$.

population. The frequency at which they should use the producer and scrounger tactics then varies depending on the extent to which their estimate of the producer's expected gain is biased (figure 2a), but the extent of bias has only a weak effect on the expected stable proportion of such learning individuals at equilibrium (figure 2b).

More precisely, the model predicts that the optimal behaviour of learning individuals at equilibrium is to never scrounge when they overestimate payoffs to producer (figure 2b). Thus, learning individuals then play a pure producer tactic, regardless of the initial conditions (figure 2a). This occurs because the first learning individual plays against a field of fixed strategies and so can optimize its allocation to producer and scrounger (figure 1). This optimal allocation always favours a mixture of tactics that is biased towards the use of producer. By using its optimal mixed strategy, the learning individual exerts a frequency-dependent effect on the payoffs to the fixed players that favours the individuals playing the scrounger tactic at the highest frequency. The consequence is that the mean proportion of time that fixed individuals play scrounger will increase in the next generation, which in turn exerts a frequency-dependent effect on the payoffs of the learning individuals that biases their optimal mixed strategy further still towards producer (figure 3). In the end, mixed strategies are no longer optimal for the learning individuals who play pure producer.

When learning individuals systematically underestimate the expected payoffs, they are expected to use an

Figure 3. The optimal frequency (α_L^*) at which the learning mutant plays producer at equilibrium in relation to the mean proportion of time fixed individuals play producer (α_F). $F_M = 12$, $a = 3$, $G = 10$, $\alpha_1 = 1$, $\alpha_2 = 0$, $\varepsilon = 0$, $N = 1$, $T = 1$, $\lambda = 2$, $D_B = 0$ and $D_W = 0$.

equilibrium mixture of producer and scrounger tactics (figure 2a). As the payoffs to producer are increasingly underestimated, the learning individuals' optimal mixture is increasingly biased towards scrounger until eventually a pure scrounger tactic is optimal when the magnitude of the underestimation becomes very large. As the payoffs to producer are increasingly underestimated, the fraction of the population made up of learning individuals is predicted first to increase, and then to decrease (figure 2b).

(b) The effect of between- and within-generation environmental variability

The frequency of learning individuals increases with increasing between-generation variation (figure 4a). However, regardless of the extent of inter-generational variability, the learning strategy never spreads to fixation (figure 4a). Surprisingly, within-generation variation has much less of an effect on the frequency of learning individuals (figure 4b). The expected proportion of learning individuals remains unaffected by population size (figure 5).

4. DISCUSSION

We show that in a game-theoretic context, players who learn to alter their strategy use can do better than fixed players, at least initially, but can never invade the whole population. Learning, therefore, is not an ESS in all game-theoretic contexts. In fact, no matter what conditions we start off with, at evolutionary stability, the majority of players is always made up of fixed strategists. These non-learning individuals are maintained in the population because the behavioural adjustment of the optimizing learning individuals buffers the selective pressure on a segment of the fixed individuals. Learners can do better than fixed producers or fixed scroungers but never better than both. We discuss the specific conditions under which such learning and non-learning polymorphisms should be expected and explore their potential as an explanation for the consistent differences in learning abilities often reported within populations.

Our result may appear to contradict cases where flexibility is an ESS in a game-theoretic context. For instance,

Figure 4. Mean equilibrium proportion ($\pm$ s.d.) of learning individuals estimated from 20 simulation runs, as a function of the magnitude of (a) between- and (b) within-generation variation. In both panels, $F_M = 12$, $a = 3$, $G = 10$, $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_{F(1)} = 0.35$, $\varepsilon = 0$, $T = 1$ and $\lambda = 2$. In (a) $N = 3$ and $D_B = 0$, while in (b) $D_W = 0$.

both asymmetric hawk–dove (Parker 1974) and war of attrition (Hammerstein & Parker 1982) fighting games are, like the PS game, characterized by strong frequency-dependent payoffs, and yet, in both these fighting games, the expected ESS is the ‘assessor’ strategy: a strategy that chooses its tactic based on an assessment of player asymmetry. We think the difference lies in exactly what must be assessed by the flexible strategist. The fighting assessor decides on the basis of an estimation of a fixed quantity, a difference between players that is unaffected by strategy choice. In the PS game, as in many behavioural games (Dugatkin & Reeve 1998), the learning individual must assess the payoff associated with each available alternative. This payoff is not a fixed quantity and depends on the strategy use of others. We predict therefore that in all cases where strategy choice depends on the assessed payoffs of each alternative, flexible learning will not be an ESS. Instead, we should expect a majority of fixed strategists with a minority of learning flexible strategists.

Other studies of flexible strategy use in a game-theoretic context have found results consistent with our own, but for different reasons. For instance, Wolf *et al.* (2008) explore whether individuals with the ability to detect the best of two habitats can do better than individuals that assort randomly. Because responsive individuals all choose to settle in the habitat that is initially better, they find that as the number of responsive individuals increases, the crowding and competition in the initially better habitat reduces the

Figure 5. Mean equilibrium proportion ($\pm$ s.d.) of learning individuals estimated from 20 simulation runs, as a function of population size. $F_M = 12$, $a = 3$, $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_{F(1)} = 0.35$, $\varepsilon = 0$, $T = 1$, $\lambda = 2$, $N = 1$, $D_W = 0$ and $D_B = 4$.

advantage of being responsive until responsive individuals do no better than unresponsive players that assort randomly. The result is a stable polymorphism of responsive and non-responsive players, the ratio of which depends on the costs of being responsive and the relative difference between the initial values of the two habitats. Their result, however, is attained by a different process from ours. Much as was true in the fighting games, the responsive individuals in Wolf *et al.*'s (2008) study decided on settlement strategy on the basis of the assessment of a fixed quantity, the initial difference between the two habitats. They did not then continue to assess as the animals settled in one or the other, and so were not affected by the frequency dependence of the payoffs but only the density dependence that results from accumulation of responsive individuals in the initially better patch. Eliassen *et al.* (2006) model patch choice in a more dynamic way, allowing learning individuals to assess throughout the patch foraging game. They too find the emergence of a learner–non-learner polymorphism, but because the dynamics of settlement become chaotic when learners become too numerous, there is no longer any gain by learning. These results, taken with our own, indicate that learning is unlikely to be an ESS in many game-theoretic contexts. While there is some evidence that learning is involved in strategy choice (Katsnelson *et al.* 2008) and in reaching the ESS (Mottley & Giraldeau 2000; Morand-Ferron & Giraldeau 2010) in the PS game, no empirical study has yet examined whether individuals differ in the rule they use to allocate to each tactic (i.e. fixed or flexible). It is not currently known whether observed group-level changes in scrounger frequencies (reviewed in Giraldeau & Dubois 2008) are the consequence of a learning process occurring in all members of a group or only in some, as predicted by recent models (Eliassen *et al.* 2006; Wolf *et al.* 2008; this paper).

Earlier explorations of the conditions that favour the evolution of learning predicted learning when environments exhibit strong inter-generational variation (Scheiner 1998; Luttbeg & Warner 1999; Collins *et al.* 2006). But in our case only a fraction of the population ever gets to evolve learning ability, even when the value of the alternatives is assessed without bias. At best, no more than 40 per cent of the population ever gets to adopt a learning strategy, even when inter-generational

variation is very high. Moreover, the fraction of learners within the population remains surprisingly little affected by within-generation variability, which other models predicted to have a strong effect on the evolution of learning (Stephens 1991; Bergman & Feldman 1995). The weak effect of within-generation variation in our study implies that the value of learning lies more with adjusting one's behaviour to changes in competitors' behaviour than to changes in resource characteristics. Our results suggest that having to adopt a strategy on the basis of an assessment of payoffs that are frequency-dependent favours the maintenance of a large fraction of non-learning individuals within a population.

Although our study addresses the evolution of learning, it does not specify exactly how the learning actually proceeds. For instance, our study is not concerned with the number of experiences required for an animal to acquire an estimate of the payoffs of each alternative. Similarly, we were not concerned with memory windows or even forgetting. In a way, our study only considers the outcome of a potentially instantaneous learning rule. In reality, learning is likely to be much less efficient a mechanism than the one we assumed operated here. Moreover, real learning probably also imposes costs in terms of sampling, and even in terms of metabolic support of the necessary learning infrastructure. However, taking all these factors into consideration would only increase the likelihood of a learning–non-learning polymorphism, with even fewer expected learning individuals at equilibrium.

Recently, behavioural ecologists have been interested in strong behavioural differences of the type we predict here and have referred to them as syndromes of behavioural plasticity (Sih *et al.* 2004; Dingemanse *et al.* 2009), flexibility (McElreath & Strimling 2006), responsiveness (Wolf *et al.* 2008) or different coping styles (Koolhaas *et al.* 1999). Personality polymorphisms are thought to be maintained as the consequence of strong trade-offs between benefits and costs across ecological contexts, as well as different selective forces acting either at different times during an organism's life history or in different environments exploited by the organism (Wilson 1998; Sih *et al.* 2004). However, the learning polymorphism we predict from frequency-dependent assessment games occurs without many of the necessary conditions normally invoked to explain the evolution of behavioural syndromes. The need to make repeated strategy decisions on the basis of frequency-dependent payoffs seems sufficient to promote the evolution of a learning–non-learning polymorphism. These individual differences in learning ability are probably correlated with behaviours normally associated with learning, such as sampling and exploration. As a result, personality types that have been described as bold and shy (Wilson *et al.* 1994; Johnson & Sih 2007), proactive and reactive (Koolhaas *et al.* 1999), and fast and slow to explore (Verbeek *et al.* 1994; Dingemanse *et al.* 2002), may in fact all be descriptors of behaviours that correspond to stable and widespread learning–non-learning dimorphism. If we are correct, then, we predict that differences in learning ability can be the cause rather than the consequence of personality traits and that individual differences in learning ability will be greatest in species that are commonly engaged in repetitive decision-making in a game context.

We thank Janne-Tuomas Seppanen, Arnem Lotem and Jeremy Kendal for their valuable comments on previous versions of this manuscript. This research was financially supported by Natural Sciences and Engineering Research Council Discovery grants to F.D. and L.-A.G. J.M.-F. was supported by a NSERC Discovery grant to L.-A.G. and by a post-doctoral fellowship from the Fonds Québécois de la Recherche sur la Nature et les Technologies.

REFERENCES

- Axelrod, R. & Hamilton, W. D. 1981 The evolution of cooperation. *Science* **211**, 1390–1396. (doi:10.1126/science.7466396)
- 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)
- Bateson, M. & Kacelnik, A. 1995 Accuracy of memory for amount in the foraging starling *Sturnus vulgaris*. *Anim. Behav.* **50**, 431–443. (doi:10.1006/anbe.1995.0257)
- Beauchamp, G. & Kacelnik, A. 1991 Effects of the knowledge of partners on learning rates in zebra finches *Taeniopygia guttata*. *Anim. Behav.* **41**, 247–253. (doi:10.1016/S0003-3472(05)80476-2)
- Bergman, A. & Feldman, M. W. 1995 On the evolution of learning: representation of a stochastic environment. *Theor. Popul. Biol.* **48**, 251–276. (doi:10.1006/tpbi.1995.1029)
- Clements, K. C. & Stephens, D. W. 1995 Testing models of non-kin cooperation: mutualism and the Prisoner's Dilemma. *Anim. Behav.* **50**, 527–535. (doi:10.1006/anbe.1995.0267)
- Collins, E. J., McNamara, J. M. & Ramsey, D. M. 2006 Learning rules for optimal selection in a varying environment: mate choice revisited. *Behav. Ecol.* **17**, 799–809. (doi:10.1093/beheco/arl008)
- Dingemanse, N. J., Both, C., Drent, P. J., Van Oers, K. & Van Noordwijk, A. J. 2002 Repeatability and heritability of exploratory behaviour in great tits from the wild. *Anim. Behav.* **64**, 929–938. (doi:10.1006/anbe.2002.2006)
- Dingemanse, N. J., Kazem, A. J. N., Réale, D. & Wright, J. 2009 Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.* **25**, 81–89. (doi:10.1016/j.tree.2009.07.013)
- Dugatkin, L. A. 1997 *Cooperation among animals: an evolutionary perspective*. Oxford, UK: Oxford University Press.
- Dugatkin, L. A. & Reeve, H. K. 1998 *Game theory and animal behavior*. Oxford, UK: Oxford University Press.
- Eliassen, S., Jørgensen, C. & Giske, J. 2006 Co-existence of learners and stayers maintains the advantage of social foraging. *Evol. Ecol. Res.* **8**, 1311–1324.
- Fretwell, S. D. & Lucas, J. R. 1969 On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheor.* **19**, 16–36. (doi:10.1007/BF01601953)
- Giraldeau, L.-A. 1984 Group foraging: the skill pool effect and frequency-dependent learning. *Am. Nat.* **124**, 72–79. (doi:10.1086/284252)
- Giraldeau, L.-A. & Beauchamp, G. 1999 Food exploitation: searching for the optimal joining policy. *Trends Ecol. Evol.* **14**, 102–106. (doi:10.1016/S0169-5347(98)01542-0)
- Giraldeau, L.-A. & Caraco, T. 2000 *Social foraging theory*. Princeton, NJ: Princeton University Press.
- Giraldeau, L.-A. & Dubois, F. 2008 Social foraging and the study of exploitative behavior. *Adv. Stud. Behav.* **38**, 59–104. (doi:10.1016/S0065-3454(08)00002-8)
- Giraldeau, L.-A. & Gillis, D. 1988 Do lions hunt in group sizes that maximize hunters' daily food returns? *Anim.*

- Behav.* **36**, 611–613. (doi:10.1016/S0003-3472(88)80037-X)
- Giraldeau, L.-A., Hogan, J. A. & Clinchy, M. J. 1990 The payoffs to producing and scrounging: what happens when patches are divisible? *Ethology* **85**, 132–146. (doi:10.1111/j.1439-0310.1990.tb00393.x)
- 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)
- Grant, J. W. A. 1993 Whether or not to defend? The influence of resource distribution. *Mar. Behav. Physiol.* **23**, 137–153. (doi:10.1080/10236249309378862)
- Hammerstein, P. 1998 What is evolutionary game theory? In *Game theory and animal behavior* (eds L. A. Dugatkin & H. K. Reeve), pp. 3–15. Oxford, UK: Oxford University Press.
- Hammerstein, P. & Parker, G. A. 1982 The asymmetric war of attrition. *J. Theor. Biol.* **96**, 647–682. (doi:10.1016/0022-5193(82)90235-1)
- Harley, C. B. 1981 Learning the evolutionarily stable strategy. *J. Theor. Biol.* **89**, 611–633. (doi:10.1016/0022-5193(81)90032-1)
- Houston, A. & Sumida, B. H. 1987 Learning rules, matching and frequency dependence. *J. Theor. Biol.* **126**, 289–308. (doi:10.1016/S0022-5193(87)80236-9)
- Johnson, J. C. & Sih, A. 2007 Fear, food, sex and parental care: a syndrome of boldness in the fishing spider, *Dolomedes triton*. *Anim. Behav.* **74**, 1131–1138. (doi:10.1016/j.anbehav.2007.02.006)
- 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)
- Koolhaas, J. M., Korte, S. M., Boer, S. F. D., Vegt, B. J. V. D., Reenen, C. G. V., Hopster, H., Jong, I. C. D., Ruis, M. A. W. & Blokhuis, H. J. 1999 Coping styles in animals: current status in behavior and stress-physiology. *Neurosci. Biobehav. Res.* **23**, 925–935. (doi:10.1016/S0149-7634(99)00026-3)
- Luttbeg, B. & Warner, R. R. 1999 Reproductive decision-making by female peacock wrasses: flexible versus fixed behavioral rules in variable environments. *Behav. Ecol.* **10**, 666–674. (doi:10.1093/beheco/10.6.666)
- Maynard-Smith, J. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
- Maynard-Smith, J. & Price, G. R. 1973 The logic of animal conflict. *Nature* **246**, 15–18.
- 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)
- Milinski, M. 1984 Competitive resource sharing: an experimental test of a learning rule for ESSs. *Anim. Behav.* **32**, 233–242. (doi:10.1016/S0003-3472(84)80342-5)
- Morand-Ferron, J. & Giraldeau, A. 2010 Learning behaviorally stable solutions to producer-scrounger games. *Behav. Ecol.* **21**, 343–348. (doi:10.1093/beheco/arp195)
- Mottley, K. & Giraldeau, L.-A. 2000 Experimental evidence that group foragers can converge on predicted producer-scrounger equilibria. *Anim. Behav.* **60**, 341–350. (doi:10.1006/anbe.2000.1474)
- Parker, G. A. 1974 Assessment strategy and evolution of fighting behavior. *J. Theor. Biol.* **47**, 223–243. (doi:10.1016/0022-5193(74)90111-8)
- Scheiner, S. M. 1998 The genetics of phenotypic plasticity. VII. Evolution in a spatially-structured environment. *J. Evol. Biol.* **11**, 303–320.
- 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)
- 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)
- Stephens, D. W. & Clements, K. C. 1998 Game theory and learning. In *Game theory and animal behavior* (eds L. A. Dugatkin & H. K. Reeve), pp. 239–260. Oxford, UK: Oxford University Press.
- Stephens, D. W., Brown, J. S. & Ydenberg, R. C. 2007 *Foraging: behavior and ecology*. Chicago, IL: Chicago University Press.
- Verbeek, M. E. M., Drent, P. J. & Wiepkema, P. R. 1994 Consistent individual differences in early exploratory behavior of male great tits. *Anim. Behav.* **48**, 1113–1121. (doi:10.1006/anbe.1994.1344)
- Vickery, W. L., Giraldeau, L.-A., Templeton, J. J., Kramer, D. L. & Chapman, C. A. 1991 Producers, scroungers, and group foraging. *Am. Nat.* **137**, 847–863. (doi:10.1086/285197)
- Wilson, D. S. 1998 Adaptive individual differences within single populations. *Phil. Trans. R. Soc. Lond. B* **353**, 199–205. (doi:10.1098/rstb.1998.0202)
- Wilson, D. S., Clark, A. B., Coleman, K. & Dearstyne, T. 1994 Shyness and boldness in humans and other animals. *Trends Ecol. Evol.* **9**, 442–446.
- Wolf, M., Doorn, G. S. V. & 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)