TASK SWITCHING AND NATURAL PROJECTIBILITY

CHRISTIAN TORSELL

Abstract. A persistent problem in the philosophy of induction concerns the variety of analogies among experiences that might guide one's learning. Given that any observational record will exhibit a multiplicity of regularities, which ones should we attend to in forming empirical beliefs and expectations? Call this the *projectibility problem.* Nelson Goodman (1983) dramatized the problem with his "new riddle of induction," and David Lewis (1969) recognized the challenges it raises for the possibility of establishing conventions by precedent. Learners in nature face a version of the projectibility problem inasmuch as they must learn which regularities in their environments are useful guides for action and prediction in order to survive. The present paper considers one method by which this problem might be solved: trial-and-error learning. The strategy is to begin with an experiment from comparative psychology in which subjects face a version of the projectibility problem, then to develop a concrete model of a learner that can solve that problem. In the model, the dispositions of three separate modules, or *subagents*, responsible for the learner's attentional and behavioral responses coevolve under a simple reinforcement learning dynamics. On simulation, the model reliably learns to attend to the practically relevant regularities in its environment.

1. Introduction

Many models of learning assume an agent facing a sequence of identical trials of a well-defined decision problem. But the practical lives of humans and other natural learners are rarely, if ever, so simple. When we learn in the course of repeated experience in a certain kind of problem, the individual instances we encounter are embedded in a varied stream of choice situations presented by our environment. And in navigating these situations, nature does not specify a unique basis on which to make judgments of similarity between them. No two choice settings will be identical in every respect, and any two choice settings will be analogous to each other in some respects. We must *learn* which observable similarities between different choice settings are indicative of a common underlying structure, such that treating those settings as identical for the purposes of learning and action is conducive to practical success.

The core of this idea appears in David Lewis' discussion of the role of precedent in establishing conventions. Lewis notes that, whenever we are guided by some particular analogy between a present situation and a past situation in classifying both as instances of the same kind of coordination problem, there will be "innumerable alternative analogies" we could have used to classify them instead. He writes: "Were it not that we happen uniformly to notice some analogies and ignore others—those we call 'natural' or 'artificial,' respectively—precedents would be completely ambiguous and worthless" (Lewis 1969, 38).

Date: December 10, 2025.

I would like to thank Jeff Barrett, Brian Skyrms, Cailin O'Connor, audiences at the Self-Assembling Games Workshop and Social Dynamics Seminar at UC Irvine, and the reviewers who read an earlier version of this paper for their comments and suggestions.

Lewis' discussion echoes Nelson Goodman's famous treatment of projectibility and the "new riddle of induction." Goodman highlighted that the inductive inferences we draw from our observations depend on the similarities we are looking for. He dramatized the problem by introducing the predicate *grue*, which applies to an object just in case it is green and observed before some future time t or blue and not observed before t. According to a standard model of enumerative induction, the statements "All emeralds are grue" and "All emeralds are green" are equally confirmed by a collection of pre-t observations of green emeralds, but they make conflicting predictions about future observations.

More generally, for any body of observations, we can use gerrymandered predicates to construct conflicting universal generalizations of which those observations count as instances. Goodman summarizes the problem memorably: "To say that valid predictions are those based on past regularities, without being able to say which regularities, is quite pointless. Regularities are where you find them and you can find them anywhere" (Goodman 1983, 81). It is now well known that Goodman's recognition of the dependence of induction on the scheme guiding a learner's judgments of similarity and identity of observations was prefigured by both Hume and De Finetti (see Skyrms 2012).

Both Lewis and Goodman (along with Hume and De Finetti) clearly recognize the central role of judgments of similarity in precedent-setting and inductive inference, respectively. And both see that there are difficulties raised by the multiplicity of available bases for those judgments. But neither has a compelling story about how inductive learners or precedent-followers might come to notice and condition their action on the relevant similarities. What is needed is an account of how agents might *learn* to attend to practically relevant regularities in their experience in order to learn and act successfully.

In a critical commentary on Goodman's treatment of projectibility, Patrick Suppes comes closer to framing the problem in our terms. Suppes highlights that Goodman's grue argument only applies to a very special kind of learning situation. He writes:

> Artificial predicates like grue . . . are no problems for animals, because they are not considered, but projectibility is. Much learning occurs in relation to features of the environment that are partly stationary, and those stationary features have nice properties of projection. What is important about learning, however, is dynamic adaptability. . . . For animals that want to survive, there is a problem of projectibility, but it is not exactly Goodman's problem. Can what they have learned in the past be flexible enough to permit them to adapt to a changing environment in the future? (Suppes 1994, 264)

In his treatment of projectibility, Suppes turns our attention to learning as it occurs in nature. Animals inhabit environments that present a variety of recurrent choice situations and which include many variable features learners might attend to in distinguishing between these situations. Successful action is possible only if they attend to features that are informative about the conditions for practical success in the problems they frequently encounter, and if they do not overgeneralize what they learn in one setting such that they cannot adjust to changing conditions in the future. The present paper considers how natural learners might solve a version of this problem by means of a simple kind of trial-and-error learning.

Experiments on *task switching* offer examples of animals learning to notice some similarities among choice settings and ignore others in navigating a changing environment.[1] In task

[1] See Monsell (2003) and Kiesel et al. (2010) for an overview.

switching problems, a subject is initially trained to perform two different tasks individually. Then, they face a series of trials in which they are sometimes required to perform the first task and sometimes required to perform the second task. A subject can reliably succeed in these problems only if she learns to flexibly toggle between distinct sets of dispositions, each appropriate to one task, depending on the state of a designated task cue. This, in turn, requires that the subject notice and condition her action on the task cue, rather than ignore it or attend to some other, uninformative feature of her environment to decide which task to execute.

We model how an agent might learn both to execute and distinguish between tasks in a simulated training environment based on experiments on task switching in rhesus macaques reported in Avdagic et al. (2014). In the model, learning is characterized in terms of the interactions among three subagents who are responsible for determining the agent's attentional and behavioral responses. The subagents' interests are aligned: each gets a positive payoff only when the learning system comprising all of them successfully performs the relevant task. The subagents' dispositions evolve by a simple form of reinforcement learning.

The goal is to show how the problem of projectibility illustrated by task switching problems might be solved by low-rationality means widely available to natural learners. As is discussed in greater detail in Section 3, we are thus concerned to invoke a psychologically plausible and cognitively undemanding model of learning. We do not, however, claim to be describing how the monkeys in fact learned in Avdagic et al.'s experiments. They may well have made use of more sophisticated (or, at any rate, qualitatively different) cognitive machinery than we allow in the model. What matters for our purposes is that the simple, naturalistically motivated dynamics we consider is *enough* to learn to manage the special demands of task switching problems.

As the modeled agent's attention gradually comes to be focused on the practically relevant regularities in her environment, she becomes a more effective learner in the problems she repeatedly faces. The phenomenon modeled here is thus an instance of *metalearning*, or *learning how to learn*, in which an agent's method of learning is itself adaptively modified by experience (see, e.g., Harlow (1949), Torsell and Barrett (2024), Barrett (2023)). Moreover, the basic mechanism behind the metalearning is a form of higher-order reinforcement learning, in which reinforcements at the attentional level shape the agent's dispositions to choose different task-specific behavioral strategies. Models like this, in which reinforcements accrue to full behavioral strategies rather than individual acts, are not without precedent. Erev and Barron's (2005) RELACS (REinforcement Learning Across Cognitive Strategies) is a well-known example. In their model, agents choose from among three decision rules to guide their first-order choices in an iterated binary decision problem, with their propensities over decision rules evolving by reinforcement learning. In a related project, Dahl (1996), building on Rosenthal's (1993) work on "rules of thumb" in games, considers reinforcement learning over strategic heuristics in a p-beauty contest, where the rules under consideration belong to the family of "k-step" best-response rules developed in Nagel (1995). In both Dahl's and Erev and Barron's models, the first-order strategies over which the agents learn are fully specified in advance by the modeler. By contrast, in our model, the task-specific behavioral strategies themselves emerge from a reinforcement learning process.

2. Monkey Learning

Avdagic et al. (2014) investigate task switching in rhesus macacques facing a series of irregularly shifting *simultaneous chaining* (SimChain) tasks. On each trial of a SimChain task, a subject is presented with a collection of stimulus items which vary along some discriminable dimension (e.g., size or color). If the subject successively selects (e.g., by touching or pointing) the individual items in an order reflecting their ranking with respect to the designated dimension, then it receives a reward. The reward condition may require the subject to select the items in either ascending or descending order. A SimChain task may, for example, require subjects to successively touch individual colored cards in light-to-dark order in order to receive a food reward. Subjects are only rewarded upon successfully ordering the entire collection of items. No rewards are received in trials in which the subject selects some but not all items in the task-appropriate order.

Avdagic et al.'s experiments presented three rhesus monkeys—named Lashley, Oberon, and MacDuff—with a series of trials which shifted randomly between two SimChain tasks. In both tasks, a subject was presented simultaneously with k many stimulus items on a screen ($3 \leq k \leq 6$). The items were randomly scattered spatially, and each consisted of a white rectangle with a grey circle in the middle. The circles varied with respect to two discriminable dimensions: luminosity (ranging from light grey to black) and radius. In the first task, subjects were rewarded for successively touching the stimuli in an order reflecting their luminosity ranking. Call this the *luminosity task*. In the second task, subjects were rewarded for ranking the stimuli according to radius. Call this the *radius task*.

With respect to the luminosity task, we call luminosity the *target dimension* and radius the *distractor dimension*. These designations are flipped for the radius task. In both tasks, Lashley and Oberon were rewarded for making selections reflecting increasing order (darkest-to-brightest in the luminosity task; smallest-to-largest in the radius task), while MacDuff was rewarded for selections reflecting decreasing order (brightest-to-darkest or largest-to-smallest).

The *training stage* of Avdagic et al.'s experiment consisted of three phases. In each phase, stimulus items were always presented on a *blue* background for the luminosity task trials and on a *red* background for the radius task trials. In the first phase, each monkey faced only radius task trials, and the task was simplified by allowing stimulus items to vary only with respect to radius (the reward-relevant dimension)—luminosity was held constant across items. If a subject correctly ranked the items according to the radius, it received a food reward immediately after the final item was selected. If it selected an item out of order, the trial was ended immediately and the subject experienced a timeout.

Each subject first faced repeated sessions of 50 three-item trials until it reached an 80 percent performance criterion (i.e., until it succeeded on 40 of the 50 trials in a session), then progressed to sessions of four-item trials, working up to six-item trials with the number of items increasing by one each time the subject reached the 80 percent criterion. This phase ended when each subject reached a 50 percent performance criterion on six-item trials.[2] The second training phase proceeded exactly like the first but with the luminosity task as the target task; here, stimuli varied with respect to luminosity but not radius.

In the third training phase, trials randomly switched between the luminosity task and the radius task, each task being chosen with equal probability on every trial. As in the first

[2] Avdagic et al. do not explain why the performance criterion changed for the final segment of this phase.

and second phases, the stimulus items presented varied only with respect to the reward-relevant dimension for a given trial. The monkeys had to learn to flexibly switch between the patterns of behavior they had learned for the luminosity task and the radius task to consistently receive rewards. They could achieve this either by conditioning their choices on the color of the background or on which feature varied across objects. Again, sessions of 50 trials were repeated until an 80 percent performance criterion was reached. But in this phase there was no progression from smaller to larger collections of stimulus items. Instead, the number of stimulus items varied between three and six from trial to trial in every session.

The training stage was followed by the *experimental stage*, in which stimulus items varied in both dimensions in every trial. The experimental stage consisted of two phases. In the first phase (lasting 36 sessions, or 1800 trials), subjects faced *biased* versions of the luminosity task and the radius task, in which the differences between stimulus items with respect to the distractor dimension were reduced relative to differences in the target dimension. This was meant to increase the salience of the target dimension. Trials were randomly and uniformly distributed with respect to task type (A or B) and number of items to be ranked (three or four). The second phase (lasting 48 sessions, or 2400 trials) proceeded exactly like the first but without bias in the magnitudes of the differences among stimuli along the two dimensions.

All three of Avdagic et al.'s subjects successfully completed all phases of the training stage. However, the authors do not provide data on the monkeys' performance in this stage. In the experimental stage, all three monkeys succeeded on more than 75% of three-item problems (Avdagic et al.). Their performance on four-item problems varied considerably between the first and second phases of the experimental stage. For four-item problems in the first phase, the relative frequency of successful trials ranged from 39.2% (MacDuff) to 58.8% (Lashley) for radius trials and from 32.6% (MacDuff) to 40% (Lashley) for luminosity trials. All subjects were significantly more accurate than chance performance of 1.7% (Avdagic et al. 623-4). In the second experimental phase, performance on four-item problems was less accurate but still considerably better than chance. Accuracy ranged from 19.8% to 43.5% on luminosity trials and from 13.5% to 20.8% on radius trials. As in the first phase, Lashley performed most accurately on both tasks; MacDuff performed least accurately on second-phase luminosity problems and Oberon set the low bar for radius problems.

What is important for our purposes is the basic structure of the problem the monkeys faced and the means Avdagic et al. furnished for them to solve that problem. Consider what the monkeys must learn in the training phase in order to perform accurately in the experimental stage. First, they must learn to perform each of the simultaneous chaining tasks individually. Second, they must learn to notice and condition their final behavioral response on the background color of the screen. Third, they must learn to associate each background color with the appropriate task, i.e. to attend to the size of the stimulus items when the background is red and to attend to their brightness when the background is blue. Success in Avdagic et al.'s task switching setup requires coordinating these attentional and behavioral dispositions to enable the learner to accurately and flexibly switch between task-specific sets of responses when called for by changes in their environment. In training phases 1 and 2, they have the opportunity to learn each task individually in the presence of the cue used to signal task switches in later stages (background color), gradually moving through problems involving successively larger collections of items. And in phase 3, they have a chance to practice sequences of randomly-switching problems.

3. Reinforcement Learning

To model the kind of learning accomplished by Avdagic et al.'s monkeys, we need to specify a learning dynamics. The family of reinforcement learning rules native to the social and behavioral sciences offers a promising pool of candidates.[3] The leading idea that these dynamics formalize is that the probability with which an agent chooses a given type of action depends on the history of associations between past actions of that kind and positive or negative outcomes that closely followed those actions. Edward Thorndike, a pioneer in experimentally investigating learning in animals, captured this idea in his *law of effect*:

> Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. (Thorndike 1911, 244)

In psychology and, later, economics, a variety of mathematical models of learning formalizing Thorndike's basic idea have been developed (see, e.g., Bush and Mosteller 1955, Roth and Erev 1995). These reinforcement learning models have a few significant virtues. There is strong empirical evidence that, in many contexts, both human and animal learning obeys the law of effect (Roth and Erev 1995; Erev and Roth 1998; Herrnstein 1970). Another virtue is that these rules typically describe very simple patterns of adaptive behavior that can be implemented with minimal cognitive machinery. What's more, even very primitive forms of reinforcement learning are efficacious in a large class of learning problems (see Beggs 2005, Argiento et al. 2009).

Here, we use *Moran learning*, a relatively underexplored reinforcement learning dynamics introduced in Barrett (2006), to model the monkeys' learning. This dynamics treats learning as a *Moran process*, a kind of stochastic process commonly used in biological models of evolution in finite populations (see Moran 1958). Moran learning can be modeled with balls and urns. Suppose an agent repeatedly faces a choice problem in which it must choose one of k many acts $a_1, a_2, ..., a_k$. The agent chooses by drawing a ball at random and without bias from an urn containing balls labeled with numbers from 1 to k. If it draws a ball with an i label ($1 \leq i \leq k$), then it chooses a_i.

Suppose the agent faces n trials of the choice problem, and let t_j stand for the jth repetition in the sequence. Suppose further that the urn initially contains c many balls. We assume that the outcomes of the learner's choice are bivalent: a given choice results in either *success* or *failure*. If the agent is a Moran learner, then it updates the urn's contents in the following way. If the agent chooses a_i in t_j and succeeds, then it first *removes* r ($r < c$) many balls at random.[4] Then r many balls of type i are added to the urn. If instead the agent's choice of a_i in t_j results in failure, then the contents of the urn remain unchanged. If the agent learns by Moran learning *with punishment*, then successful actions are reinforced in

[3] The label "reinforcement learning" subsumes a variety of models of learning (see, e.g. Sutton and Barto 2018). Here we are concerned with reward-based reinforcement learning models of the sort psychologists and economists have developed to model associative learning in which choice behavior is shaped by rewards and punishments.

[4] We assume the urn is well-mixed so that, for each ball removed and each label-type i, the probability that that ball is an i-type is equal to the proportion of i-type balls in the urn.

the way just described, but if a chosen action a_i results in failure, the learner first removes p many balls of the unsuccessful type (where $p < c$) and then adds p many balls such that for each type l, the probability that a given newly-added ball is an l-type is equal to the proportion of l balls in the urn after the initial removal of balls.

The variant of Moran learning with punishment used in our model involves a small modification to the description above. If removing balls in either a reinforcement or punishment event reduces the number of balls of some type in the urn to zero, then a new ball of that type is added and a random ball of a different type (which has more than one representative in the urn) is removed. This ensures that no act type goes to extinction, i.e., that the agent always chooses each act with positive (but possibly very small) probability.

A key motivation for using Moran learning in our context concerns its treatment of *forgetting*. A learning dynamics is *forgetful* if more recent experience exerts a greater influence than less recent experience in determining the agent's choice dispositions.[5] There are numerous ways to introduce forgetting into a reinforcement learning dynamics. One way involves building in an adjustable reference point, where forgetting is introduced by allowing the learner to "get used to" a given payoff level, such that the magnitude of the agent's reinforcement on a given action-outcome pair depends on how often outcomes with similar payoffs have occurred in the recent past.[6] In this model, the risk of lock-in on suboptimal dispositions is countered by the exploration encouraged by the learner's tendency to treat a given objective payoff level as less valuable (and possibly eventually as a positive *loss*) as it comes to reliably receive payoffs at or above that level. Other implementations of forgetfulness involve periodically throwing out a certain number of randomly selected balls from the learner's urn.

Under Moran reinforcement learning, forgetting is implemented by allowing reinforcement events to replace a constant *proportion* of the learner's urn's contents, rather than simply adding a constant number of balls upon success. This allows a Moran learner to adapt its dispositions to new practical contexts more nimbly than in models in which propensities are unbounded and reinforcements matter cumulatively for choice probabilities (as in, e.g., Roth and Erev's (1995) model). What recommends this learning rule for our purposes is just that it exhibits the basic structure of reinforcement learning with forgetting. The particular mechanism by which forgetfulness is implemented is not theoretically significant. That said, there are pragmatic considerations that favor Moran learning. In particular, it is a relatively simple learning rule, having just three free parameters[7] (the reinforcement and punishment levels, along with the capacity of the urn) with straightforward psychological interpretations. And Moran processes are well understood mathematically and familiar across a range of statistical, biological, and social scientific disciplines. An interesting extension of our project would consider how our model would perform under other forgetful dynamics.

Forgetting is particularly useful for reinforcement learners in task switching problems. In these problems, the practical demands confronting an agent change often, despite that agent facing a qualitatively similar environment in each trial. Of course, the risk of premature lock-in at the level of task-specific behavioral dispositions is mitigated by the availability of a task cue the learner can use to choose which such dispositions to activate on a given trial. But there remains a serious risk of such lock-in at the level of attentional dispositions,

[5] For an extensive discussion of the role of forgetting in learning, see Barrett and Zollman 2009.

[6] See Bereby-Meyer and Erev (1998) for discussion of this kind of learning.

[7] This assumes the urn's initial contents are fixed.

assuming that those dispositions also evolve by reinforcement learning. In the lab, subjects are typically trained on each task individually before facing mixed trials, and there is no extrinsic reward for attending to the task cue in early single-task training stages. If a learner habitually ignores the task cue in those stages, then without some form of forgetting, the force of past successes may make it very difficult to correct course and eventually learn to attend to that cue.

4. The Model

The model of the monkeys' learning involves three *subagents*, each corresponding to a distinct functional role in determining the learner's final behavioral output. The first is the *metasalience controller.* Recall that in the Avdagic et al. experiment, the background color of the screen in a given trial reliably indicates which task the monkey will be rewarded for executing. If the monkey attends to this feature and notices when it changes, it may learn to switch between tasks precisely when necessary to consistently receive rewards. But it doesn't *have to* pay attention to background color. It might ignore the screen's color on any given trial. In the model, the metasalience controller is the subagent responsible for determining whether the learner attends to background color.

Assuming that the metasalience controller has just these two options—*attend to* or *ignore* background color—amounts to supposing the learner has two available ways of categorizing the repeated problems she faces: she can either treat all stimulus displays as instances of the same kind of problem, or she can distinguish between two kinds of problem corresponding to the two possible background colors. Of course, we could envision more complex environments featuring "distractor" cues the learner must learn to ignore in order to consistently succeed in the problem at hand.[8] Such distractions could be straightforwardly implemented in the model by making more actions available to the metasalience controller, where each action corresponds to a different feature of the environment the agent could attend to in distinguishing problems.[9] For the sake of simplicity, they are excluded here.

Recall also that in the experiment the stimulus items presented in each trial differ with respect to two features, radius and luminosity. The *salience controller* determines which of these features the learner will condition on in choosing an order in which to select the stimulus items.[10] If the metasalience controller chooses to attend to background color, then the salience controller conditions this choice on whether the background is red or blue; otherwise, background color is ignored in choosing to attend to luminosity or radius.

[8] A concrete example: Perhaps there is a lamp in the corner of the laboratory within the learner's view. Its state (*on* or *off*) is uncorrelated with the task the learner will be rewarded for performing. Nevertheless, the agent might notice and condition its action on the state of the lamp, effectively categorizing the problems they face as "lamp-on problems" and "lamp-off problems." Consistent success in the problem at hand would require the agent to learn to ignore the lamp and instead attend to background color.

[9] In terms of the balls-and-urns interpretation given below, each distractor feature would correspond to a distinct type of ball in the metasalience urn. For each such feature, the *salience controller* (whose behavior is described in detail below) would be equipped with a distinct urn for each possible state of that feature. Upon drawing a ball from the metasalience urn, the metasalience controller would tell the salience controller to draw from the urn corresponding to the observed state of the feature specified by the type of ball drawn. Then, the dynamics would work exactly as described below for the model without distractor cues.

[10] The model of attentional learning in the present paper bears much in common with extant work on the evolution of salience in self-assembling games. For more on salience learning, see Herrmann and VanDrunen (2023), Barrett (2020), and Torsell and Barrett (2024). For an introduction to the framework of self-assembling games, see Barrett and Skyrms (2017) and Barrett (2025).

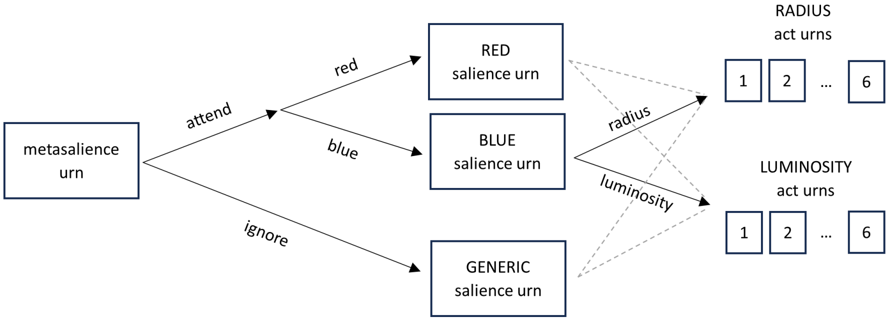

Figure 1. The urn model

The *action controller* is responsible for determining the order in which items are selected. If the salience controller chooses luminosity as the salient feature, the action controller distinguishes among objects according to their rank with respect to luminosity; if the salience controller chooses radius, the action controller distinguishes among objects according to their rank with respect to radius length. The action controller can choose the same object multiple times in a given trial; it must learn that repeated selections lead to failure in the problems at hand. In the model, each sub-agent's dispositions—corresponding to noticing or failing to notice the color cue, selecting a feature of the stimuli on which to condition one's actions, and choosing a behavioral response to those stimuli—evolve by Moran reinforcement learning.[11] The total model of choice is illustrated in figure 1.

For a more complete characterization of the model, consider a single trial in Avdagic et al.'s experiment. For concreteness, suppose it is a trial in the third training phase. First, the target task is selected by a random and unbiased choice between the luminosity task and the radius task, the background color of the screen is set accordingly, and the stimulus items are presented to the learner. (In the first two training phases, the task is chosen deterministically.) Then, the number of items is determined. In Avdagic et al.'s experiment, the number of items varies between three and six throughout the third training phase. As a simplifying assumption, our simulations fix the number of items at four for this phase.

Another respect in which our simulated experiment diverges from Avdagic et al.'s setup is in assuming that the stimulus items vary in *both* radius and luminosity throughout all training phases (recall that for Avdagic et al., items vary in only one dimension until the first experimental stage.) In a k-item trial, a random permutation f of $\{1, 2, ..., k\}$ is chosen, where $f(i)$ $(1 \leq i \leq k)$ gives the rank with respect to the distractor dimension of the ith-ranked object with respect to the task-relevant dimension. If, for example, in a three-object radius task we have it that $f(2) = 3$, this is interpreted to mean that the second-smallest stimulus item is also the brightest.

[11] The assumption the reinforcements and punishments occur for all three subagents in lockstep turns out to be inessential to the model's success. See Appendix figure 1 for results for a variant of the model in which the metasalience and salience controllers' acts are only positively reinforced on successful trials (and propensities are unchanged on unsuccessful trials), but the action controller's dispositions evolve by both reinforcement on successful trials and punishment on unsuccessful ones.

The first step in determining the learner's response to the presented stimuli is executed by the metasalience controller. The metasalience controller draws a ball from a well-mixed urn—the *metasalience urn*—containing balls labeled "attend" and balls labeled "ignore." The metasalience controller's draw determines which of three urns the salience controller will draw from to determine which feature of the stimulus items the learner will condition its final behavioral response on. If an "attend" ball is drawn, then the metasalience controller observes the background color of the screen and reports the color to the salience controller, who draws from an urn labeled "BLUE" if the background is blue and draws from an urn labeled "RED" if the background is red. Call these the *blue salience urn* and the *red salience urn*. If an "ignore" ball is drawn, then the salience controller draws from an urn labeled "GENERIC". The red, blue, and generic salience urns each contain balls labeled "radius" and balls labeled "luminosity." Note that in the first training phase, the blue salience urn is never drawn from, since the metasalience controller never observes a blue background; similarly, the red salience urn is never active in the second training phase.

The salience controller's draw determines which of two sets of urns the action controller will draw from to determine the order in which to select the stimulus items. One set of urns is labeled "RADIUS" and the other is labeled "LUMINOSITY." Call these the *radius act urns* and *luminosity act urns*. Each set contains six urns, labeled with digits from 1 to 6 subscripted with an r or an l to distinguish urns in the luminosity set from urns in the radius set (so that, e.g., the radius act urns include an urn labeled "2_r" and the luminosity act urns include an urn labeled "2_l"). Each of these act urns contains balls labeled with digits ranging from one to six.

In considering how the action controller determines the final behavioral response, there are two possible cases. The first is the case in which the salience controller draws a ball labeled with the task-relevant dimension—suppose, without loss of generality, that it is luminosity—and so the learner attends to luminosity in choosing an order in which to select the items. In that case, the action controller first draws from the 1_l urn to determine which item it will select first. If the learner draws a 1-labeled ball, then its first selection will be of the least luminous object—the unique task-appropriate choice—and it will then determine its second selection by drawing from the 2_l urn. If that selection is successful (i.e., if the learner draws a 2-ball, and so selects the item ranked second with respect to luminosity), it will make its third selection by drawing from the 3_l urn, and so on. If it correctly ranks all the presented stimulus items according to luminosity rank by drawing balls whose labels match the corresponding urn labels, then the trial is successful. The agent receives a reward, and *every urn* that was drawn from in that trial is updated by Moran reinforcement with a reinforcement magnitude of r (i.e., first a random selection of r many balls is removed from the urn, then r balls of the type just drawn are added to the urn).

In the second kind of case, the salience controller draws a ball corresponding to the task-irrelevant dimension, and so the learner's behavioral response will be determined by draws from the radius act urns (we maintain the assumption that the task-relevant dimension is luminosity). Then, since the learner attends to the distractor dimension, it will succeed on that trial just if it draws an $f(1)$-ball from the 1_r urn, an $f(2)$ ball from the 2_r urn, and so on (because f relates the rank of a given object with respect to the task-relevant dimension to its rank with respect to the distractor dimension). If the agent does successfully rank all the items, it updates all urns drawn from in that trial by Moran reinforcement just as in the case where the task-relevant dimension is salient.

Regardless of whether the agent attends to the task-relevant or distractor dimension, if it selects an object out of order, punishment proceeds as follows: the final act urn drawn from (i.e., the urn drawn from in determining the incorrect selection), the salience urn drawn from, and the metasalience urn are updated by Moran punishment (i.e., p many balls of the type just drawn are removed and replaced according to the rule described above). Notice that the contents of the act urns drawn from prior to the incorrect selection remain unchanged.

Recall that the number of stimulus items presented (k) varies in some phases of Avdagic et al.'s experiments. In our model, this is accommodated by allowing the action controller to first observe the number of stimulus items on the screen in a given trial and then remove, from all urns labeled with a number no larger than k, all balls labeled with numbers larger than k. Suppose, for example, that the learner faces a trial in which $k = 3$. Since only three stimulus items are presented, the learner will consult only the first three urns in the set of act urns corresponding to the salience controller's draw. All balls labeled with numbers greater than 3 are removed from these urns. The idea is that the learner's disposition to choose, e.g., the fifth-most luminous item in a collection of stimuli will not be activated in a trial in which it sees just three items; only the learner's propensities to select the first, second, and third brightest stimuli (represented by the number of 1-, 2-, and 3-labeled balls, respectively, in the relevant urn) should count in determining its behavior.

Of course, the model does not capture all potentially relevant features of the monkeys' learning and cognition or all the details of Avdagic et al.'s experimental protocol. It is particularly significant that the radii and luminosities of the stimuli are represented only in terms of their ordinal rank—there is no way to compare the differences between pairs of stimuli with respect to these features. Recall that the experimenters facilitated the monkeys' learning in the first experimental phase by reducing the degree to which the stimulus items varied in the distractor dimension relative to the task-relevant dimension. Since the model represents stimulus items only in terms of their ordinal ranks, it cannot capture this strategy for inducing an attentional bias in favor of the task-relevant dimension. Nor does the model enable us to represent how reduced differences in the distractor dimension relative to the task-relevant dimension would influence the monkeys' learning, as the differences between the monkeys' performance in the first and second phases of the experimental stage clearly indicate it did.

Furthermore, in all of Avdagic et al.'s training phases, it is possible for the monkeys to learn to perform accurately by conditioning their behavior on *which feature varied across stimulus objects in a given trial*, rather than background color, since stimuli varied only with respect to the task-relevant dimension throughout the training stage. For simplicity, no parameter for the salience of this feature of the subjects' environment is included. As a result, the model cannot capture how limiting variation to the task-relevant dimension might assist the monkeys' learning, and so the model assumes that objects vary with respect to both dimensions in each trial of the training stage.

These simplifying assumptions—that stimulus items are fully characterized by their ordinal rank with respect to the two dimensions and that the items presented always vary in both dimensions—have the consequence that the model cannot capture the differences between the protocols of the third training phase and those of the two experimental phases. In the model, learning to successfully navigate the final training phase *just is* learning to flexibly switch between the two SimChain tasks in the constant presence of variation with respect to both luminosity and radius, with only background color indicating which task will

training phase	***items/problem***	***mean no. sessions to 80% criterion***
1	3	12.73
	4	7.44
	5	9.96
	6	12.77
2	3	62.72
	4	7.51
	5	10.07
	6	12.99
3	4	7.6

FIGURE 2. mean time-to-criterion for each simulation phase

be rewarded in a given trial; in Avdagic et al.'s experiment, the monkeys did not face "complete" task switching problems of this kind until the experimental stage. For this reason, simulations were run only on the three simulated training phases.

5. Results and Discussion

Simulations were run in which the learner faced the three training phases in succession. As in Avdagic et al.'s experiment, the learner progressed by successively meeting the 80% performance criterion on problems with increasingly large collections of stimulus items. A simulated run was counted as a success if the learner progressed through both single-task training phases and reached the 80% criterion in the third phase.

Initially, the salience urns and the metasalience urn each contained 100 balls (50 of each possible type). Similarly, each of the action controller's urns contained 50 balls of each possible type (for a total of 300 balls in each urn). Recall the structure of the first two training phases: subjects progress through blocks of problems with an increasing number of stimuli as they successively meet the performance criterion. Because the action controller ignores balls and urns labeled with numbers greater than the number of stimulus items presented, we can think of each of the first two training phases as beginning with the action controller working with just three urns in each set (i.e., the luminosity act urns and the radius act urns), each of these containing 150 balls (initially: fifty 1-balls, fifty 2-balls, and fifty 3-balls). Each time the learner meets the performance criterion for the k-item task, a new urn is added to both the luminosity and radius sets (containing fifty balls for each label from 1 to $k + 1$), and fifty k+1-labeled balls are added to each of the k act urns that were available in the preceding trial, until the criterion is met for the maximum number of items.

Simulations were executed in batches of 1000 runs for all reinforcement and punishment values $(r, p) \in \{1, 2, 3\} \times \{0, 1, 2, 3\}$. For each batch, r and p were held constant across and within runs. On all parameter values tested except $(1, 0)$, all runs ended in success. The results reported in this section are for 1000 runs of the entire training sequence with $r = 2$ and $p = 1$. There is nothing special about these values. While the speed of learning varied across parameter settings, with higher values for both r and p corresponding to faster learning, the qualitative patterns reported below held across all parameter settings tested. For results on alternative parameter settings, see Appendix figure 2.

Figure 2 reports the mean number of 50-trial sessions it took for the agent to reach the performance criterion for each phase of the training stage. Avdagic et al. do not report analogous results for the training phases of their experiments, so no direct comparison between the performance of the model and the performance of the monkeys is possible. However, it is worth noting that the values for average sessions-to-criterion for our simulations never exceed 63—and in fact, with the exception of the first segment of phase 2 (which is discussed in greater detail below), all values are less than 13. Given that the experimental phases consisted of 36 and 48 fifty-trial sessions each, this suggests that, at least on our chosen parameter settings, the modeled agent is able to progress through the training stage within a reasonable timeframe for an experiment of this kind.

Notice that, although the number of stimulus items increases in each successive segment of the first two phases, the number of sessions-to-criterion did not uniformly increase as more items were added. This is because the structure of the model allows for significant transfer of learning from earlier segments to later ones. Once the learner has evolved up dispositions allowing it to achieve $\geq 80\%$ accuracy on, say, three-item radius problems in phase one, it has already achieved much of the learning necessary to reliably complete four-item problems. It is already disposed to choose a "radius" ball from the salience urn with high probability, to choose a "1"-ball from the 1_r urn with high probability, to choose a "2"-ball from the 2_r urn with high probability, and to choose a "3"-ball from the 3_r urn with high probability.

Moving into four-item problems, the learner does face some new challenges: the fifty "4" balls added to the three act urns used in the previous segment interfere with its evolved dispositions, and the contents of the newly-activated 4_r urn are uniform across ball types (meaning that initially the learner will make a correct fourth-item selection with chance 0.25). But the learning the agent has already accomplished is a powerful aid to meeting these challenges. In particular, the favorable dispositions built up in the three previously-active act urns, even when diluted by the new "4" balls, give the learner many more opportunities to attempt, and learn from, drawing from the 4_r urn than it initially would have were it to start with uniform proportions of types in all the relevant act urns. Were the agent to enter the four-item problems with its untrained uniform propensities, it would have the opportunity to make a fourth-item selection (i.e., draw from its fourth act urn) only on those trials in which its random guessing produced three correct choices in succession.

The agent's learning in the first two training phases thus exhibits a kind of *bootstrapping*, in which the learning accomplished in earlier, simpler problems facilitates faster learning in later, more difficult problems. The structure of these phases—that the number of items subjects are required to rank gradually increases from an initially small number—is important to the agent's success in working up to the task switching problems of the third phase insofar as it allows the agent to take advantage of this learning transfer.

A striking feature of the results in figure 2 is the large difference in time-to-criterion between the first segment of the first training phase and the first segment of the second training phase. The explanation of this feature has to do with the metasalience controller's dispositions. Figure 3 plots, for each run, the final probability with which the learner attended to background color for each of the three training phases, given by the proportion of "attend" balls in the metasalience urn at the end of each phase. Results are reported at a precision of 0.05. Notice that the distribution for phase 1 is bimodal, with peaks close to zero and close to one. On about half of runs, the learner had either evolved to reliably attend to background color or to reliably ignore it by the time it reached the performance criterion for six-item radius problems. For the other half, the learner exited the first phase with an intermediate probability of attending to background color.

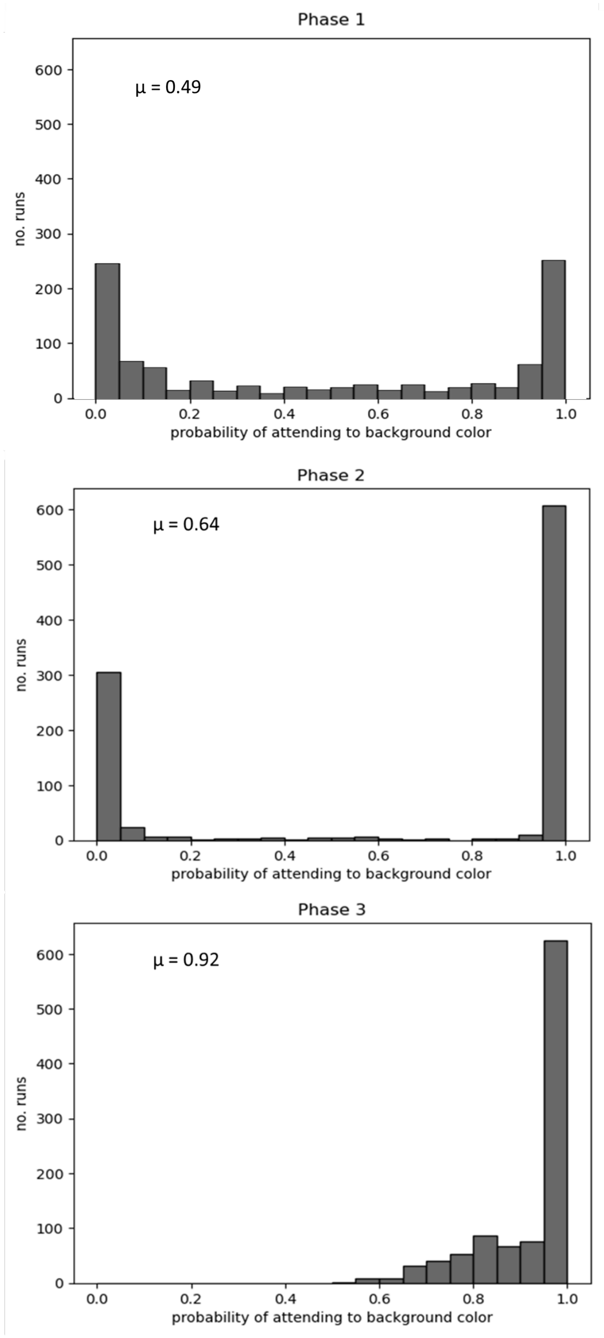

Figure 3. Final metasalience choice probabilities, phases 1-3

How reliably the learner had evolved to attend to background color by the end of the first phase of a run strongly influenced the difficulty of adjusting to the new practical demands of phase 2 on that run. When the learner evolves to reliably attend to background color, working up to the performance criterion for the three-item problems of phase 2 is no more challenging than reaching the criterion in the first segment of phase 1. But when the learner evolves to ignore background color, considerable new difficulties arise. In this case, the learner draws from the same set of salience and act urns she had trained up in the first stage, saddling her with a strong bias in favor of attending to the task-irrelevant dimension.

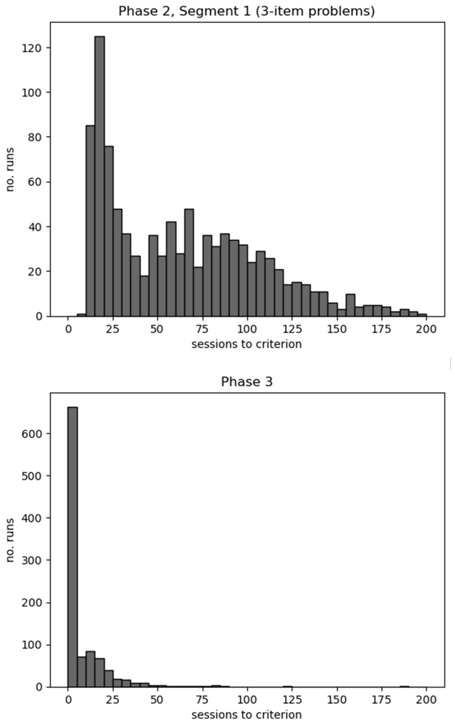

FIGURE 4. Sessions-to-criterion for phase 3 and the first segment of phase 2

She must retool to meet the demands of the new environment by either gradually unlearning the bias of the generic salience urn or else learn to attend to the background color and develop task-appropriate attentional dispositions via the blue salience urn. This is what explains the large mean sessions-to-criterion value for the first segment of phase 2.

There are a few more things to note about the metasalience probabilities recorded in figure 3. The distribution of final metasalience probabilities after phase 2 indicates a shift in favor of attending to background color, with nearly twice as many runs resulting in a probability close to one than resulting in a probability close to zero. Compared with the data for phase 1, the mean probability of attending to background color has increased by nearly 0.15. Additionally, the number of runs ending with an intermediate probability of attending to background color has decreased.

As in the transition from phase 1 to phase 2, the probability with which the learner attended to background color influenced the difficulty of learning to adjust to the new challenges of phase 3. When the learner exited phase 2 disposed to attend to background color with high probability, it entered phase 3 already disposed to condition its behavior on the feature of its environment that would reliably signal which task it would be rewarded for performing in phase 3. On runs in which the agent had learned to attend to background color with high probability in *both* the first and second phase, it entered phase 3 with a strong metasalience bias in favor of attending to background color, a strong salience bias in favor of attending to the task-relevant dimension conditional on the background color, and act-level dispositions accurate enough to successfully complete the training phases for each individual SimChain task—precisely the dispositions necessary for reliable success in phase 3. In the time-to-criterion distribution shown in the bottom panel of figure 4, these are the runs accounting for the large spike in the leftmost bin of the histogram. By contrast, when the learner exited phase 2 disposed to ignore background color with high probability, it had to *learn* to notice shifts in background color and to associate each color with the appropriate set of salience- and act-level dispositions in the third phase. The maximum time-to-criterion for phase 3 was 185 sessions, though in the overwhelming majority of runs, the phase was completed in fewer than 50 sessions.

Overall, these results indicate that the model reliably learns to cope with the inductive challenges distinctive of task switching problems with a high degree of success. It is important to emphasize again the simplicity of the model's basic structure: three simple subagents responsible for selecting what the learner pays attention to and how it behaves evolve by reinforcing successful dispositions and punishing unsuccessful ones. In a simulated environment in which the learner first has an opportunity to learn two distinct ranking tasks in isolation and then confronts a series of trials that randomly alternate between the two tasks, the operation of this simple dynamics gradually assembles a system that can deftly navigate the task switching problems with a high degree of success. The model thus illustrates one way in which a learner might accomplish what Avdagic et al.'s subjects achieved in the lab. It is significant that the model's success arises from a simple kind of reward-based association learning operating on the agent's attentional and behavioral dispositions.

6. Conclusion

Natural learners face a problem of projectibility inasmuch as they must learn which regularities in their environment are useful guides for action and prediction in order to survive. We considered a concrete laboratory setting in which subjects face a version of this problem, using a model to show how that problem might be solved by means of a simple form of reinforcement learning. In the model, the dynamics operates simultaneously on two kinds of attentional dispositions and first-order behavioral dispositions. As these dispositions coevolve, the agents learn to perform two distinct tasks and to selectively activate the dispositions suited to each task depending on the state of a task cue. On simulation, the learner reliably comes to perform both tasks with a high degree of accuracy and to attend to the task cue to distinguish between them.

The model thus shows how the problem of projectibility that arises in task-switching problems might be solved by means of a particularly simple form of learning. An important virtue of this model is that it describes a learning process in which the mechanism responsible for learning which regularities to project—salience learning by means of reinforcement—operates *simultaneously* with the mechanisms responsible for learning the first-order tasks the agent must learn to discriminate. That is, the agent learns which regularities in its environment to project for the purpose of distinguishing the tasks even as it learns the tasks themselves. What's more, all the relevant dispositions evolve by the *same* simple reinforcement mechanism.

The resulting story is one on which a highly structured learning system emerges endogenously from what is initially a largely unstructured collection of attentional and behavioral dispositions. The model might thus be interpreted as capturing the *self-assembly* of a learning method adapted to the demands of the task-switching problem at hand. In particular, it captures a process that reliably assembles a learning method that projects the practically relevant regularities in a particular kind of task-switching problem.

Of course, not all the relevant structure is learned endogenously; some is baked in. We assume, for example, that our learner starts out "knowing that" either brightness or size is what matters in the problem at hand, that she is not distracted by potential cues other than background color in individuating tasks, and that she represents stimulus items according to their order along the two relevant dimensions. The self-assembly involved in the model is not self-assembly *ex nihilo*. An interesting question for further work concerns how the structural features of the agent stipulated at the start might themselves be learned or evolved. Moreover, we studied just one form of reinforcement learning as the mechanism of self-assembly. Another useful step in the direction of greater generality would be to investigate whether other forms of reinforcement learning would be similarly successful.

While canonical philosophical treatments of projectibility center on justifying the regularities agents attend to in inductive inference, here we have been concerned with learning, not justification. The model above does not show that the monkeys are justified in projecting the regularities they come to attend to. Rather, it shows that simple, reward-based learning suffices to guide them to project regularities that are in fact reliable guides to successful action in the problems they repeatedly face, even when those problems are complex and frequently-changing. Of course, there is no *guarantee* that the patterns they look for will continue indefinitely to provide helpful practical guidance. The modeled learners are nevertheless disposed to persist in projecting them because doing so has led to successful action in the past and their psychological nature disposes them to repeat operations that have been rewarded.

References

[1] Argiento, R., Pemantle, R., Skyrms, B., & Volkov, S. (2009). Learning to signal: Analysis of a micro-level reinforcement model. Stochastic processes and their applications, 119(2), 373-390.

[2] Avdagic, E., Jensen, G., Altschul, D. et al. (2014). Rapid cognitive flexibility of rhesus macaques performing psychophysical task switching. *Animal Cognition* 17, 619–631. https://doi.org/10.1007/s10071-013-0693-0

[3] Barrett, J. A. (2006). Numerical simulations of the Lewis signaling game: Learning strategies, pooling equilibria, and the evolution of grammar. Institute for Mathematical Behavioral Sciences. Paper 54.

[4] Barrett, J. A. (2024). Humean learning (how to learn). *Philosophical Studies*, 181(1), 281-297.

[5] Barrett, J.A. (2025). *Self-assembling games: language, learning, and inquiry.* forthcoming, Oxford University Press.

[6] Barrett, J.A. (2020). Self-assembling games and the evolution of salience. *British Journal for the Philosophy of Science.* https://www.journals.uchicago.edu/doi/10.1086/714789

[7] Barrett, J. A. and Brian Skyrms (2017). Self-assembling games. *The British Journal for the Philosophy of Science*, 68(2), 329—353

[8] Barrett, J., & Zollman, K. J. (2009). The role of forgetting in the evolution and learning of language. *Journal of experimental & theoretical artificial intelligence*, 21(4), 293-309.

[9] Beggs, A. W. (2005). On the convergence of reinforcement learning. *Journal of economic theory*, 122(1), 1-36.

[10] 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. *Journal of mathematical psychology*, 42(2-3), 266-286.

[11] Erev, I. and A. E. Roth (1998). Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* 88: 848—81.

[12] Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review*, 112(4), 912.

[13] Goodman, N. (1983). *Fact, fiction, and forecast.* Harvard University Press.

[14] Herrmann, D., & VanDrunen, J. (2022) Sifting the signal from the noise. *British Journal for the Philosophy of Science.*

[15] Herrnstein, R. J. (1970). On the law of effect. *Journal of the experimental analysis of behavior*, 13(2), 243-266.

[16] Huttegger, S. M. (2017). *The probabilistic foundations of rational learning.* Cambridge University Press.

[17] Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological bulletin*, 136(5), 849.

[18] Lewis, D. (1969). *Convention: A philosophical study.* John Wiley & Sons.

[19] Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, 7(3), 134-140.

[20] Moran, P. A. P. (1958). Random processes in genetics. *Mathematical proceedings of the Cambridge Philosophical Society*, 54 (1): 60–71.

[21] Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American economic review*, 85, 1313–1326.

[22] Rosenthal, R. W. (1993). Rules of thumb in games. *Journal of Economic Behavior & Organization*, 22(1), 1-13.

[23] Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, 8(1), 164-212.

[24] Skyrms, B. (2012). *From Zeno to arbitrage: essays on quantity, coherence, and induction.* OUP Oxford.

[25] Suppes, P. (1994). “Learning and projectibility." in *Grue!*, D. Stalker, ed. Open Court Publishing.

[26] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.

[27] Torsell, C., & Barrett, J. A. (2024). Learning how to learn by self-tuning reinforcement. *Synthese*, 203(6), 209.

[28] Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* Macmillan Press. https://doi.org/10.5962/bhl.title.55072