Understanding the evolution of learning by explicitly modeling learning mechanisms

Michal ARBILLY[*]

Department of Psychology, University of California, Berkeley, CA 94720, USA

Abstract   Models of the evolution of learning often assume that learning leads to the best solution to any task, and disregard the details of the learning and decision-making process along with its potential pitfalls. These models therefore do not explain instances in the animal behavior literature in which learning leads to maladaptive behaviors. In recent years a growing number of theoretical studies use explicit models of learning mechanisms, offering a fresh perspective on the issue by revealing the dynamics of information acquisition and biases arising from it. These models have pointed out possible learning rules and their adaptive value, and shown that the value of learning may crucially depend on such factors as the layout of the physical environment to be learned, the structure of the payoffs offered by different alternatives, the risk of failure, characteristics of the learner and social interactions. This review considers the merits of explicit modeling in studying the evolution of learning, describes the kinds of results that can only be obtained from this modeling approach, and outlines directions for future research [*Current Zoology* 61 (2): 341–349, 2015].

Keywords   Agent-based simulation, Learning bias, Learning rule, Decision rule, Cognition

1   Introduction

Traditionally, theoretical models of behavior are concerned with optimality, assuming that evolved behaviors provide the organisms with the best solution to any given situation (e.g. Charnov, 1976; Stephens and Krebs, 1986). However, experiments showing animals making what appears to be suboptimal and maladaptive behaviors suggest that is not necessarily the case (e.g. Bateson et al., 2002; Marsh and Kacelnik, 2002; Shafir et al., 2002). This has led to a re-examination of the theoretical approach. Notably, it has been suggested that rather than being optimal in any specific situation, evolved decision-making mechanisms are optimal on average, across situations, considering the ecological and evolutionary circumstances they have developed in (Schuck-Paim et al., 2004; Frankenhuis and Del Giudice, 2012). It has been recognized that in order to understand these behaviors, we must not simply assume optimal behavior, but consider the evolution of the mechanisms that produce the behavior (McNamara and Houston, 2009), and try to use more complex environments in our models in an effort to bring them closer to the real world (Fawcett et al., 2014). Furthermore, behavioral ecologists have been called to abandon the treatment of behaviors as the sole target of selection,

and instead consider the learning and decision-making mechanisms that generate them (Fawcett et al., 2013).

Pioneering theoretical work on the evolution of learning has used mostly analytical models to study questions such as the adaptive value of learning (Stephens, 1989, 1991), learning rules efficiency (McNamara and Houston, 1985, 1987b) and an evolutionarily stable learning rule (Harley, 1981; Houston and Sumida, 1987; Tracy and Seaman, 1995). These models have laid the foundations for studying many questions in the evolution of learning; however, they did not delve into the details of the learning process.

Learning is a dynamic process. The decision-making process is often probabilistic and results in stochastic outcomes; every decision can affect future evaluations and choices, breeding further stochasticity and making its exact consequences difficult to predict. While mathematical analysis of a learning rule can provide an approximation of the value-estimation process and an expectation of the mean result, deriving more than that would be analytically quite difficult and calls for other methods (Stephens, 1993; Dridi and Lehmann, 2014).

Agent-based computer models simulate the behavior of individuals and lets them interact with the model environment, among themselves, or both, to study questions concerning the outcomes of these interactions. The

---

∗ Corresponding author. E-mail: michalarbilly@gmail.com

model usually defines some rules governing the behavior of the agents and the dynamics of the system, and can leave much room for stochasticity by assuming some rules are probabilistic. The behavior of agents can be recorded for later analysis, and they themselves may have memories of previous events in the simulation, upon which they may base future behavior. These are in fact artificial societies, attempting to capture aspects of the real world *in silico* (Epstein and Axtell, 1996; Peck, 2004; Epstein, 2009). Although the high level of detail in these simulations interferes with covering the full parameter space, making them less general than analytical models, it is also one of their great strengths. They may be viewed as an experimental system, with an advantage of being highly manipulable compared to real-life ones (Peck, 2004). All these properties make agent based simulations a prime tool in studying a process such as learning. Combined with analytical models, they may provide us with a wide and deep understanding of its evolution.

The use of computer simulations in studying the evolution of learning is not new. Some pioneering work in the field goes back to the 1980s (Regelmann, 1984; Bernstein et al., 1988), but such studies were rather rare. Recent years, however, have seen a great increase in the use of computer simulations for that aim (Beauchamp, 2000; Beauchamp and Fernández-Juricic, 2004; Amano et al., 2006; van der Post and Hogeweg, 2006, 2009; Eliassen et al., 2007, 2009; Groß et al., 2008; Hamblin and Giraldeau, 2009; van der Post et al., 2009; Rendell et al., 2010a; Arbilly et al., 2010, 2011; Katsnelson et al., 2012; Trimmer et al., 2012; Afshar and Giraldeau, 2014; Kolodny et al., 2014; Lotem and Biran-Yoeli, 2014; Arbilly and Laland, 2014). This is perhaps part of a wider trend, acknowledging the power of simulations in studying complex problems (Peck, 2004; Buchanan, 2009; Epstein, 2009), and also due to technological advancements, making feasible what was once considered too computationally demanding.

Learning models typically involve two rules: a learning rule, used for updating newly acquired information in the learner's memory, and a decision rule, which uses the information stored in memory to choose between available alternatives (see Glossary for definition of these and other related terms). Although the two rules are at times considered to be one unit, each has a separate effect on behavior (as we discuss below) and therefore a distinction seems appropriate (see box 1 for a summary of commonly used rules). An explicit model, as referred to here, is a model that simulates an individual's learning and decision-making process in detail, tracking the flow of information and its effects on the individual's behavior step by step. In many cases, costs can be an emerging property of the learning process (e.g. the time it may take to sample the environment) rather than imposed externally as a deduction from the learners, total gains. In evolutionary simulations, learning and decision rules, or some of their parameters, are usually defined as genetically determined, and are the heritable variation on which natural (or more precisely modeled) selection operates, allowing the carriers of the most successful rules/parameter values to reproduce and increase their frequency in the simulated population.

Models studying the evolution of learning mechanisms via computer simulations can be roughly divided into three types, based on the question they are investigating. One is models concerned with a foraging problem, which can also be defined as a single- or multi-armed bandit (Robbins, 1952): learning to choose between a number of food alternatives (Beauchamp and Fernández-Juricic, 2004; van der Post and Hogeweg, 2006; Eliassen et al., 2007, 2009; Groß et al., 2008; Trimmer et al., 2012; Kolodny et al., 2014). This is perhaps the most ubiquitous problem in animal behavior, referring to a fundamental issue all organisms face; furthermore, this multiple choice scenario can, with slight modifications, refer to other resources sought by the organism (e.g. mates, nesting area, shelter).

The second type are models concerned with learning the evolutionarily stable strategy (ESS); these often involve a producer-scrounger game (Beauchamp, 2000; Hamblin and Giraldeau, 2009; Katsnelson et al., 2012; Afshar and Giraldeau, 2014). Learners in these models live in a social environment, where the outcome of their own behavior is affected by the behavior of others in the population. This presents a complex learning task due to a highly dynamic environment: learners acquire information about the behavior of others and consequently may modify their own behavior, but since this is true for all learners, previously acquired information may become useless as a result. The use of agent-based simulations in models of learning in a social context offers a double advantage, in capturing the stochasticity of both the learning process and of social interactions.

The third type are models combining features from the first and second types: foraging models taking place in a social environment (Beauchamp and Fernández-Juricic, 2004; van der Post and Hogeweg, 2009; van der Post et al., 2009; Arbilly et al., 2010, 2011; Rendell et al., 2010a; Arbilly and Laland, 2014). Individuals in

these models are learning to choose between a number of food alternatives, but their food-finding benefits directly depend on the behavior of others. This, for example, is the case of models studying the adaptive use of individual learning versus that of social learning: the goal is to learn where to find food, which can be done by either learning independently or by copying others' choices. An individual learning independently may pay some cost (e.g. time or energy), avoided by individuals who copy from it; copiers, on the other hand, depend on the availability of individuals to copy from and on the quality of their decisions (e.g. Rendell et al., 2010a).

In what follows, we will consider recent models, issues emerging from the modeling approach, the kind of results they offer, and their contribution to the understanding of the evolution of learning. Finally, we will outline some possible future directions. Since the focus of this review is explicit models as a tool in studying the evolution of learning, a discussion of general topics in the field of learning (e.g. Reinforcement Learning, Bayesian Learning) is outside its scope; the interested reader may find thorough reviews of these elsewhere (e.g. Sutton and Barto, 1998; Tenenbaum et al., 2006).

2 Learning does not Always Lead to the Best Decisions

A central contribution of explicit models to the study of the evolution of learning is in showing that learning does not always lead to the right answer, as often assumed. Simulating the learning process reveals that biases emerging from learning and decision rules may hamper a learner's route to making the optimal decision: these biases interfere with acquiring the full scope of information, especially in variable or dynamic environments. Furthermore, explicit models allow costs to arise directly from the cognitive process rather than have them modeled as an external property, thus unveiling their effects, which are often not trivial. Just like externally imposed costs, emerging costs result from the assumptions of the model, but since they are a result of the process they are, at least to some extent, less arbitrary. Still, some costs of learning, such as the costs of neural development and maintenance remain difficult to model as anything other than externally imposed.

One example of the effects of learning biases is the triumph of random foraging over learning when lifespan is short, due to learning biases (Arbilly et al., 2010). In an environment where the highest payoffs, on average, are also the most rare – due to depletion, for example – a learner searching for food in the best patch may not

find anything on its first attempt. On its next foraging attempts it is therefore not likely to search that patch again, a phenomenon known as risk aversion, or "the hot stove effect" (March, 1996; Denrell and March, 2001). An individual who does not learn, however, and forages at random, will not suffer this effect, and is thus more likely than the learner to visit the best patch again, increasing the probability of retrieving that high payoff early on. Eventually, given a flexible enough decision rule, learners will revisit the best patch, but if life is short, they are unlikely to ever forage there again, making learning an unsuccessful strategy (Arbilly et al., 2010; see further discussion in section 4).

Another obstacle for the evolution of learning may lie in the dynamics of the learning objective. Change has long been observed as the main reason for learning to evolve over any innate behavior (Stephens, 1991), but highly stochastic environments present a real challenge for learners, who invest in acquiring and storing information which may soon become useless. Indeed, simulation studies have shown a complex relationship between learning and ecological conditions (Eliassen et al., 2007, 2009). Frequency-dependent strategy games, where individuals learn which strategy to use, offer an even greater challenge: the acquired information is the behavior of others, who may be learners themselves and may change their behavior based on their own experience. Still, it has been shown that individuals in such games in the real world do learn and modify their strategy choice based on experience (Morand-Ferron and Giraldeau, 2010; Belmaker et al., 2012). How does learning evolve under these dynamic conditions?

Computer simulations of strategy learning in the producer-scrounger game have demonstrated that learning is far from being the most adaptive solution for finding the evolutionarily stable strategy (Katsnelson et al., 2012). When competing against innately fixed strategies, i.e., individuals who use the producer or scrounger strategy at a set frequency, learning only evolves given at least one of two conditions: (1) a high enough rate of environmental change, leading to a large enough change in the ESS; or (2) a high enough phenotypic asymmetry leading individuals to perform better as either producers or scroungers, therefore making it beneficial to learn your strength. If environmental change or phenotypic asymmetry are not high, the time invested in strategy-learning simply does not pay off. Interestingly, in finite populations where learning is favored, it appears to have an advantage over innate pure strategies simply because it may lead to a mixed strategy, and not neces-

sarily because it allows for better choice of strategy; this phenomenon (demonstrated also by the result that learning does not do better than an innate mixed strategy) is due to the 'flexibility' of mixed/learned strategies. While pure strategies may be selected against as their frequency increases, mixed strategies offer the ESS frequencies of the two strategies and therefore do not face this problem (Katsnelson et al., 2012). This model's findings contrast the results of a mathematical model (Dubois et al., 2010) that investigated the same question and found that learners will invade a population of non-learners, but will not reach fixation, suggesting that the ability of some individuals to learn prevents learning from evolving in others. There are several differences between the two models' assumptions, but perhaps most significantly, Dubois et al. assume idealized learning while Katsnelson et al. model the learning process explicitly; more than anything the dissonance between the two results emphasizes the implications of choosing either modeling approach (see discussion in Katsnelson et al., 2012).

3 Finding the Adaptive Value of Learning

As mentioned above, an explicit learning model includes a learning rule and a decision rule; however, it is not quite clear which rules are used by learners in the real world. Studied mostly in analytical models, in recent years there have also been some suggestions as to which learning rule is the most adaptive using explicit learning models. These studies either compared the success of learning and decision rules in a foraging context (Groß et al., 2008) or simulated strategy-learning in a game, pitting different learning rules against each other and analyzing their success (Beauchamp, 2000; Hamblin and Giraldeau, 2009).

Many explicit learning models use a linear operator rule (Bush and Mosteler, 1955) as their learning rule (Regelmann, 1984; Bernstein et al., 1988; Beauchamp, 2000; Beauchamp and Fernández-Juricic, 2004; Eliassen et al., 2007, 2009; Groß et al., 2008; Hamblin and Giraldeau, 2009; Arbilly et al., 2010, 2011; Katsnelson et al., 2012; Afshar and Giraldeau, 2014; Arbilly and Laland, 2014; Lotem and Biran-Yoeli, 2014). This rule updates memory using a moving average, weighing recent and past experience by a pre-defined memory factor (see box 1). While compatible with some real-world observations, its evolution remains unclear. Trimmer et al. (2012) performed a rigorous examination of the adaptive value of the Rescorla-Wagner rule, a modification of the linear operator rule for classical conditioning (Rescorla and Wagner, 1972). They used genetic programming to evolve learning rules that, like the Rescorla-Wagner rule, update the associative strength between a conditioned stimulus and an unconditioned stimulus. Comparing the properties of successful learning rules to the Rescorla-Wagner rule and to an optimally performing rule (which produces better estimates than the Rescorla-Wagner rule), the authors have been able to identify a number of characteristics favoring the Rescorla-Wagner rule under natural selection, suggesting an evolutionary explanation for its prevalence (Trimmer et al., 2012).

Learning rules offer a basic formulation of how a connection between a cue and a payoff is established. In the real world, organisms may learn far more complex relationships in the environment surrounding them, creating representations such as spatial maps of the environment and social hierarchies. A step towards understanding this learning of the "greater picture" was recently introduced in a model looking into the evolution of continuous learning – a learning strategy based on constant absorption of environmental information and identification of statistical regularities within it (Kolodny et al., 2014). This model compared, using agent-based simulations, the performance of three learning strategies, all constructed from the same basic building blocks of reinforcement learning: (1) Simple reinforcement learning – learning a simple association of a cue and primary reinforcer (e.g., if honey is the primary reinforcer, a beehive is a likely cue); (2) Reinforcement learning with chaining - where once a cue is associated with the primary reinforcer, this cue can act as a secondary reinforcer and other cues can be associated with it, creating a chain of cues (following the aforementioned example, a beehive may become a secondary reinforcer; tree holes, a likely location of beehives, may become a reinforcer following this association; foraging bees may later become yet another reinforcer in this chain); and (3) Continuous learning – where associations between all elements of the environment are learned regardless of a primary reinforcer, turning into a chain of reinforcers once the primary reinforcer is found (here, hollow trees, foraging bees, and all other elements in the environment – whether with some connection to honey (citrus blossom, dandelions) or irrelevant to it (cats and dogs) will be associated with each other based on co-occurrence. Once hollow trees are associated with honey, elements associated with hollow tress, elements associated with those elements, and so on, will form a chain of reinforcers). The model results suggest that

continuous learning is successful under specific conditions, such as short learning period, scarce food, and well-structured environments, which allow learners to predict the presence of food more than one step ahead (Kolodny et al., 2014). Using this modeling approach the authors have shown how a complex learning ability can be broken down into simple processes, thus suggesting a plausible evolutionary trajectory for its emergence. This is arguably a path theoreticians should consider in tackling the evolution of seemingly complex cognitive capacities.

4 The Problem of Sticky Learning

A recurring issue encountered in explicit models is the problem of "sticky learning", when learners stop sampling all alternatives in their learning environment and adhere to a single or a small set of choices instead (e.g. Groß et al., 2008; Hamblin and Giraldeau, 2009; Arbilly et al., 2010). This situation arises when learners experience a few alternatives as highly rewarding compared to the rest. As choosing the most rewarding alternative is the ultimate goal of learning, this may hardly seem like a problem, but in fact, what registers as the best alternative in the learner's experience may not be the best available alternative in the environment if the learner had limited sampling experience (Hamblin and Giraldeau, 2009), if alternatives offer variable payoffs and require repeated sampling to find their expected value (Arbilly et al., 2010), or if the environment is subject to within-generation temporal change (Groß et al., 2008).

This hurdle can be overcome in several ways. One is defining a sampling (or training) period which precedes the decision-making (or testing) period ("Random sampling rule" in Katsnelson et al., 2012). This allows learners to sample their environment without the biases that form with experience, but presents a problem in a changing environment where information decays over time. Another way to avoid stickiness is by using a learning rule which updates not only the learner's memory about its current choice, but also updates all other choices by a certain factor, improving their value over time and increasing their probability of being chosen (Groß et al., 2008). One can also use a rule such as the relative payoff sum rule, that sets the lowest threshold to which the value of each alternative can decrease to, diffusing, at least to some extent, the effect of repeated failure (see box 1). A fourth possibility is to use a decision rule which allows flexibility. Such is the exponential ratio decision rule, also known as the softmax rule (see Bereby-Meyer and Erev, 1998; Sutton and Barto, 1998 and box 1), which increases the probability of sampling alternatives that have been experienced as less attractive (e.g. Arbilly et al., 2010; Lotem and Biran-Yoeli, 2014). Each of these methods can be justified for different reasons, and choosing which one to use largely depends on the question and the model at hand.

It should be noted that the learner's reference point (also referred to as a-priori assumptions or expectation) of the environment, i.e. the value alternatives are assigned in the learner's memory before it acquires any experience, play an important role in how 'sticky' learning will be (Kahneman and Tversky, 1979). An "optimistic" reference point (Sutton and Barto, 1998; not to be confused with reproductive optimism discussed recently in the Behavioral Ecology literature, e.g. McNamara et al., 2011), that is high relative to the real payoffs will result in thorough sampling of the environment, as the relatively low experienced payoffs will drive the learner to try alternatives that it has not sampled yet and therefore are still perceived as offering higher payoffs. On the other hand, assigned value that is relatively low (a "pessimistic" reference point) is likely to lead to very little sampling, as the learner will perceive the alternative it sampled first as offering a relatively high payoff compared to all other alternatives (see March, 1996).

5 Uncovering the Trade-offs between Individual and Social Learning

Providing the basis for cumulative culture and the key to the enormous advances made by human societies, social learning and the conditions promoting it over individual learning have been, and still are, the subject of extensive theoretical work. Learning – both individual and social – has been mostly treated as idealized, with models often assuming that learners always choose the optimal behavior, but are (at times) burdened by external costs related to time, energy and the decay of information (e.g. Boyd and Richerson, 1985; Feldman et al., 1996; Aoki et al., 2005; Franz and Nunn, 2009; Kendal et al., 2009; Rendell et al., 2010b).

Aiming to identify when best to use social learning over individual learning, Rendell et al. (2010a) initiated a tournament of social learning strategies, yielding interesting insights not only in regard to the strategic use of social learning, but also to the nature of a successful social learning process. The strategies, entered by external participants, included a set of rules indicating when to learn individually, when to learn socially, and when to perform behaviors acquired through either learning process. They could also include rules as to how to

update newly acquired information in memory. Setting the strategies to compete against each other in an evolutionary agent-based simulation, Rendell et al. found that the most successful strategies relied heavily on social learning. These successful strategies had also employed a memory updating rule that accounted for decay of information in a changing environment; the winning strategy had gone further, discounting older information acquired socially, more so in a changing than in a stable environment (Rendell et al., 2010a).

The success of strategies which were heavily reliant on social learning in the tournament was due to the high payoff that socially learned behaviors resulted in: social learners copied behaviors that were performed by others in the population; performed behaviors were the best behaviors in learners' repertoire, thus representing the top tier of behaviors available (Rendell et al., 2010a). A different explicit model assisted in discovering a different mechanism providing an advantage to social learning over individual learning – a mechanism based on the risk aversion learners develop in environments with high variability (Arbilly et al., 2011). While risk aversion may prevent individual learners from foraging in the best patches (see discussion in section 2 above), a learning strategy that is based solely on social learning and disregards personal experience can circumvent this problem. Social learners are likely to learn from successful others rather than from individuals who failed to find food; this puts them at an advantage, since it means that when they join others, they always receive a payoff and therefore are not 'aware' of the variability in payoff within patches. In other words, social learners never experience a failure. Failure to find food in the highest paying patches are the reason learners do not realize these patches are the best foraging alternative; by learning only from others' success, social learners circumvent this problem. The visit frequency in the best patch of exclusive social learners will be much higher than that of individuals who rely on personal experience or combine both individually and socially acquired information, making it a much more successful learning strategy (Arbilly et al., 2011).

Further analysis, using the advantage of social learning in this type of risky environment to compare the performance of two social learning mechanisms, demonstrated how the quality of information acquired socially can modify the nature of information acquired individually, and how the two types of learning may co-evolve to compensate for each others' shortcomings (Arbilly and Laland, 2014).

6   Summary and Future Directions

Computer simulations are somewhere in between analytical models and real-world experiments: they are usually not as general as their analytical counterparts, and not as grounded as real-world experiments. However, as demonstrated in this review, their great power lies in the ability to reveal what might be happening "under the hood", the inner workings of the learning process.

Our ability to use simulations for that aim is limited, however, by our narrow observational perspective: we attempt to evolve learning mechanisms that generate the behaviors that we observe in the natural world. The underlying learning processes remain, therefore, a hypothesis. While a simulation producing real-world behaviors may indicate our hypotheses have merit, to gain a full understanding of the learning process we may benefit from looking into studies of learning in other fields of research. The best such example are experimental studies looking at learning at the mechanistic, neural level. These offer information regarding issues such as learning from positive versus negative reinforcement (e.g. Crestani et al., 1999; Frank et al., 2005), the mechanics of acquiring and extinguishing fear memory and its consequences for further learning (Maren and Quirk, 2004) and even individual differences in learning abilities (Schönberg et al., 2007). Evidence from such studies may inform the way we model learning processes, both in structure and in emphasis, and result in better model predictions – that can later be tested in behavioral experiments.

The relative realism of agent-based simulations can make them easy to translate into real-world experiments and test their assumptions and predictions (e.g. Amano et al., 2006; Ilan et al., 2013). While testing the predictions of evolutionary models may in some cases be impossible, it has already been shown that cognitive abilities in flies can undergo selection in the lab (Mery and Kawecki, 2002; Dunlap and Stephens, 2009). Such "mini brains" may offer a useful real-world framework, at least for some models (see Chittka and Skorupski, 2011).

Many of the explicit models mentioned above are concerned with a foraging problem. This is a classic learning problem, but learning is taking place in many other situations in animals' lives, and modeling other problems may shed light both on the evolution of learning in other contexts, and on the co-evolution of learning with other behaviors. For example, a recent explicit model of learning by nestlings begging in the nest provided new insight regarding the role of kin selection and

group selection in the parent-offspring conflict (Lotem and Biran-Yoeli, 2014). Explicit modeling of learning in behaviors such as mate choice, parental care and in host-parasite/predator-prey arms races (to name a few) may bring new understanding of observed phenomena and provide new directions for real-world experiments.

While there is still much to explore at the basic level, explicit models and the increasing accessibility of computational power also offer us the opportunity to take learning models to the next level. Perhaps, in the spirit of "Integrating mechanism and function" (McNamara and Houston, 2009), and "Exposing the behavioral gambit" (Fawcett et al., 2013), we should begin to consider modeling cognitive capacities of a higher order, building them up from the simple processes they are likely to be made of. Most importantly, an explicit modeling approach might benefit our understanding of processes that are considered a 'black box' – where our knowledge, at best, scratches the surface. Examples of such processes include Theory of Mind – the ability to attribute knowledge, intentions and beliefs to others; and mental time travel – the ability to recollect past events or anticipate future needs. While the distance between the simple learning rules described in this review and advanced abilities such as Theory of Mind and mental time travel seems irreconcilable, the apparent complexity of these abilities suggests that they have evolved gradually rather than by way of a single mutation. Modeling these abilities from the bottom up, using simple learning rules as building blocks, is perhaps the most interesting challenge facing the field today – and one that may provide great insight into the mind and its evolution.

References

Afshar M, Giraldeau L-A, 2014. A unified modelling approach for producer-scrounger games in complex ecological conditions. Anim. Behav. 96:167–176.

Amano T, Ushiyama K, Moriguchi S, 2006. Decision-making in group foragers with incomplete information: Test of individual-based model in geese. Ecol. Monogr. 76:601–616.

Aoki K, Wakano JY, Feldman MW, 2005. The emergence of social learning in a temporally changing environment: A theoretical model. Curr. Anthropol. 46: 334–340.

Arbilly M, Laland KN, 2014. The local enhancement conundrum: In search of the adaptive value of a social learning mechanism. Theor. Popul. Biol. 91: 50–7.

Arbilly M, Motro U, Feldman MW, Lotem A, 2010. Co-evolution of learning complexity and social foraging strategies. J. Theor. Biol. 267: 573–81.

Arbilly M, Motro U, Feldman MW, Lotem A, 2011. Evolution of social learning when high expected payoffs are associated with high risk of failure. J. R. Soc. Interface 8: 1604–1615.

Bateson M, Healy SD, Hurly TA, 2002. Irrational choices in hummingbird foraging behaviour. Anim. Behav. 63: 587–596.

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.

Beauchamp G, Fernández-Juricic E, 2004. The group-size paradox: effects of learning and patch departure rules. Behav. Ecol. 16: 352–357.

Belmaker A, Motro U, Feldman MW, Lotem A, 2012. Learning to choose among social foraging strategies in adult house sparrows *Passer domesticus*. Ethology 118: 1111–1121.

Bereby-Meyer Y, Erev I, 1998. On learning to become a successful loser: A comparison of alternative abstractions of learning processes in the loss domain. J. Math. Psychol. 42: 266–86.

Bernstein C, Kacelnik A, Krebs JR, 1988. Individual decisions and the distribution of predators in a patchy environment. J. Anim. Ecol. 57: 1007–1026.

Boyd R, Richerson PJ, 1985. Culture and the Evolutionary Process. Chicago: University of Chicago Press.

Buchanan M, 2009. Meltdown modelling. Nature 460: 680–682.

Bush RR, Mosteler F, 1955. Stochastic Models for Learning. Oxford: John Wiley & Sons, Inc.

Charnov EL, 1976. Optimal foraging, the marginal value theorem. Theor. Popul. Biol. 9: 129–136.

Chittka L, Skorupski P, 2011. Information processing in miniature brains. Proc. R. Soc. B Biol. Sci. 278: 885–8.

Crestani F, Lorez M, Baer K, Essrich C, 1999. Decreased GABA$_A$- receptor clustering results in enhanced anxiety and a bias for threat cues. Nat. Neurosci. 2: 833–839.

Dall SRX, Giraldeau L-A, Olsson O, McNamara JM, Stephens DW, 2005. Information and its use by animals in evolutionary ecology. Trends Ecol. Evol. 20: 187–93.

Denrell J, March JG, 2001. Adaption as information restriction : The hot stove effect. Organ. Sci. 12: 523–538.

Dridi S, Lehmann L, 2014. On learning dynamics underlying the evolution of learning rules. Theor. Popul. Biol. 91: 20–36.

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 Biol. Sci. 277: 3609–16.

Dunlap AS, Stephens DW, 2009. Components of change in the evolution of learning and unlearned preference. Proc. R. Soc. B Biol. Sci. 276: 3201–3208.

Eliassen S, Jørgensen C, Mangel M, Giske J, 2007. Exploration or exploitation: Life expectancy changes the value of learning in foraging strategies. Oikos 513–523.

Eliassen S, Jørgensen C, Mangel M, Giske J, 2009. Quantifying the adaptive value of learning in foraging behavior. Am. Nat. 174: 478–489.

Epstein JM, 2009. Modelling to contain pandemics. Nature 460: 687.

Epstein JM, Axtell RL, 1996. Growing Artificial Societies: Social Science from the Bottom Up. Washington: Brookings Institution Press.

Fawcett TW, Fallenstein WB, Higginson AD, Houston AI, Mall-

press DEW et al., 2014. The evolution of decision rules in complex environments. Trends Cogn. Sci. 18: 153–61.

Fawcett TW, Hamblin S, Giraldeau L-A, 2013. Exposing the behavioral gambit: The evolution of learning and decision rules. Behav. Ecol. 24: 2–11.

Feldman MW, Aoki K, Kumm J, 1996. Individual versus social learning: Evolutionary analysis in a fluctuating environment. Anthropol. Sci. 104: 209–231.

Frank MJ, Woroch BS, Curran T, 2005. Error-related negativity predicts reinforcement learning and conflict biases. Neuron 47: 495–501.

Frankenhuis WE, Del Giudice M, 2012. When do adaptive developmental mechanisms yield maladaptive outcomes? Dev. Psychol. 48: 628–42.

Franz M, Nunn CL, 2009. Rapid evolution of social learning. J. Evol. Biol. 22: 1914–1922.

Groß R, Houston AI, Collins EJ, McNamara JM, Dechaume-Moncharmont F-X et al., 2008. Simple learning rules to cope with changing environments. J. R. Soc. Interface 5: 1193–202.

Hamblin S, Giraldeau L-A, 2009. Finding the evolutionarily stable learning rule for frequency-dependent foraging. Anim. Behav. 78: 1343–1350.

Harley CB, 1981. Learning the evolutionarily stable strategy. J. Theor. Biol. 89: 611–633.

Houston AI, Sumida BH, 1987. Learning rules, matching and frequency dependence. J. Theor. Biol. 126: 289–308.

Ilan T, Katsnelson E, Motro U, Feldman MW, Lotem A, 2013. The role of beginner's luck in learning to prefer risky patches by socially foraging house sparrows. Behav. Ecol. 24: 1398–1406.

Kacelnik A, Bateson M, 1997. Risk sensitivity: Crossroads for theories of decision-making. Trends Cogn. Sci. 1: 304–309.

Kahneman D, Tversky A, 1979. Prospect theory: An analysis of decision under risk. Econometrica 47: 263–291.

Katsnelson E, Motro U, Feldman MW, Lotem A, 2012. Evolution of learned strategy choice in a frequency-dependent game. Proc. R. Soc. B Biol. Sci. 279: 1176–84.

Kendal J, Giraldeau L-A, Laland K, 2009. The evolution of social learning rules: Payoff-biased and frequency-dependent biased transmission. J. Theor. Biol. 260: 210–219.

Kolodny O, Edelman S, Lotem A, 2014. The evolution of continuous learning of the structure of the environment. J. R. Soc. Interface 11: 20131091.

Lotem A, Biran-Yoeli I, 2014. Evolution of learning and levels of selection: A lesson from avian parent-offspring communication. Theor. Popul. Biol. 91: 58–74.

March JG, 1996. Learning to be risk averse. Psychol. Rev. 103: 309–319.

Maren S, Quirk GJ, 2004. Neuronal signalling of fear memory. Nat. Rev. Neurosci. 5: 844–852.

Marsh B, Kacelnik A, 2002. Framing effects and risky decisions in starlings. Proc. Natl. Acad. Sci. U. S. A. 99: 3352–5.

McNamara JM, Dall SRX, 2010. Information is a fitness enhancing resource. Oikos 119: 231–236.

McNamara JM, Houston AI, 1987a. Memory and the efficient use of information. J. Theor. Biol. 125: 385–95.

McNamara JM, Houston AI, 2009. Integrating function and mechanism. Trends Ecol. Evol. 24: 670–5.

McNamara JM, Houston AI, 1987b. Memory and the efficient use of information. J. Theor. Biol. 125: 385–395.

McNamara JM, Houston AI, 1985. Optimal foraging and learning. J. Theor. Biol. 117: 231–249.

McNamara JM, Trimmer PC, Eriksson A, Marshall JAR, Houston AI, 2011. Environmental variability can select for optimism or pessimism. Ecol. Lett. 14: 58–62.

Mery F, Kawecki TJ, 2002. Experimental evolution of learning ability in fruit flies. Proc. Natl. Acad. Sci. U. S. A. 99: 14274–14279.

Morand-Ferron J, Giraldeau L-A, 2010. Learning behaviorally stable solutions to producer-scrounger games. Behav. Ecol. 21: 343–348.

Peck SL, 2004. Simulation as experiment: A philosophical reassessment for biological modeling. Trends Ecol. Evol. 19: 530–4.

Regelmann K, 1984. Competitive resource sharing: A simulation model. Anim. Behav. 32: 226–232.

Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K et al., 2010a. Why copy others? Insights from the social learning strategies tournament. Science 328: 208–213.

Rendell L, Fogarty L, Laland KN, 2010b. Rogers' paradox recast and resolved: Population structure and the evolution of social learning strategies. Evolution 64: 534–48.

Rescorla RA, Wagner AR, 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In: Black AH, Prokasy WF eds. Classical Conditioning II: Current Research and Theory. New York: Appleton-Century-Crofts, 64–99.

Robbins H, 1952. Some aspects of the sequential design of experiments. Bull. Am. Math. Soc. 58: 527–535.

Schönberg T, Daw ND, Joel D, O'Doherty JP, 2007. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. J. Neurosci. 27: 12860–7.

Schuck-Paim C, Pompilio L, Kacelnik A, 2004. State-dependent decisions cause apparent violations of rationality in animal choice. PLoS Biol. 2: e402.

Shafir S, Waite TA, Smith BH, 2002. Context-dependent violations of rational choice in honeybees *Apis mellifera* and gray jays *Perisoreus canadensis*. Behav. Ecol. Sociobiol. 51: 180–187.

Shettleworth SJ, 2010. Cognition, Evolution and Behavior. 2nd edn. New York: Oxford University Press.

Stephens DW, 1991. Change, regularity, and value in the evolution of animal learning. Behav. Ecol. 2: 77–89.

Stephens DW, 1989. Variance and the value of information. Am. Nat. 134: 128–140.

Stephens DW, 1993. Learning and behavioral ecology: Incomplete information and environmental predictability. In: Papaj DR, Lewis AC eds. Insect Learning: Ecology and Evolutionary Perspectives. London: Chapman & Hall, 195–218.

Stephens DW, Krebs JR, 1986. Foraging Theory. Princeton: Princeton Univesity Press.

Sutton RS, Barto AG, 1998. Reinforcement Learning: An Introduction. Cambridge, Massachusetts: MIT Press.

Tenenbaum JB, Griffiths TL, Kemp C, 2006. Theory-based Bayesian models of inductive learning and reasoning. Trends Cogn. Sci. 10: 309–18.

Tracy ND, Seaman JW, 1995. Properties of evolutionarily stable learning rules. J. Theor. Biol. 177: 193–198.

Trimmer PC, McNamara JM, Houston AI, Marshall JAR, 2012. Does natural selection favour the Rescorla-Wagner rule? J.

Theor. Biol. 302: 39–52. Elsevier.

Van der Post DJ, Hogeweg P, 2006. Resource distributions and diet development by trial-and-error learning. Behav. Ecol. Sociobiol. 61: 65–80.

Van der Post DJ, Hogeweg P, 2009. Cultural inheritance and diversification of diet in variable environments. Anim. Behav. 78: 155–166.

Van der Post DJ, Ursem B, Hogeweg P, 2009. Resource distributions affect social learning on multiple timescales. Behav. Ecol. Sociobiol. 63: 1643–1658.

Box 1: Learning and Decision Rules in Explicit Learning Models

1 Learning Rules

Rules are given for updating the memory value M of alternative i with payoff Y, received for alternative i at time t.

The linear operator

$$M_{i,t} = \alpha M_{i, t-1} + (1 - \alpha)Y_{i, t}$$

Where $0 < \alpha < 1$ is the memory factor, the weight given to past payoffs compared to the most recent (Bush and Mosteler, 1955; McNamara and Houston, 1987a).

Relative payoff sum

$$M_{i,t} = \alpha M_{i, t-1} + (1 - \alpha)r_i + Y_{i, t}$$

Where $0 < \alpha < 1$ is the memory factor, the weight given to past payoffs compared to the most recent, and r_i is the residual of alternative i – the lowest value M_i goes to in the case of repeated failures in alternative i (Harley, 1981; but see Tracy and Seaman, 1995).

Perfect memory

$$M_{i, t} = \frac{\alpha + S_{i,t}}{\beta + N_{i,t}}$$

Where α and β are constants, $S_{i,t}$ is the sum of all payoffs received for alternative i until time t, and $N_{i,t}$ is the number of times alternative i has been chosen until time t (Houston and Sumida, 1987).

2 Decision Rules

Rules specify the probability p of choosing alternative i at time $t+1$, from alternatives $j = 1 \ldots k$.

Matching rule

$$P_{i, t+1} = \frac{M_{i,t}}{\sum_{j=1}^{k} M_{j,t}}$$

Maximize rule

This rule dictates choosing the alternative that has the highest value at time t.

Softmax / Exponential ratio rule

$$P_{i, t+1} = \frac{e^{M_{i,t}}}{\sum_{j=1}^{k} e^{M_{j,t}}}$$

Glossary

Cue: The media encoding the information (McNamara and Dall, 2010).

Learning: A change in behavior due to experience (Shettleworth, 2010). In this review, the term is used pragmatically to refer to the process of updating information in memory.

Decision-making: The process of choosing between alternatives.

Individual learning: The acquisition of information through direct interaction with the environment (Dall et al., 2005).

Social learning: The acquisition of information by observing the behavior of others.

Bias: A tendency towards choosing one alternative over the other.

Risk: Unpredictable variation in outcomes (Kacelnik and Bateson, 1997); while the term 'risk' may at times refer to risk of predation, its common use in the literature is in reference to the possibility of obtaining (or failure to obtain) a payoff.

Explicit modeling: detailed simulation of a process (rather than assuming an end result). For example, an explicit model of learning simulates the process of sampling, updating the information in memory and its use in later decision-making.