George, David N., Dwyer, Dominic Michael , Haselgrove, Mark and Le Pelley, Mike 2024. Apparent statistical inference in crows may reflect simple reinforcement learning. Quarterly Journal of Experimental Psychology 10.1177/17470218241305622

Publishers page: https://doi.org/10.1177/17470218241305622

George, D. N., Dwyer, D. M., Haselgrove, M., & Le Pelley, M. (2024). Apparent statistical inference in crows may reflect simple reinforcement learning. Quarterly Journal of Experimental Psychology. © Experimental Psychology Society 2024.
https://doi.org/10.1177/17470218241305622

Apparent statistical inference in crows may reflect simple reinforcement learning

David N. George[1], Dominic M. Dwyer[2], Mark Haselgrove[3], Mike E. Le Pelley[4]

[1]School of Psychology, University of Hull, Hull, UK

[2]School of Psychology, Cardiff University, Cardiff, UK

[3]School of Psychology, University of Nottingham, Nottingham, UK

[4]School of Psychology, UNSW Sydney, Sydney, Australia

Author Note

David N. George https://orcid.org/0000-0002-0038-2999

Dominic M. Dwyer https://orcid.org/0000-0001-8069-5508

Mark Haselgrove https://orcid.org/0000-0001-8981-1181

Mike E. Le Pelley https://orcid.org/0000-0002-5145-5502

All code to reproduce our simulations and model fitting is available from the Open Science Framework and can be accessed at https://osf.io/n2cf5/.

Correspondence concerning this article should be address to: David N. George, School of Psychology, University of Hull, Hull HU67RX. Email: d.george@hull.ac.uk

Abstract

Johnston et al. (2023) report results which they argue demonstrate that crows engage in statistical inference during decision-making. They trained two crows to associate a set of stimuli with different reward probabilities (from 10% to 90%) before choice tests between pairs of stimuli. Across most pairwise combinations, and in a control task in which the number of rewards was equated between probabilities, both crows preferred the stimulus associated with higher reward probability. The magnitude of this preference was affected by the absolute difference between the two probabilities, although (contrary to a claim made by Johnston et al.) preference did not reflect the ratio of prior probabilities independently of absolute differences. Johnston et al argue that preference for the stimulus with the higher reward probability is “the signature of true statistical inference” (p.3238), implemented by an analogue magnitude system that represents the reward probability associated with each stimulus. Here, we show that a simple reinforcement learning model, with no explicit representation of reward probabilities, reproduces the critical features of crows’ performance – and indeed better accounts for the observed empirical findings than the concept of statistical inference based on analogue magnitude representations, because it correctly predicts the absence of a ratio effect that would reflect magnitudes when absolute distance is controlled. Contrary to Johnston et al.’s claims, these patterns of behaviour do not necessitate retrieval of calculated reward probabilities from long-term memory and dynamic application of this information across contexts, or (more specifically) require the involvement of an analogue magnitude system in representing abstract probabilities.

Keywords: reinforcement learning, associative learning, statistical inference, distance effect, magnitude effect, analogue magnitude system

Introduction

Given a choice between two stimuli that differ in their history of reinforcement, animals will preferentially respond to the stimulus that is associated with the higher rate of reinforcement. For example, Roberts et al. (2018) gave pigeons 10 days of training in which a red or a green key light was presented on different trials; pecking 10 times upon the light could yield a food reinforcer. For one key, reinforcement was delivered after 10 pecks with a probability of 75%, and for the other key reinforcement was delivered with a probability of 25%. Probe trials were also presented during each session in which both stimuli were presented side by side, providing the pigeons with a choice. It was found that, at first, pigeons' preference for the 75% key over the 25% key was around 70% but, after 10 sessions, this preference had increased to around 95% (Experiment 1). This result was not merely the consequence of one key being paired more often with reinforcement, for when the number of pairings of each key with reinforcement was equated, a similar (but slightly smaller) preference was still apparent (Roberts et al, Experiment 2).

The explanation provided by Roberts et al. (2018) for these results is that the pigeons computed a probability of reinforcement based upon an approximate number system. Specifically, they proposed that the numbers of reinforcers and non-reinforcers obtained on each key were represented as individual magnitudes, and that ratios were computed based upon these magnitudes for each key. The key with the larger ratio of reinforcement magnitude to non-reinforcement magnitude would then be preferred. The functional argument here is that pigeons, and other animals, need to *compute probabilities* for successful foraging – to exploit environments with higher densities of food, or potential mates, and lower densities of predators. In a recent article published in *Current Biology*, Johnston et al. (2023) agreed with the idea that the ability to *relate* different reward probabilities is a necessary feature of statistical inference, and expanded on the definition by arguing that a demonstration of true statistical inference further requires that reward probabilities are used flexibly across different contexts. The argument here is that a site which affords a 60%

success rate for foraging is optimal only when the alternative is a site with a lower probability of successful foraging (e.g. 20%). Compared to a site that affords, for example, a 90% success rate for foraging, choosing the 60% site is sub-optimal. Johnston et al. investigated this issue in two experiments designed to determine whether birds can learn to associate multiple reward probabilities with arbitrary stimuli. The findings of these experiments were taken as evidence that, “remarkably, crows demonstrated statistical inferences immediately after learning the sign–reward associations” (p3240).

Of course, drawing the conclusion that crows were demonstrating statistical inference in these experiments requires one to be able to define what statistical inference is: i.e., what would constitute an example, and what would not. As outlined above, Johnston et al. (2023) equate statistical inference with the ability to use limited information to draw conclusions about the likelihood of an event. In particular, they define the type of statistical inference that their crows—and Roberts et al.’s (2018) pigeons—were employing as a “sample-to-population" inference. Here, a conclusion is drawn about the population using sample information which must be stored and updated each time sampling takes place. This class of statistical inference is argued to be more cognitively demanding than “population-to-sample” statistical inference, in which a conclusion is drawn about a sample based upon population information that is available to the subject at the time of, or just prior to, decision making (e.g., Denison & Xu, 2014).

Notably, Johnston et al. (2023) go further by arguing that statistical inference is “inextricably tied to quantity competence” (p3241), wherein abstract quantities are compared via an analog magnitude system (AMS). Here the suggestion is that abstract reward probabilities are compared as though they were analog magnitudes (like weight, or height, or time), and hence that resulting inferences will show similar patterns to those seen for analog magnitudes: specifically, that discriminations between quantities are determined by the *ratio*, rather than the absolute difference, between them – a pattern that Johnston et al. describe as “a hallmark of the magnitude effect” (p3239). Finally, Johnston et al. draw a distinction between true statistical inference and “low-level associative learning strategies", arguing that associative learning cannot explain the high accuracy observed from the start of the test sessions in their forced choice paradigm.It is our contention, however, that there is really nothing remarkable about the behaviour of the crows in Johnston et al.'s experiments. Instead, the behaviour demonstrated there, and also in Roberts et al.'s prior study, can in fact be explained by simple principles of reinforcement learning – exactly the kind of low-level associative learning process that Johnston et al. argued they had ruled out. Here we (1) re-analyze Johnston et al.'s results to show that there is in fact no magnitude effect (independent of distance), and (2) describe the output of computational modelling of the experiments reported by Johnston et al. which accurately fits the reported performance of the crows (including the absence of the magnitude effect independent of distance). In doing so, we show that the conclusions drawn by Johnston et al. for what they argue their crows are (1) doing and (2) not doing, are both undermined. We begin by outlining the task used by Johnston et al.

Overview of Johnston et al. (2023)

Experiment 1: Methods

Training 1. Two carrion crows initially received daily sessions of training in which responses to nine different-coloured and shaped cues, presented upon a touch-screen monitor, were probabilistically (10% - 90%) followed with reinforcement. On each trial, a cue was presented in the centre of the monitor, and a peck on the monitor during a subsequent "response period" (of up to 8 seconds) was followed with bird-seed reinforcement according to the programmed probability of reinforcement for that trial. If a response was made prior to the response period then the trial was terminated and reinforcement was not delivered; the trial was also terminated without reinforcement if no response was made during the response period. Each crow received either 10 or 11 days of this training.

Training 2. During the second training phase, two-alternative choice responding was trained with two new cues over the course of 3 or 4 days.

Test Sessions. Following the two training phases, the crows were given 10 sessions of choice trials, in which the timing and structure of the trials was the same as in *Training 2*. Nine consecutive test sessions were given immediately following training, and one test session was given a month later. During testing, there were 72 unique choice trials which corresponded to every possible pairwise combination of the nine cues from *Training 1* in two spatial arrangements. Responding on either cue was reinforced with its assigned probability.

Experiment 1: Results

The results of the test sessions are shown in Figure 1. These figures were produced from the test data provided by Johnston et al. at the Dryad open science repository (https://datadryad.org/stash/share/iAogVMtOcm53SjMNZbNVsMVR4UWmKTutydJWiSPvz-w). Figure 1A shows, for each crow individually, the percentage of choices that were optimal (i.e. to the stimulus associated with the higher probability) as a function of the difference in probability between the two test stimuli. As can be seen, and as Johnston et al. noted, the percentage of optimal responses increases as a function of this difference. However, as Johnston et al. also point out, choice performance between two stimuli of equal difference is not always the same. The top panels of Figure 1B show optimal-choice percentages for both crows as a function of the lower- and higher-reward probabilities of the two choice stimuli. To illustrate their point, Johnston et al. highlighted an example in the data from crow 2: here we see that the optimal choice percentage between stimuli associated with 10% and 20% reinforcement was around 20%; however, optimal choice increased to almost 90% when the choice was between stimuli associated with 50% and 60% reinforcement, even though the absolute difference between options was the same (10%) in each case. Johnston et al. consider this effect to be the “hallmark of the magnitude effect” (p. 3239), which states that discrimination between quantities is determined by their ratio, rather than the difference between them (akin to the Weber-Fechner Law: Fechner, 1966). That is, the smaller the ratio between two choice quantities (where this ratio is calculated as lower probability divided by higher probability), the easier the discrimination should be. Worryingly, we run into a problem at this point, since the example noted above (and highlighted by Johnston et al.) actually shows the *opposite* pattern to this so-called hallmark of a magnitude effect: in this example, the crow's performance is *poorer* when the ratio is smaller (10%/20% = 0.5) than when it is larger (50%/60% = 0.83). Putting this particular problematic example aside for the moment, Figure 1C plots, for each crow, the percentage of optimal choice as a function of the ratio—again, calculated as lower probability divided by higher probability. This figure shows a systematic decrease in discrimination performance as ratio increases, and these are the key data that Johnston et al. interpret as revealing that "both crows made more optimal choices when the ratio between reward probabilities was smaller" (p. 3240).

<Figure 1 about here>

Experiment 2: Methods

Training**.** The same two crows from Experiment 1 subsequently took part in Experiment 2. Two novel cues were presented on separate trials with responding to one cue reinforced on 80% of trials and responding to the other cue reinforced on 40% of trials. In order to match the total number of reinforcements that were delivered following these two cues, there were twice as many 40% trials as 80% trials.

Testing. Following the training phase, the crows were given 1 session of choice trials between the 80% and 40% cues. The timing and structure of these test trials were similar to the testing phase of Experiment 1.

Experiment 2: Results

The results of the test session from Experiment 2 are shown in Figure 1D. Again, this figure was produced from data provided by Johnston et al. (2003). This figure shows the percentage of optimal choices, for each crow. During this test, both birds chose the stimulus that was associated with the 80% reinforcement probability significantly more than chance.

Conclusions of Johnston et al. (2023)

In discussing these results, Johnston et al. draw a number of conclusions. Following 10 sessions of training comprising more than 5,200 conditioning trials, they first conclude that “crows successfully learned to associate nine unique stimuli with reward probabilities ranging from 10% to 90%” (p. 3240), and second that the crows demonstrated “statistical inferences” both immediately after learning these associations, and after one month without any further exposure to the task. They contend that “statistical inference is at work if subjects use the relative reward frequency during decision making” (p. 3240) and point to the results of Experiment 2, in particular, as showing that crows use relative rather than absolute frequency of reinforcement as the signature of this statistical inference. Based on the similarity of the performance to the 80% and 40% stimuli across Experiment 1 and 2, they infer that the crows used the same “relative reward reasoning for both experiments”. Like Roberts et al. (2018), in interpreting their findings Johnston et al. refer to the notion of an analogue magnitude system (AMS), wherein values (in this case, reward probabilities) are represented on a mental scale such that errors in estimation increase in proportion to the size of the value being estimated. Critically, they argue that evidence for the operation of an AMS “presents itself via the distance and magnitude effects, whereby discrimination improves with an increasing distance between two values and, at a given numerical distance, worsens with an increasing difference in ratio difference, respectively” (p.3241). Finally, they conclude that their results provide the first evidence of the AMS in a “complex retrospective statistical inference task in which the probabilistic information is context dependent” and that this AMS is important for representing “highly abstract probabilistic value” (p. 3242).

A re-evaluation of Johnston et al. (2023)

The claims made by Johnston et al. are notable and eye-catching: if it were indeed the case that crows are capable of abstract and context-sensitive statistical inference based on values represented in an analogue magnitude system, this would be a remarkable finding that underscores the complexity of animal cognition. However, we believe that Johnston et al.’s findings do not, in fact, provide diagnostic evidence of abstract statistical inference via an AMS. Our refutation of their conclusions is based on two key points: (1) Their empirical results do not provide evidence that discrimination performance decreases as a function of ratio when absolute difference is held constant; and (2) Simulations show that a simple model of reinforcement learning can provide a full explanation of their findings without requiring notions of abstract statistical inference. We expand on these points below.

The absence of a ‘true’ magnitude effect

As noted earlier, a central piece of evidence that Johnston et al. pointed to as a “hallmark” implicating the operation of an AMS is the finding that crows’ performance decreased as a function of the ratio of probability values when the absolute distance between these values was held constant. However, the data that they took as supporting this claim (Figure 1C) are critically flawed because this analysis confounds the ratio of reward probabilities with the absolute difference in probabilities (i.e., as plotted, datapoints with larger ratio values typically have smaller differences). When instead we re-plot the data in a manner that explicitly deconfounds the ratio between reward probabilities and the difference between reward probabilities, it becomes immediately clear that there was no ‘true’ magnitude effect in Experiment 1, independent of absolute difference (see Figure 2). For example, across the pairs of stimuli that had a difference in reward probability of 10% (top left panel), there was either little consistent relationship between performance and ratio of probabilities (crow #1) or performance improved as the ratio between reward probabilities increased (crow #2) - which is the opposite to the predictions of a magnitude effect as based on the operation of an AMS. In none of the panels representing a given difference in reward probability does performance decrease as the ratio between the higher and lower reward probabilities increases. When examined in this manner it is clear that the diagnostic claim made by Johnston et al.—that for a given difference in reward probability, performance decreases as the ratio between the higher and lower reward probabilities increases—is false.

<Figure 2 about here>

Reinforcement learning explains apparent statistical inference

In the previous section we noted that Johnston et al.'s empirical data do not demonstrate a true magnitude effect, which had constituted a key pillar of their claimed evidence for the operation of statistical inference via an AMS. In the current section we show that patterns of behaviour that *were* exhibited by their crows can be explained in terms of simple principles of reinforcement learning embodied in well-established theories of conditioning, and hence do not provide diagnostic evidence of abstract statistical inference.

Theories of conditioning assume that when a stimulus is paired with a reward an association develops between mental representations of the two events. As a result, subsequent presentation of the stimulus will lead to an expectation of the reward. A unifying principle across many influential models of conditioning, going back over seventy years, is that learning of stimulus–reward associations is *error-correcting* (Bush & Mosteller, 1951; Mackintosh, 1975; Rescorla & Wagner, 1972; Rumelhart et al., 1986; Sutton & Barto, 1981; Wagner, 1981) – a suggestion that receives support at the neural level in humans and animals (Fletcher et al., 2001; Tobler et al., 2006; Waelti et al., 2001). This concept is perhaps most clearly expressed in the model of Rescorla and Wagner (Rescorla & Wagner, 1972), who proposed a rule to describe how the strength of the stimulus-reward association is modified by experience of prediction error (δ) – the difference between the reward that is experienced (r), and the reward that is expected (predicted) based on the current associative strength of the stimulus (v):

$$\delta = r - v \tag{1}$$

In Rescorla and Wagner's model, following a conditioning trial, the strength of the stimulus-reward association is updated using the learning rule

$$v \leftarrow v + \varepsilon\delta \tag{2}$$

where ε is a learning rate parameter in the range $0 \leq \varepsilon \leq 1$. Hence, when a stimulus is followed by an unexpected reward (i.e., $r > v$), δ will be positive and the association will be strengthened. When the reward is expected (i.e., $r \approx v$), δ will be small and there will be little change in the strength of the association. If an expected reward is omitted (i.e., $r < v$), δ will be negative and the association will weaken. The learning rate ε determines the impact of prediction error on changes in associative strength: if ε is large then associative strength will change rapidly in response to experience of prediction error, whereas if ε is small then associative strength will change more gradually.

In Johnston et al.'s experiments, after training with individual stimuli, crows were presented with a choice between pairs of stimuli. Hence to model performance in this task we require a *choice rule* that specifies how to translate associative strength (v) into an index of behavioural responding. Many different choice rules have been proposed (Sutton & Barto, 2018). Among these, one of the most popular and influential (across behavioural psychology, cognitive science, and economics) is the softmax function, which implements Luce's choice axiom (Luce, 1959). Here, when presented with a choice between n stimuli, the probability of choosing stimulus i, P_i, is given by

$$P_i = \frac{e^{\beta v_i}}{\sum_{j=1}^{n} e^{\beta v_j}} \tag{3}$$

where β is a bias parameter which affects the degree of variability in choices. When β is large, the stimulus with the highest associative strength will be selected with a probability approaching 1; by contrast, when $\beta = 0$ all options are equally likely to be chosen regardless of associative strength. In reinforcement learning, this variation in behaviour is described as a balance between *exploiting* current knowledge about stimulus values, and *exploring* the problem space to determine whether current knowledge provides a good fit to the actual values of rewards associated with the stimuli, with β setting the balance point (high values promote exploitation; low values promote exploration). Applying the softmax rule to the case of two stimuli presented side-by-side (as in Johnston et al.'s experiments), the probabilities of selecting the left ($\mathrm{P_L}$) or right ($\mathrm{P_R}$) stimulus are given by Equations 4 and 5 where $\mathrm{P_L} + \mathrm{P_R} = 1$.

$$P_L = \frac{e^{\beta v_L}}{e^{\beta v_L} + e^{\beta v_R}} \tag{4}$$

$$P_R = \frac{e^{\beta v_R}}{e^{\beta v_L} + e^{\beta v_R}} \tag{5}$$

Equation 4 may be rewritten as

$$P_L = \frac{1}{1 + e^{-\beta(v_L - v_R)}} \tag{6}$$

from which it is evident that choice performance is determined by the difference in the associative strengths of the two alternatives, and not by their ratio. That is, in direct contrast to the predictions Johnston et al. make on the basis of statistical inference via an analogue magnitude system, the reinforcement learning model must predict that there will be no systematic relationship between performance and the ratio between reward probabilities when the difference between reward probabilities is held constant. And as noted in the previous section, Johnston et al.'s empirical data show little evidence of any such relationship, despite their claims to the contrary.

Simulation of Johnston et al. (2023) – Experiment 1

We ran simulations of Johnston et al.'s experiments using this simple reinforcement learning model. Simulations were conducted in R version 4.3.3 (R Core Team, 2024) using the packages Metrics (Hammer & Frasco, 2018), readxl (Wickham & Bryan, 2019), and tidyverse (Wickham et al., 2019). Model parameters were fit to real data using the optim and nls functions, and goodness of fit was assessed through linear regression using the lm function. Code to reproduce our simulations and model fitting is available at https://osf.io/n2cf5/.

The performance of the two crows in Johnston et al.'s Experiment 1 was simulated by training the model using the sequences of rewarded and non-rewarded stimulus presentations that the crows experienced (obtained from the data file made publicly available by the authors). Since all rewards in this experiment had equal magnitude (a birdseed pellet), we employed a binary code so that $\mathrm{r} = 1$ when choice yielded a reward and $\mathrm{r} = 0$ when it did not. For simplicity, each stimulus was represented separately, with no overlapping features; i.e., we assumed no generalization of associative strength between stimuli. For each simulation run, nine stimuli were initialised with $\mathrm{v} = 0$. On each simulated trial of the 'Training 1' phase, the appropriate stimulus was selected, and its associative strength updated according to Equations 1 and 2. Simulated crow #1 was trained across a total of 5,229 trials. The experienced number of trials and associated probability of reward for the nine stimuli were: 578 (11%), 588 (18%), 584 (32%), 570 (38%), 578 (53%), 587 (62%), 576 (70%), 582 (82%), and 586 (91%). Simulated crow #2 was trained for 5,202 trials. The experienced number of trials and reward probabilities for individual stimuli were: 576 (9%), 581 (19%), 573 (31%), 580 (44%), 573 (56%), 580 (61%), 580 (70%), 580 (82%), 579 (91%). The Pearson correlation coefficient between programmed and experienced reward probabilities was .998 for crow #1 and .997 for crow #2. Following training, choice probabilities were calculated for each pairwise combination of stimuli using Equation 6.

We found that the best fits between the model and real data were obtained with the following parameters: $\varepsilon = .003$, $\beta = 6.0$ for crow #1, and $\varepsilon = .005$, $\beta = 6.0$ for crow #2. The results of simulations with these values are shown alongside data from the real crows in Figures 1A-C. The model explained 72% of variance in the performance of crow #1 and 80% of variance in the performance of crow #2. Notably, these simulations reproduced the critical patterns in crows' behaviour that Johnston et al. (2023) took as evidence of abstract and flexible statistical inference: proportion of optimal choices increased with the absolute size of the difference in reward probability of the two options (Figure 1A and 1B), and decreased as the ratio of reward probabilities increased (in an analysis that confounded ratio with absolute difference: Figure 1C).

To assess the typical performance of the reinforcement learning model, we ran additional simulations of a cohort of 100 crows using random trial order and stochastic reward delivery for each simulation. Each stimulus was presented on 581 trials (the average number of times each stimulus was presented to each crow during the training phase of Experiment 1), with programmed probabilities of reward of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. Delivery of reward on each trial was determined by a generating a random number from a uniform distribution. For each crow, the learning rate was randomly chosen in the range $.001 \leq \varepsilon \leq .01$. To investigate the influence of the softmax bias parameter on the predicted pattern of behaviour, for each simulated crow, choice probabilities were calculated separately using each of four bias values: $\beta = 0$, $\beta = 2$, $\beta = 6$, and $\beta = 12$. For each bias value, performance was averaged across the 100 simulation runs with different ε values. The results of these simulations are shown in Figures 3A and 3B. At very high values of the softmax bias parameter the model's performance becomes essentially deterministic and performance approaches ceiling; when bias is zero, the model effectively guesses on all trials and choice is at chance. However, for a wide range of values of the softmax bias between these extremes (and across a wide range of simulated learning rates) the model correctly predicts the key pattern of better discrimination performance as a function of absolute difference.

Finally, we tested the predictions of the model for these different bias values in an ideal system where the associative strength of the stimuli at the end of training exactly matched the programmed reward probabilities (i.e., 10% to 90% in increments of 10%). These predictions were in line with the results of our other simulations and are represented by the dots in Figures 3A and 3B.

Simulation of Johnston et al (2023) – Experiment 2

In Johnston et al.'s Experiment 2, two new stimuli were paired with reward with a probability of 40% or 80%. In contrast to Experiment 1, each stimulus was paired with reward approximately the same number of times, but the number of non-rewarded presentations was varied. As a result, the total number of times the two stimuli were presented differed. For our initial simulations of Experiment 2, we again used the trial sequences experienced by the real crows. For each simulation, two stimuli were initialised with $\mathrm{v} = 0$. We used the same parameter values as for Experiment 1: $\varepsilon = .003$ for crow #1, and $\varepsilon = .003$ for crow #2. For both crows, we set $\beta = 6, \text{and } \mathrm{r} = 1$ for rewarded trials and $\mathrm{r} = 0$ for non-rewarded trials. There was a total of 490 trials for simulated crow #1, comprising 335 trials with one stimulus (137 rewarded, giving reward probability of 41%) and 155 with the second stimulus (130 rewarded, reward probability 84%). Simulated crow #2 was trained over a total of 440 trials, 293 with one stimulus (120 rewarded, reward probability 41%), and 147 with the other (120 rewarded, reward probability 82%). At the end of training, choice preference between the two stimuli was calculated for each crow using Equation 6; resulting data are shown in Figure 1D. Both simulated crows showed a preference for the stimulus associated with a higher relative reward frequency, mirroring Johnston et al.'s empirical findings. This preference was larger for simulated crow #2 than for simulated crow #1, a difference that was also observed for the real birds.

As for Experiment 1, we ran an additional set of 100 simulations of Experiment 2 to determine how robust the model's predictions were to variations in parameter values. For these simulations, we first calculated the average number of rewarded trials with each stimulus experienced by the two real crows (127). For each simulated crow, the number of trials in which each stimulus was presented was obtained by dividing this number by the programmed probability of reward. Hence, there were 318 trials with one stimulus (reward probability 40%) and 159 with the other (reward probability 80%). Other details of these simulations were the same as those for Experiment 1: learning rate was randomly chosen in the range $.001 \leq \varepsilon \leq .01$, and each simulation was tested with bias values of $\beta = 0$, $\beta = 2$, $\beta = 6$, and $\beta = 12$. The results of these simulations are shown in Figure 3C. The model predicts a tendency to choose the stimulus associated with a higher relative reward probability over a wide range of values for both learning rate and bias.

<Figure 3 about here>

Discussion

In their recent article, published in a prominent journal, Johnston et al. (2023) made the eye-catching claim that they had demonstrated that “Crows flexibly apply statistical inferences based on previous experience”. Here, we have shown that a simple model of reinforcement learning can provide a full account of their key findings. Indeed, given the absence of a relationship between performance and the ratio of reward probabilities *independent* of the difference in reward probabilities, the reinforcement learning model, in fact, corresponds more closely to their results than does the idea of statistical inference based on the use of an analogue magnitude system as described by Johnston et al. So where does this leave the idea of statistical inference in crows?

One recourse would be to argue that models of reinforcement learning *are* models of statistical inference. In a sense this is not an unreasonable suggestion: for example, the Rescorla-Wagner model (on which the current model is closely based) can be seen as effectively implementing linear regression, establishing which stimuli are the most accurate predictors of outcome events and which are less useful. And patterns of behaviour that derive from reinforcement learning will vary across different contexts, as suggested under Johnston et al.’s definition of ‘true’ statistical inference: if associative strengths of three stimuli increase in the order A < B < C, then given a choice between B and A the model will choose B, but if given a choice between B and C the model will avoid B (and choose C instead).

On this view—that reinforcement learning is a process of statistical inference—the fact that Johnston et al.’s findings concord closely with a learning-based account supports the claim that crows can perform statistical inference. But research into animal learning has been going on for well over a century: we already know very well that behaviour driven by reinforcement learning can be observed across a wide range of species: sea slugs, snails, bees, birds, rats, cats and so on and on. And so, this demonstration that crows are capable of learning stimulus–reward contingencies does not advance our understanding of animal cognition in any particularly meaningful way: in effect, if reinforcement learning is defined as statistical inference, then we already knew that animals were capable of statistical inference long before Johnston et al.'s study of crows.

We should stress, however, that this view of "reinforcement learning as statistical inference" deviates profoundly from the claims made by Johnston et al., who argued that their findings implicated a much more complex form of flexible and abstract statistical inference wherein the reward probability of each stimulus is represented in an analogue magnitude system, with values retrieved and compared when a choice between stimuli must be made. It does not seem unreasonable to suggest that this was the claim that made their study newsworthy (e.g., Bastos, 2023; Quaglia, 2023; Wascher, 2023), and set it apart from previous work. However, as we have shown here, it is also a claim that does not receive diagnostic support from their data. By contrast, our reinforcement learning model does not encode or represent the probability of reward predicted by each stimulus at all. Instead, in this model the strength of the stimulus–reward association reflects the current expectation of reward given the presence of a stimulus. The model is 'amnesic' in the sense that it has no memory of how it arrived at the current reward expectation, or the distribution of reward experiences that led it there.

We can look at this distinction in a different way. Johnston et al. argued that their findings provided evidence that crows were performing a process of "sample-to-population" inference (also known as inductive reasoning): that is, taking information about samples (trial-by-trial experiences of each stimulus in the training phase) and integrating this information to draw a conclusion about the overall probability of reward associated with each stimulus (see: Introduction). On their account, it is this population-level statistic that is used to guide subsequent choices, with new instances seen as being drawn from that population-level distribution. However, it is important to note that the crows were never actually queried about these population-level statistics, and hence there is no direct evidence that they are making inferences about the population – or even that they are representing the population at all. Instead, the reinforcement learning account developed here sees what is being measured in Johnston et al. as sample-to-*sample* inference: crows have a series of experiences with stimuli in the training phase, and then encounter new examples of those same stimuli in the test phase, and these new instances are able to elicit conditioned responding on the basis of their similarity to training items. But at no point does the reinforcement learning model use the training examples to draw (and encode) a conclusion about the distributional nature of the theoretical underlying population.

As Lind (2018) notes "It is an intriguing paradox that associative [reinforcement] learning is acknowledged for producing complex flexible behaviour within AI research, but is often dismissed and neglected as a model for flexible behaviour in biological systems (both humans and non-human animals)." Here we address this neglect as a step towards resolving Lind's paradox: we show that a reinforcement learning model can produce complex, flexible behaviour; in the current case producing the appearance of statistical inference in crows.

References

Bastos, A.P.M. (2023). Crows make optimal choices based on relative probabilities. *Learning & Behavior*. https://doi.org/10.3758/s13420-023-00612-1

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58*(5), 313-323. https://doi.org/10.1037/h0054388

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition, 130(3)*, 335-347. https://doi.org/10.1016/j.cognition.2013.12.001

Fechner, G. (1966). *Elements of psychophysics. Vol. 1.* Holt, Rinehart and Winston: New York.

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., Papadakis, N., & Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience, 4*(10), 1043-1048. https://doi.org/10.1038/nn733

Hammer, B., & Frasco, M. (2018). *Metrics: Evaluation metrics for machine learning.* https://CRAN.R-projects.org/package=Metrics

Johnston, M., Brecht, K. F., & Nieder, A. (2023). Crows flexibly apply statistical inferences based on previous experience. *Current Biology, 33*(15), 3238-3243. https://doi.org/10.1016/j.cub.2023.06.023

Lind, J. (2018). What can associative learning do for planning? *Royal Society Open Science, 5*, 180778. https://doi.org/10.1098/rsos.180778

Luce, R. D. (1959). *Individual choice behavior*. John Wiley.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*(4), 276-298. https://doi.org/10.1037/h0076778

Quaglia, S. (2023, July 5). Crows can understand probability like primates do. *New Scientist*. https://www.newscientist.com/article/2381335-crows-can-understand-probability-like-primates-do

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). Appleton.

Roberts, W. A., MacDonald, H., & Lo, K. H. (2018). Pigeons play the percentages: computation of probability in a bird. *Animal Cognition, 21*(4), 575-581. https://doi.org/10.1007/s10071-018-1192-0

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536. https://doi.org/10.1038/323533a0

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks. *Psychological Review, 88(2)*, 135-170. https://doi.org/10.1037/0033-295X.88.2.135

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning* (2nd ed.). MIT Press.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology, 95*(1), 301-310. https://doi.org/10.1152/jn.00762.2005

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412*(6842), 43-48. https://doi.org/10.1038/35083500

Wagner, A. R. (1981). A model of automatic memory processing in animal behaviour. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Lawrence Erlbaum Associates, Inc.

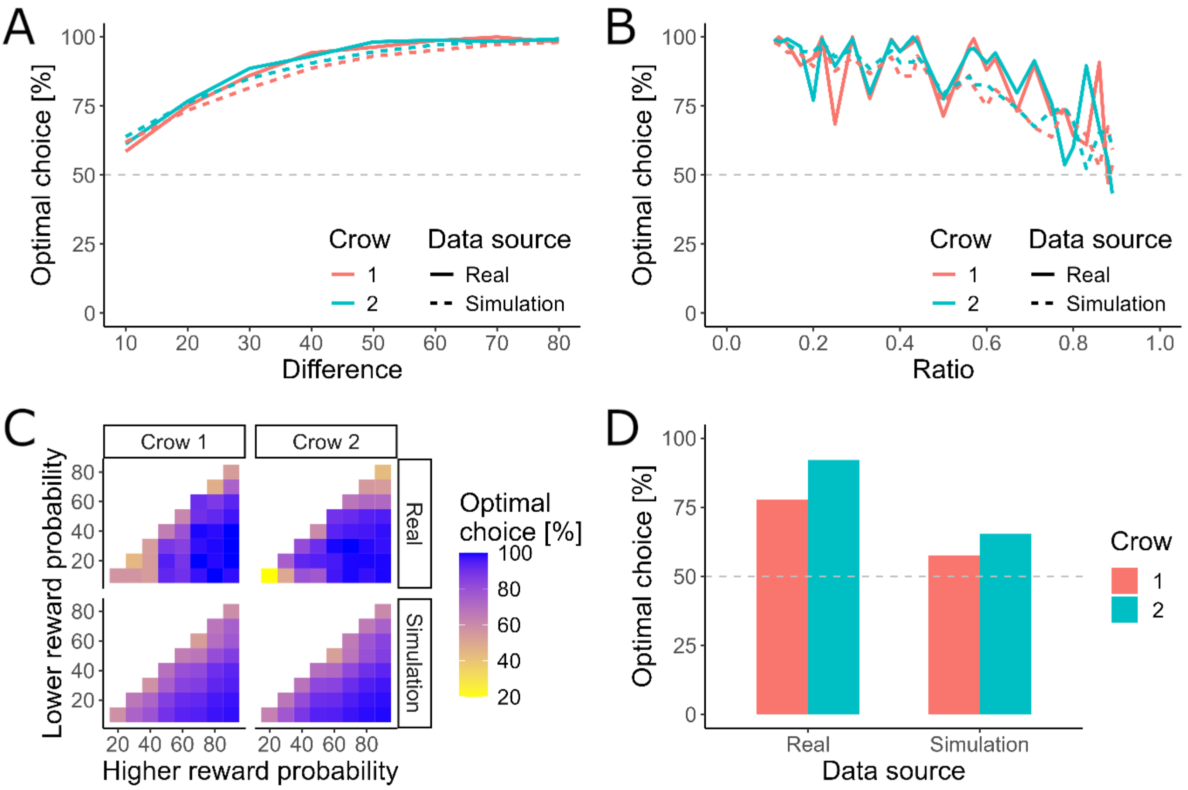
Wascher, C. A. F. (2023). Cognition: Crows are natural statisticians. *Current Biology,* 33(15), R808-R810. https://doi.org/10.1016/j.cub.2023.06.073

Wickham, H. & Bryan, J. (2019). *readxl: Read Excel files*. https://CRAN.R-project.org/package=readxl

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

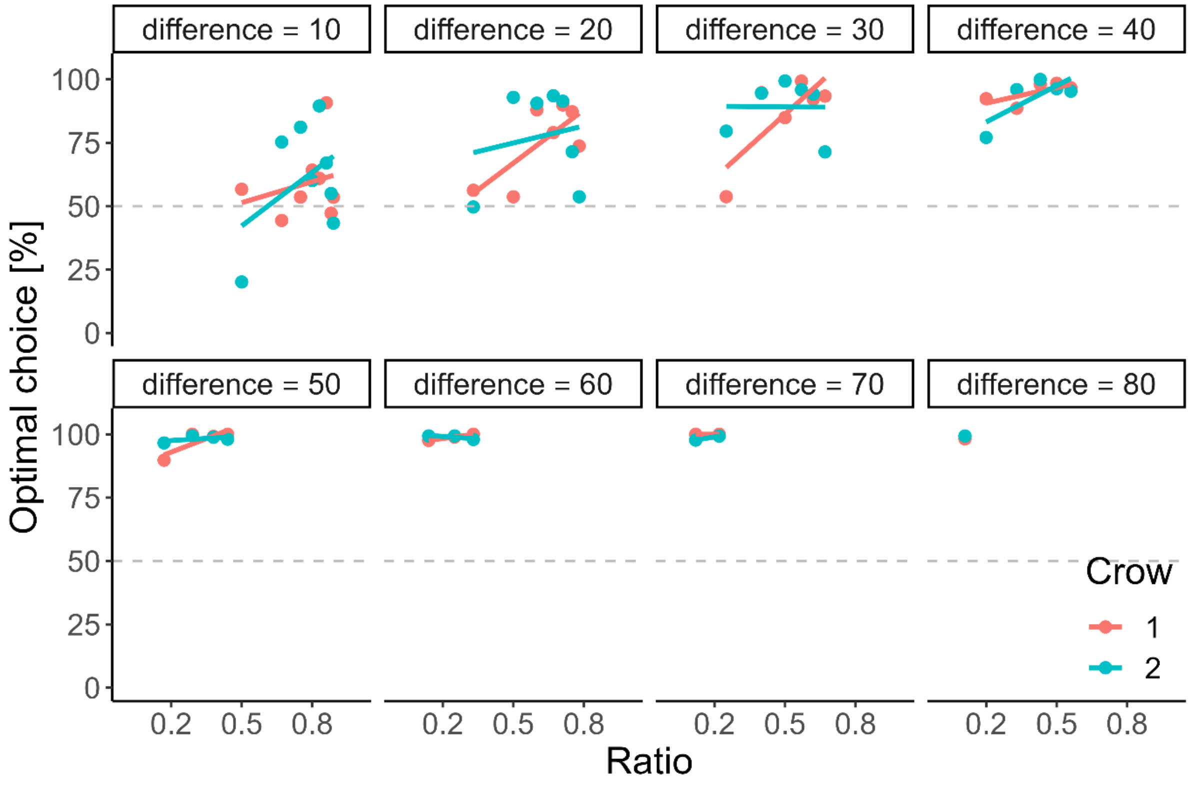
Figures

Figure 1. Simple reinforcement learning provides a good fit to the performance of real crows in a statistical inference task (Johnston et al., 2023).



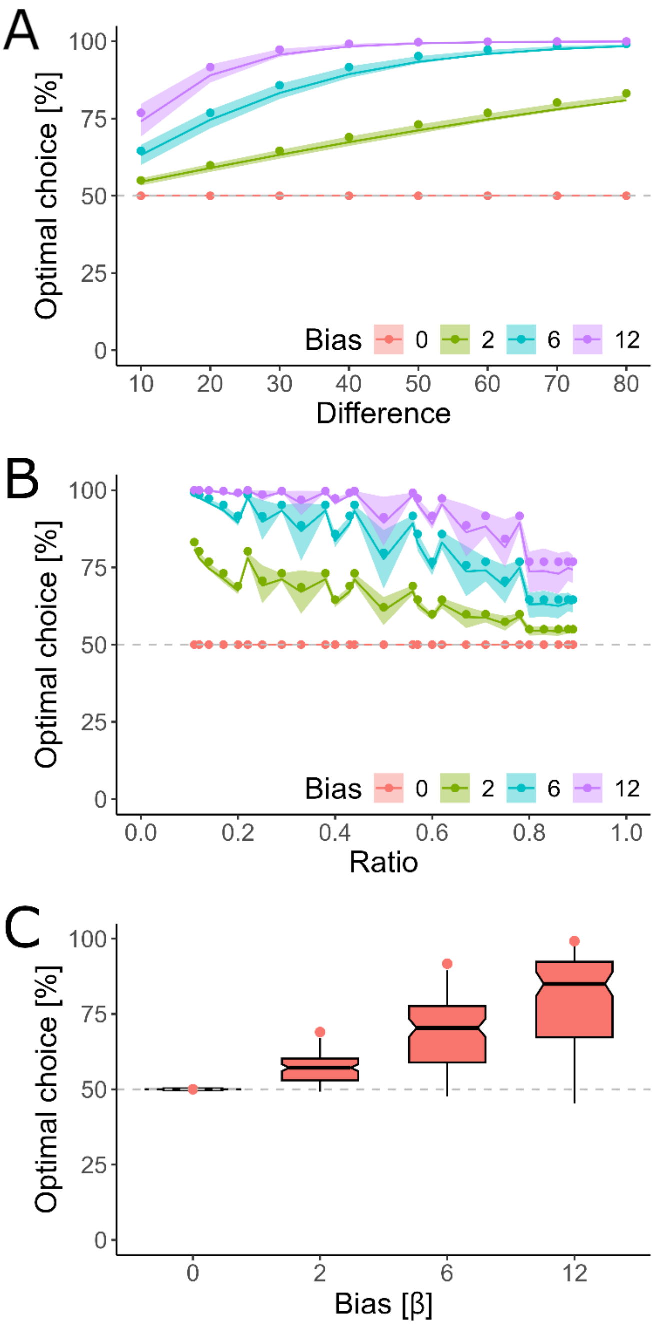
Note. (A) The probability of making an optimal choice [i.e., choosing the stimulus with higher reward probability] is a function of the absolute difference in the reward probabilities associated with two stimuli for real crows and simulated crows given equivalent training. (B) Real and simulated crows display similar variability in choice preference across the different stimulus pairings. (C) Both real and simulated crows show an apparent magnitude effect where optimal choice performance is generally better when the ratio between reward probabilities is low, but this relationship is not monotonic. (D) In a control task, two stimuli were paired with reward equally often, but differed in the number of times they were presented without reward – such that one stimulus had a reward probability of 40%, and the other had a reward probability of 80%. Both real and simulated crows were more likely to select the stimulus associated with higher reward probability [optimal choice > 50%]. Data for real crows are taken from Johnston et al. (2023), and correspond to data shown in Figures 3 and 4 (p.3241 and 3242) of their article.

Figure 2. Johnston et al's (2023) crows did not display a true magnitude effect.



Note. Each panel shows the performance of Johnston et al's (2023) crows as a function of the ratio of probability values of the stimuli in the pair (calculated as lower probability divided by higher probability) for all comparisons between stimuli with the same absolute difference in reward probability. For example, the first panel shows performance as a function of ratio for all comparisons with an absolute difference of 10% (i.e., 10% vs 20%, 20% vs 30%, 30% vs 40% etc). Dots in each panel show the probability of making an optimal choice for specific pairings of stimuli, and the solid lines are regression lines for each crow. A magnitude effect would be expressed as poorer performance for comparisons with higher ratios between reward probability within each panel – that is, the regressions lines should have a negative slope. In fact, the relationship between performance and ratio is flat or positive for both crows across the range of differences.

Figure 3. Simulations of the reinforcement learning model for Experiments 1 and 2 of Johnston et al (2023)



Note. (A and B) Performance of 100 simulated crows in Johnston et al.'s (2023) Experiment 1. Lines show the average performance of the 100 simulations, and shaded areas are their interquartile ranges. (C) Performance of 100 simulated crows in Experiment 2. Dots in all panels indicate performance of an 'ideal' system where the associative strength of each stimulus is exactly equal to its programmed reward probability.