Crows flexibly apply statistical inferences based on previous experience

Graphical abstract

Highlights

- Crows learned to associate multiple reward probabilities with arbitrary stimuli

- Crows remembered probabilities to maximize reward during forced choice

- Crows represented probabilities as abstract magnitudes

- Relative rather than absolute reward frequency usage implies statistical inference
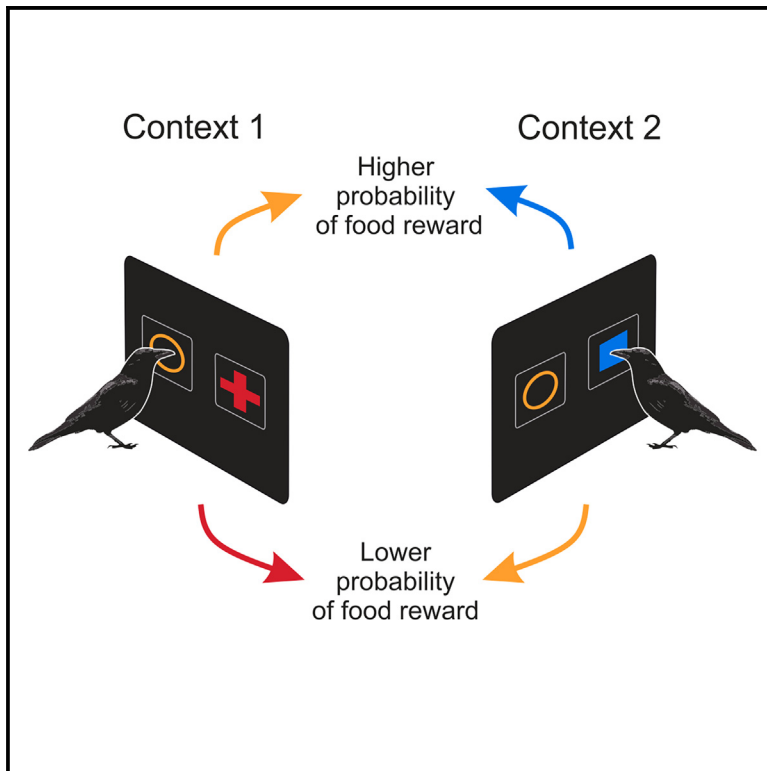
Authors

Melissa Johnston, Katharina F. Brecht, Andreas Nieder

Correspondence

milliej.johnston@gmail.com (M.J.),
andreas.nieder@uni-tuebingen.de (A.N.)

In brief

Johnston et al. show that crows associate different stimuli with specific reward probabilities. When confronted with new stimulus combinations, the crows memorize the associated probabilities and use relative reward frequency to arrive at reward-maximizing decisions, a sign of true statistical inference.

CelPress

Report

Crows flexibly apply statistical inferences based on previous experience

Melissa Johnston,[1,*] Katharina F. Brecht,[1] and Andreas Nieder[1,2,*]
[1]Animal Physiology Unit, Institute of Neurobiology, University of Tübingen, 72076 Tübingen, Germany
[2]Lead contact
*Correspondence: milliej.johnston@gmail.com (M.J.), andreas.nieder@uni-tuebingen.de (A.N.)
https://doi.org/10.1016/j.cub.2023.06.023

SUMMARY

Statistical inference, the ability to use limited information to draw conclusions about the likelihood of an event, is critical for decision-making during uncertainty. The ability to make statistical inferences was thought to be a uniquely human skill requiring verbal instruction and mathematical reasoning.[1] However, basic inferences have been demonstrated in both preliterate and pre-numerate individuals,[2–7] as well as non-human primates.[8] More recently, the ability to make statistical inferences has been extended to members outside of the primate lineage in birds.[9,10] True statistical inference requires subjects use relative rather than absolute frequency of previously experienced events. Here, we show that crows can relate memorized reward probabilities to infer reward-maximizing decisions. Two crows were trained to associate multiple reward probabilities ranging from 10% to 90% to arbitrary stimuli. When later faced with the choice between various stimulus combinations, crows retrieved the reward probabilities associated with individual stimuli from memory and used them to gain maximum reward. The crows showed behavioral distance and size effects when judging reward values, indicating that the crows represented probabilities as abstract magnitudes. When controlling for absolute reward frequency, crows still made reward-maximizing choices, which is the signature of true statistical inference. Our study provides compelling evidence of decision-making by relative reward frequency in a statistical inference task.

RESULTS

Oftentimes we must make decisions about the future with only limited information. Statistical inference is the process of drawing conclusions about a population based on sample data and vice versa. For example, when presented with a jar containing a mixture of preferred and non-preferred candy, what is the likelihood of getting the preferred candy if only one item is randomly selected from the jar? In such statistical inference tasks, subjects can make their decision based on information immediately available to them, that is, they can *see* the probabilities in front of them.[9] Oftentimes, however, one must rely on previous experiences alone when making statistical inferences. In animals, for example, a popular foraging site might yield food on some occasions but not others. Therefore, the decision of whether to visit a particular site will be dependent on the frequency in which it successfully yielded food in the past. Roberts et al.[10] therefore tested whether pigeons could use previous experience alone for decision-making based on learned probabilistic information. In their experiment, pigeons were trained in to associate reward probabilities (25% and 75%) with two arbitrary stimuli. When faced with making a choice between the two stimuli in a forced-choice paradigm, the pigeons overwhelmingly selected the stimulus that predicted the higher chance of reward even after controlling for absolute versus relative reward associations.

While the pigeon data by Roberts et al.[10] is impressive, the ability to make true statistical inferences requires one to relate

different reward probabilities and use probabilistic information flexibly across contexts. That is, one must recognize that a given probability (e.g., reward or successful foraging) may be considered "high," and thus an optimal choice, in one context, but "low" in another. For example, a foraging site with a 60% success rate is higher than that of a 20% success rate but lower than one with a 90% success rate; thus, it is not always the optimal choice. Therefore, the question arises: can birds learn to associate multiple reward probabilities to arbitrary stimuli (i.e., learn multiple sign–reward associations by experience alone) and flexibly apply such knowledge in a reward-maximizing way during a decision-making task as a signature of true statistical inference?

To address this question, we trained two carrion crows to associate nine unique stimuli with reward probabilities ranging in increments of 10% from 10% to 90% reward probability before presenting them with a forced-choice paradigm. Crows represent an ideal bird for investigating such behavior as they make abstract visual associations,[11] differentiate reliable and unreliable stimuli,[12] and demonstrate future planning,[13] all of which are important when making statistical inferences about prospective events. If the crows can successfully combine these components, they should learn to associate multiple reward probabilities with arbitrary stimuli and choose the stimulus with the higher reward probability during the forced-choice paradigm. Additionally, as with other orderly magnitude representations,[14–16] discriminations between reward probabilities should

A

B

C
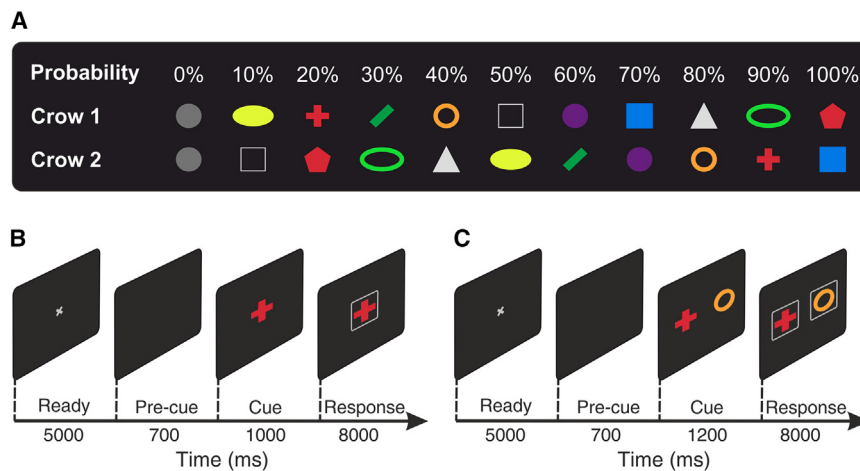
Figure 1. Stimulus reward probability assignments and behavioral protocol

(A) Stimulus–reward probability mappings for each crow.

(B) Schematic of training paradigm. A trial was initiated through activation of the light barrier during a ready period. A subsequent 700 ms pre-cue period was followed by the presentation of cue (1,000 ms), which was associated with a reward probability. During an 8,000 ms response period, a gray border appeared around the cue, indicating the subject could make a response. Appropriately responding to the stimulus (by pecking it) resulted in reward with the assigned probability associated with the stimulus, i.e., crows were rewarded 50% of the time after pecking the stimulus associated with the 50% reward probability.

(C) Schematic of forced-choice paradigm used for the experiments in which the crows were required to respond to one of two stimuli.

reflect psychophysical signatures such as 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. Our results confirm all these hypotheses in crows.

Overall performance

We tested whether crows make statistical inferences by selecting the optimal stimulus (i.e., the stimulus with the larger reward probability) in a forced-choice paradigm (Figure 1). Both Crow 1 and Crow 2 selected the optimal choice significantly more than chance (50%) in every session (Figure 2A; Crow 1: performance $\geq$ 76% optimal choice, all exact binomial tests ps $\leq$ 4.68 $\times$ 10^{-41}; Crow 2: performance $\geq$ 76% optimal choice, all exact binomial tests ps $\leq$ 1.50 $\times$ 10^{-39}). A t test ascertained that there was no significant difference between the overall performance between the two crows ($t(17.251) = 1.6668$, $p = 0.11$). We collapsed performance across the two crows for a repeated measures one-way ANOVA, which revealed no significant differences in performance across sessions ($F(9,10) = 2.1498$, $p = 0.12$)—therefore, all subsequent analyses are collapsed across sessions. That means crows not only reliably chose the optimal stimulus from day one of testing but also maintained the learned probabilities a month later. This supports the notion that crows can learn and flexibly apply probabilistic information learned through previous experience when making statistical inferences.

Learning effects

To ensure performance on the forced-choice paradigm was not a result of associative learning within the first few trials of each condition (i.e., learning to respond left or right based on stimuli configurations), we took a closer look at the first 20 trials of each condition. Upon first presentation of the 72 conditions, Crow 1 made the optimal choice 51 times, which was significantly more than chance (performance = 70.8%, exact binomial test $p = 2.67 \times 10^{-4}$; Figure 2B). Similarly, Crow 2 made the optimal choice 52/72 times, which was also significantly more than chance (performance = 72.2%, exact binomial test $p =$

1.04×10^{-4}). There was a significant moderate correlation between performance and trial number for Crow 2 ($r(18) = 0.63$, $p = 3.19 \times 10$) but no significant correlation in Crow 1 ($r(18) = 0.27$, $p = 0.243$). Overall, low-level associative learning strategies in the forced-choice paradigm cannot explain the high accuracy right from session start.

Distance and magnitude effects

While looking at performance in depth, we found that optimal choices increased as a function of the higher reward probability associated with each condition. That is, crows made the optimal choice more often when the condition included a high ($\geq$50%) reward probability stimulus. However, this could be a result of the distance effect, whereby performance increases as the distance between two stimuli increases—as larger probabilities inherently result in larger differences (e.g., the 90% reward probability stimulus being paired with the 10% stimulus). Indeed, for both crows, optimal choices increased as a function of increasing distance (Figure 3A). At the same time, performance was above chance for all distances (Crow 1: average performance $\geq$ 58.5% optimal choice, all exact binomial tests $p \leq$ 2.99 $\times$ 10^{-11}; Crow 2: average performance $\geq$ 61.3% optimal choice, all exact binomial tests $p \leq 8.06 \times 10^{-18}$). Performance as a function of the distance between reward probabilities with session and bird as random effects was modeled using a binomial general linear model and showed that the distance between the two reward probabilities was predictive of performance, with the larger distances predicting more optimal choices (slope = 0.0764, standard error = 0.002, $p < 8.716 \times 10^{-16}$).

It is important to note that performance for choices of equal distance were not always the same (Figure 3B). For example, performance on a choice between 10% and 20% for Crow 2 was not significantly better than chance (performance = 20%, $p = 1$), whereas performance on a choice between 50% and 60% was (performance = 89%, $p = 4.63 \times 10^{-27}$). Differences in performance across equal distances is a hallmark of the magnitude effect, which states that discriminations between two quantities (e.g., reward probabilities) are determined by the ratio, rather than absolute difference, between them. That
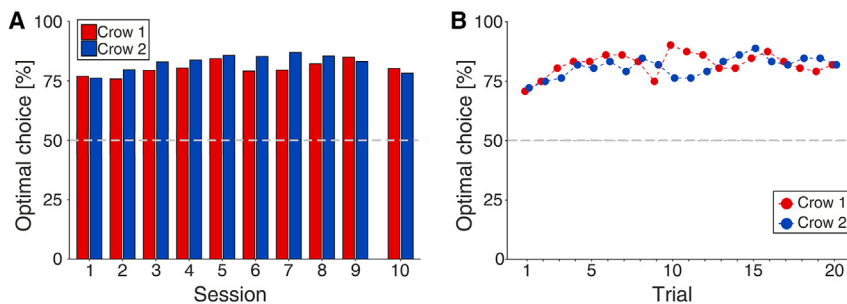
Figure 2. Crows' overall performance
(A) Percentage of trials in each session in which the crows made the "optimal" choice by responding to the stimulus associated with the higher reward probability during the forced-choice paradigm.
(B) Percentage of optimal choices made across the first 20 presentations of each condition. Data points for each bird are offset for visualization purposes only.
For both (A) and (B), dashed line represents chance performance.

is, smaller ratio differences between comparison quantities result in easier discriminations. We found that both crows made more optimal choices when the ratio between reward probabilities was smaller (Figure 3C). Performance was above chance for all ratios (Crow 1: average performance $\geq$ 64.2% optimal choice, all exact binomial tests $ps \leq 8.48 \times 10^{-5}$; Crow 2: average performance $\geq$ 60.1% optimal choice, all exact binomial tests $ps \leq 4.26 \times 10^{-3}$) until 0.875 for Crow 1 and 0.778 for Crow 2, at which point performance was no longer significantly above chance (Crow 1: average performance $\leq$ 53.6% optimal choice, all exact binomial tests $p \geq 0.187$; Crow 2: average performance $\leq$ 55% optimal choice, all exact binomial tests $p \geq 0.078$). Using a binomial general linear model, we examined performance as a function of the ratio between reward probabilities with session and bird as random effects. Indeed, for both crows, the ratio between the two reward probabilities was predictive of performance with the smaller ratios predicting more optimal choices (slope = –2.886, standard error = 0.116, $p < 9.02 \times 10^{-137}$).

Control task

For the control task, we tested whether crows use absolute or relative frequency of reward when choosing a stimulus in a forced-choice paradigm. Here, crows were trained and tested with two new stimuli: one associated with an 80% reward probability and the other a 40% reward probability but presented twice as often during training (Figure 4A). During the test, both birds chose the stimulus associated with the 80% reward probability significantly more than chance (exact binomial test: Crow 1, $p < 4.9 \times 10^{-26}$; Crow 2, $p < 9.8 \times 10^{-48}$, Figure 4B). This finding indicates crows use relative, rather than absolute, frequency when making statistical inferences.

DISCUSSION

### Crows learn probabilistic information from experience

In the current study, we investigated whether crows could learn to associate multiple reward probabilities to arbitrary stimuli (i.e., learn multiple sign–reward associations by experience alone) and flexibly apply such knowledge in a reward-maximizing way during a decision-making task. With minimal training, crows successfully learned to associate nine unique stimuli with reward probabilities ranging from 10% to 90%. When presented with pairs of stimuli in a forced-choice paradigm, crows transferred the learned probabilistic information, such that they selected the reward-maximizing optimal choice significantly more than

chance. Remarkably, crows demonstrated statistical inferences immediately after learning the sign–reward associations. This is true for both the first session and the first presentation of each condition. Although mild improvements were identified in performance across the first 20 presentations of each condition, it is unlikely the crows were using low-level associative learning strategies in the forced-choice paradigm as there were 72 unique pairings (when controlling for onscreen location of stimuli) and only small differences in the number of optimal choices made across the first 20 trials, and crows performed significantly above chance during the very first presentation of each condition. In addition, we show that both crows retained sign–reward associations over time. After 1 month without any further exposure to the task, performance was no different to that of previous sessions. These findings provide conclusive support for the notion that crows learn sign–reward associations from experience alone, use such associations in a combinatorial way to apply relative probabilistic information in a reward-maximizing fashion, and retain flexibility in making statistical inferences over long time scales.

Relative reward frequency versus quantity heuristics

While basic statistical inferences have been demonstrated across many species, it sometimes remains unclear whether subjects use simple quantity heuristics (e.g., avoiding least rewarding stimulus) rather than true statistical inference. Statistical inference is at work if subjects use the relative reward frequency during decision-making. In Experiment 2, we show that our crows used relative rather than absolute frequency of reward during decision-making as a clear signature of statistical inference. Here, the crow experienced the same absolute amount of reward from each stimulus (40% stimulus presented twice as often as 80% stimulus) but made the optimal choice (80%) significantly more than chance. Given that performance on the same pair (80% versus 40%) was similar between Experiment 1 and Experiment 2 (control task), we reason the crows used the same relative reward reasoning for both experiments.

In both human infants and animals, studies on statistical inference at a non-symbolic level are somewhat inconclusive. Denison and Xu[4,5] suggest infants as young as 11 months reason by relative reward frequency when making prospective inferences; however, when Girotto et al.[17] adopted the same paradigm with 3–5-year-olds, only those aged 5 used relative frequency during decision-making. Similarly, Denison et al.[2] and Xu and Garcia[6] both suggest infants can make use of relative reward frequencies; however the studies did not include a
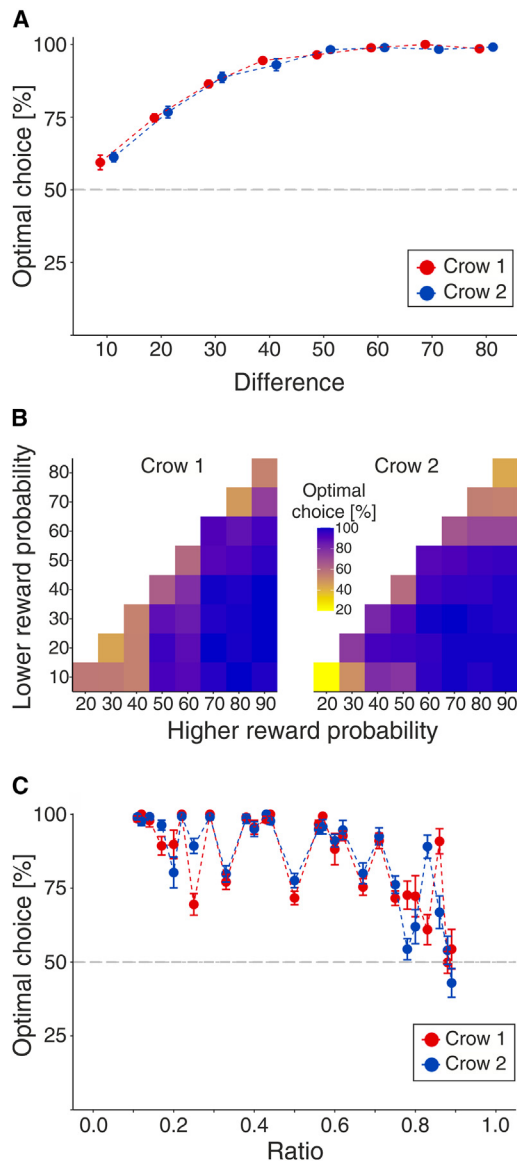
Figure 3. Testing the distance and magnitude effects

(A) Performance as a function of the distance between the two reward probabilities. For both crows, optimal choices increased as a function of increasing distance. Data points for each bird are offset for visualization purposes only, and error bars represent the standard error of the mean.

(B) Performance for choices of equal distance (e.g., 10%) were not always the same across the 36 unique reward probability pairings.

(C) Performance as a function of the distance between the two reward probabilities. Crows made more optimal choices when the ratio between reward probabilities was smaller. Error bars represent the standard error of the mean. For both (A) and (C), dashed line represents chance performance.

control to rule out the possibility that subjects were using quantity heuristics. In monkeys and great apes, the evidence is equally conflicting for decision-making by relative reward frequency tasks.[18–24]

Depending on the temporal trajectory of inference, true statistical inference comes in two flavors: population-to-sample or sample-to-population. For population-to-sample inferences,

one draws a conclusion about a sample based on population information that is either visible to the subject immediately prior to, or at the time of, decision-making. Bastos and Taylor provide an excellent example of prospective inference in a bird species.[9] The kea in their experiment were shown to use relative reward frequency when selecting tokens sampled from two containers. Moreover, kea integrated both physical (barriers in containers) and social (biased sampling from experimenter) information during decision-making.

For sample-to-population inferences, however, one draws a conclusion about the population using sample information, which must be mentally stored and updated for each new sample. Consequently, cognitive demands are significantly higher during sample-to-population inferences and are therefore more difficult to make.[25] In the current task, crows made demanding sample-to-population inferences as they had to retrieve the reward probabilities assigned to arbitrary visual stimuli from long-term memory before they could compare them with other stimuli (and their associated reward probabilities) during the ongoing task. Similarly, pigeons in the Roberts et al. study made sample-to-population inference; they selected the one of two stimuli they had learned to associate with higher reward.[10] However, because the pigeons were trained and tested with only two stimuli and their associated reward probabilities of 25% and 75%, respectively, the pigeons were not required to flexibly relate different reward probabilities and use probabilistic information dynamically across contexts.

We show that crows can indeed update reward probabilities associated with visual stimuli during stimulus–reward learning and use this information flexibly on a single-trial basis when later tested. With the current study, we extend the ability to make sophisticated inferences based on relative reward frequency to a more cognitively demanding sample-to-population inference task. Therefore, our study provides compelling evidence of decision-making by relative reward frequency in statistical inference tasks in animals.

Statistical inferences and analog magnitudes

While language and mathematical reasoning are not prerequisites for making statistical inferences, it is inextricably tied to quantity competence, as one is essentially discriminating between two abstract magnitudes. Underlying quantity discriminations, such as spatial extent, temporal duration, or number, is the analog magnitude system (AMS).[26] The AMS allows for the approximate estimation of quantities and relational magnitudes without language or symbols.[27] When discriminating between multiple relational magnitudes, the 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. Indeed, the AMS has already been well established in numerical discrimination tasks in birds,[14–16,28,29] primates,[30–34] and even elephants[35] and fish.[36,37]

Most of the previous statistical inference literature has used only two reward (outcome) probabilities and has therefore been unable to examine the relationship between the AMS and statistical inference abilities. An exception to this is a study by Eckert and colleagues[19] using chimpanzees and adult humans as subjects. Subjects were given a prospective inference task using 16
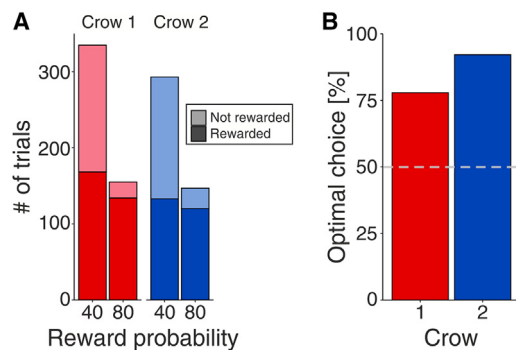
Figure 4. Testing absolute versus relative reward frequency in the control task

(A) The absolute frequency crows were rewarded during training with two new stimuli. One stimulus was associated with an 80% reward probability and the other with a 40% reward probability, with the latter being presented twice as often.

(B) Percentage of trials in which the crows made the "optimal" choice by responding to the stimulus associated with the higher reward probability during the forced-choice paradigm with the newly trained stimuli. Dashed line represents chance performance.

containers differentially weighted (with rewards) and paired to create eight conditions with difference ratios ranging from 0.25 to 0.816. As expected, both humans and chimpanzees selected the optimal choice more often when difference ratios were smaller (i.e., magnitude effect). While Eckert et al.[19] provide the first evidence of the AMS in a prospective statistical inference task by having each container paired with only one other, there were relatively few conditions for investigating the AMS, and subjects did not need to use probabilistic information in a flexible manner.

In our study, we paired each of our nine reward probabilities with one another resulting in 36 conditions with difference ratios ranging from 0.11 to 0.88. Moreover, given reward probabilities from 20% to 80% could serve as both the higher and lower reward probability, crows were required to use probabilistic information flexibly across contexts. Performance by crows was subject to both the distance and magnitude effects. Moreover, we found that both the distance and ratio differences between probabilities was predictive of performance using a binomial general liner model. Therefore, we provide the first evidence of the AMS in a complex retrospective statistical inference task in which the probabilistic information is context dependent. Perhaps more importantly, we expand our understanding of how quantity information might be represented, as we demonstrate that the AMS is important for not only quantities represented as an absolute value[19] but also as a highly abstract probabilistic value. Finding the AMS at work for statistical inference in such distantly related species as primates and crows suggests that it is of adaptive value in the animal kingdom irrespective of phylogenetic relationship.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Subjects
- METHOD DETAILS
  - ○ Apparatus
  - ○ Behavioral protocol
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Behavioral analysis

REFERENCES

1. Piaget, J., and Inhelder, B. (1975). The Origin of the Idea of Chance in Children (WW Norton).
2. Denison, S., Reed, C., and Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: evidence from 4.5-and 6-month-olds. Dev. Psychol. 49, 243–249.
3. Denison, S., and Xu, F. (2010). Integrating physical constraints in statistical inference by 11-month-old infants. Cogn. Sci. 34, 885–908.
4. Denison, S., and Xu, F. (2014). The origins of probabilistic inference in human infants. Cognition 130, 335–347.
5. Xu, F., and Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. Cognition 112, 97–104.
6. Xu, F., and Garcia, V. (2008). Intuitive statistics by 8-month-old infants. Proc. Natl. Acad. Sci. USA 105, 5012–5015.
7. Fontanari, L., Gonzalez, M., Vallortigara, G., and Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. Proc. Natl. Acad. Sci. USA 111, 17075–17080.
8. Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., and Call, J. (2014). Apes are intuitive statisticians. Cognition 131, 60–68.
9. Bastos, A.P.M., and Taylor, A.H. (2020). Kea show three signatures of domain-general statistical inference. Nat. Commun. 11, 828.
10. Roberts, W.A., MacDonald, H., and Lo, K.H. (2018). Pigeons play the percentages: computation of probability in a bird. Anim. Cogn. 21, 575–581.
11. Veit, L., and Nieder, A. (2013). Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. Nat. Commun. 4, 2878.
12. Wascher, C.A.F., Hillemann, F., Canestrari, D., and Baglione, V. (2015). Carrion crows learn to discriminate between calls of reliable and unreliable conspecifics. Anim. Cogn. 18, 1181–1185.
13. Boeckle, M., Schiestl, M., Frohnwieser, A., Gruber, R., Miller, R., Suddendorf, T., Gray, R.D., Taylor, A.H., and Clayton, N.S. (2020). New

Caledonian crows plan for specific future tool use. Proc. Biol. Sci. *287*, 20201490.

14. Ditz, H.M., and Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. Proc. Biol. Sci. *283*, 20160083.

15. Bogale, B.A., Kamata, N., Mioko, K., and Sugita, S. (2011). Quantity discrimination in jungle crows, Corvus macrorhynchos. Anim. Behav. *82*, 635–641.

16. Kirschhock, M.E., Ditz, H.M., and Nieder, A. (2021). Behavioral and neuronal representation of numerosity zero in the crow. J. Neurosci. *41*, 4889–4896.

17. Girotto, V., Fontanari, L., Gonzalez, M., Vallortigara, G., and Blaye, A. (2016). Young children do not succeed in choice tasks that imply evaluating chances. Cognition *152*, 32–39.

18. Eckert, J., Rakoczy, H., Call, J., Herrmann, E., and Hanus, D. (2018). Chimpanzees consider humans' psychological states when drawing statistical inferences. Curr. Biol. *28*, 1959–1963.e3.

19. Eckert, J., Call, J., Hermes, J., Herrmann, E., and Rakoczy, H. (2018). Intuitive statistical inferences in chimpanzees and humans follow Weber's law. Cognition *180*, 99–107.

20. Eckert, J., Rakoczy, H., and Call, J. (2017). Are great apes able to reason from multi-item samples to populations of food items? Am. J. Primatol. *79*, e22693.

21. Eckert, J., Rakoczy, H., Duguid, S., Herrmann, E., and Call, J. (2021). The ape lottery: Chimpanzees fail to consider spatial information when drawing statistical inferences. Anim. Behav. Cogn. *8*, 305–324.

22. Placì, S., Eckert, J., Rakoczy, H., and Fischer, J. (2018). Long-tailed macaques (Macaca fascicularis) can use simple heuristics but fail at drawing statistical inferences from populations to samples. R. Soc. Open Sci. *5*, 181025.

23. Placì, S., Padberg, M., Rakoczy, H., and Fischer, J. (2019). Long-tailed macaques extract statistical information from repeated types of events to make rational decisions under uncertainty. Sci. Rep. *9*, 12107.

24. Tecwyn, E.C., Denison, S., Messer, E.J.E., and Buchsbaum, D. (2017). Intuitive probabilistic inference in capuchin monkeys. Anim. Cogn. *20*, 243–256.

25. Völter, C.J., and Call, J. (2017). Causal and inferential reasoning in animals. In APA handbook of comparative psychology: Perception, learning, and cognition, J. Call, G.M. Burghardt, I.M. Pepperberg, C.T. Snowdon, and T. Zentall, eds., pp. 643–671.

26. Nieder, A. (2020). The adaptive value of numerical competence. Trends Ecol. Evol. *35*, 605–617.

27. Vallentin, D., and Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. Curr. Biol. *18*, 1420–1425.

28. Rugani, R., Fontanari, L., Simoni, E., Regolin, L., and Vallortigara, G. (2009). Arithmetic in newborn chicks. Proc. Biol. Sci. *276*, 2451–2460.

29. Kobylkov, D., Mayer, U., Zanon, M., and Vallortigara, G. (2022). Number neurons in the nidopallium of young domestic chicks. Proc. Natl. Acad. Sci. USA *119*. e2201039119.

30. Beran, M.J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (Pan troglodytes). J. Comp. Psychol. *115*, 181–191.

31. Beran, M.J. (2012). Quantity judgments of auditory and visual stimuli by chimpanzees (Pan troglodytes). J. Exp. Psychol. *38*, 23–29.

32. Hanus, D., and Call, J. (2007). Discrete quantity judgments in the great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus): the effect of presenting whole sets versus item-by-item. J. Comp. Psychol. *121*, 241–249.

33. Merten, K., and Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. J. Cogn. Neurosci. *21*, 333–346.

34. Nieder, A., and Miller, E.K. (2004). Analog numerical representations in rhesus monkeys: evidence for parallel processing. J. Cogn. Neurosci. *16*, 889–901.

35. Perdue, B.M., Talbot, C.F., Stone, A.M., and Beran, M.J. (2012). Putting the elephant back in the herd: elephant relative quantity judgments match those of other species. Anim. Cogn. *15*, 955–961.

36. Agrillo, C., Miletto Petrazzini, M.E., and Bisazza, A. (2017). Numerical abilities in fish: a methodological review. Behav. Processes *141*, 161–171.

37. Messina, A., Potrich, D., Perrino, M., Sheardown, E., Miletto Petrazzini, M.E., Luu, P., Nadtochiy, A., Truong, T.V., Sovrano, V.A., Fraser, S.E., Brennan, C.H., and Vallortigara, G. (2022). Quantity as a Fish Views It: Behavior and Neurobiology. Front. Neuroanat. *16*, 943504.

38. Hoffmann, A., Rüttler, V., and Nieder, A. (2011). Ontogeny of object permanence and object tracking in the carrion crow, Corvus corone. Anim. Behav. *82*, 359–367.

39. Webster, M.M., and Rutz, C. (2020). How STRANGE are your study animals? Nature *582*, 337–340.

40. Veit, L., Pidpruznykova, G., and Nieder, A. (2015). Associative learning rapidly establishes neuronal representations of upcoming behavioral choices in crows. Proc. Natl. Acad. Sci. USA *112*, 15208–15213.

41. Ditz, H.M., and Nieder, A. (2020). Format-dependent and format-independent representation of sequential and simultaneous numerosity in the crow endbrain. Nat. Commun. *11*, 686.

42. Wagener, L., and Nieder, A. (2020). Categorical Auditory Working Memory in Crows. iScience *23*, 101737.

43. Nieder, A., Wagener, L., and Rinnert, P. (2020). A neural correlate of sensory consciousness in a corvid bird. Science *369*, 1626–1629.

44. Wagener, L., and Nieder, A. (2023). Categorical representation of abstract spatial magnitudes in the executive telencephalon of crows. Curr. Biol. *33*, 2151–2162.e5.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Corvus corone	University of Tübingen, Institute of Neurobiology	crow 1, crow 2
Software and algorithms		
NIMH Cortex	National Institute of Mental Health	c595; https://www.nimh.nih.gov/labs-at-nimh/research-areas/clinics-and-labs/ln/shn/software-projects.shtml
R version 4.2.2	R	https://www.r-project.org/

<div align="right">a</div>

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead authors, Andreas Nieder (andreas.nieder@uni-tuebingen.de).

Materials availability
This study did not generate new unique reagents.

Data and code availability

- Data and code to reproduce all statistical analyses are published under the Open Science Framework: https://datadryad.org/stash/share/iAogVMtOcm53SjMNZbNVsMVR4UWmKTutydJWiSPvz-w.
- This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects
Two male carrion crows (*Crovus corone*; Crow 1: 8 years, Crow 2: 7 years) served as subjects in the current study. The crows were genetically unrelated and had different hatching dates. Both crows were hand-raised and housed in groups of up to four conspecifics of mixed sex and age in enriched indoor aviary (360 cm long x 240 cm wide x 300 cm high) with daylight.[38] The crows were on a controlled feeding protocol such that they earned food during daily sessions and were supplemented afterwards when necessary. Water was available ad libitum in the aviaries and during daily sessions. All procedures were carried out in accordance with European law and the Guidelines for the Care and Use of Laboratory Animals from the National Institutes of Health, and were approved by the responsible national authorities (Regierungspräsidium Tübingen).

Acknowledging the STRANGE framework,[39] we add that both crows had previously participated in a number of stimulus detection tasks and (working) memory tasks involving the use of the touch screen apparatus (e.g.,[40–44]). Prior to this study, they had never worked on reward probabilities. Individuals were included in behavioral studies if they could be easily trained on the aforementioned tasks. Only the two crows that were accidentally available for this study (i.e., not engaged in other studies) were trained on the tasks, and both crows sampled for the study were used in the analyses. Due to their rearing and training history, the crows in our study may show some STRANGEness as defined by Webster and Rutz.[39]

METHOD DETAILS

Apparatus
Crows were trained and tested in darkened operant chambers. The chamber was equipped with an infrared light barrier that tracked a reflective foil attached to the crows' heads. At the front of the chamber was a touch screen monitor (3M. Microtouch, 15'', 60-Hz refresh rate) with a viewing distance of 14 cm. Crows responded by pecking stimuli presented on the touchscreen. Speakers emitted

sound to indicate the outcome of the trial. Rewards (birdseed pellets) were delivered via an automated feeder located under the touch screen.

Behavioral protocol
Experiment 1
Crows underwent two initial training phases prior to testing. For the first training phase, the crows were introduced to nine stimuli associated with various probabilities of reward (from 10–90%), with each crow having a unique stimulus mapping (Figure 1A). Each trial began with a 500 ms inter-trial-interval (ITI) followed by a 5000 ms ready period, which was signaled by a small white cross in the middle of the screen (Figure 1B). During the ready period, the crow needed to position their heads within a light barrier. Next was a 700 ms pre-cue period, followed by the presentation of the stimulus (cue period) in the middle of the screen. The stimulus was presented for 1000 ms before a gray border appeared around it for up to 8000 ms (response period). The crow was required to wait until the gray border appeared before making a response to the stimulus. Trials in which responses were made prior to the onset of the gray border were terminated immediately without reward. A trial was also terminated without reward if the crow failed to make a response within the 8000 ms response period. Appropriately responding to the stimulus (by pecking it) resulted in reward with the assigned probability associated with the stimulus, i.e., crows were rewarded 50% of the time after pecking the stimulus associated with the 50% reward probability. Training sessions lasted as long as the crow was still engaging with the task, thus each session consisted of 360–720 trials for Crow 1 and 360–701 trials for Crow 2. Crows participated in one training session per day. Crow 1 was trained for 10 days resulting in 5229 training trials, while Crow 2 was trained for 11 days resulting in 5202 training trials. The order of stimulus presentation was randomized within a session, with each stimulus being presented between 570–588 and 573–581 times for Crow 1 and Crow 2, respectively, across the entirety of the training. The correlation between experienced and ascribed probabilities during training for Crow 1 and Crow 2 were $r = 0.998$ and $r = 0.997$, respectively.

For the second phase of training, the birds were familiarized with the forced choice paradigm. Here the subjects were required to choose between two stimuli: one rewarded with 100% probability (i.e., always rewarded) and one with a 0% probability (i.e., never rewarded). As with the previous training paradigm, each trial began with a 5000 ms ready period followed by a 700 ms pre-cue period. Next, two stimuli appeared onscreen (one on the left and one on the right), with the position of the rewarded stimulus being randomized from trial-to-trial. After 1200 ms, a gray border appeared around each stimulus signaling the crow was able to make a choice. Crows had 8000 ms to make their choice. As with previous training, trials in which responses were made prior to the onset of the gray border were terminated immediately without reward. A trial was also terminated without reward if the crow failed to make a response within the 8000 ms response period. Crows were rewarded for responding to the 100% stimulus. Crow 1 received 952 trials across three days of training (237–373 per session), while Crow 2 received 1202 trials across four days of training (129–401 per session).

Upon the completion of training, crows were given a forced-choice paradigm using the stimuli from the first training phase (Figure 1C) to test whether they could make statistical inferences based on experience alone. As with the training paradigms, each trial began with a 500 ms ITI followed by a 700 ms pre-cue period. Next, two stimuli appeared. After 1200 ms, a gray border appeared around each stimulus signaling the crow was able to make a choice (response period). Crows had 8000 ms to make their choice. As with initial training, trials were rewarded with the assigned probability, thus, there were no "correct" or "incorrect" trials, only "optimal choices" whereby the stimulus with the higher reward probability was selected. If the crows had acquired statistical knowledge associated with the stimuli, they would respond to the stimulus with the higher reward probability, i.e., respond to the stimulus associated with 60% reward probability when presented alongside the 40% reward probability stimulus. All possible stimulus combinations (stimulus pairings and onscreen location of higher reward probability stimulus) were used, resulting in 72 unique stimulus pairings (conditions). The order of condition presentation was randomized from trial-to-trial, with each condition being presented between 59–122 (Crow 1) and 23–125 (Crow 2) times per session. Testing sessions lasted as long as the crow was still engaging with the task, thus each session consisted of 461–722 trials for Crow 1 and 536–652 trials for Crow 2. Crows participated in one session per day for nine consecutive days and one additional day a month later. Across all 10 days, Crow 1 received 5706 trials and Crow 2 received 6024 trials. The correlation between experienced probability and ascribed probabilities for Crow 1 and Crow 2 were $r = 0.999$ and $r = 0.997$, respectively.

Experiment 2
Following Experiment 1, we introduced a control task as Experiment 2 to determine whether the crows were using the absolute or relative frequency of reward during decision-making. That is, were the crows simply selecting the stimulus they received more rewards from or avoiding those that resulted in less reward. To that end, crows were trained with two new stimuli associated with 80% and 40% reward probabilities, with the latter presented twice as often resulting in the same absolute number of rewards as the 80% reward probability stimulus. A yellow triangle and pink vertical bar served as the 80% and 40% reward probability stimuli, respectively for Crow 1, while the opposite was true for Crow 2. The general training protocol was the same as that used during initial training, however, here we used a 400 ms ITI, 5000 ms ready period, 700 ms pre-cue period, 1000 ms cue period, and 8000 ms response period. Additionally, the cue was presented on either left or right side of the screen (rather than center), with the order randomized. Crows participated in one training session resulting in 490 and 440 trials for Crow 1 and Crow 2, respectively. After training, crows were tested on a forced choice paradigm with the newly trained 80% and 40% reward probability stimuli. For the test, we used a 500 ms ITI, 5000 ms ready period, 700 ms pre-cue period, 1200 ms cue period, and 8000 ms response period. Crows completed one testing session with 343 and 245 trials for Crow 1 and Crow 2, respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analysis

All data were analyzed using R version 4.2.2. Data was organized using the tidyr (v1.2.1) and dplyr (v1.0.10) packages. Base R functions and the lme4 package (v1.1.13) were used for statistical tests. Figures were plotted using the ggplot2 (v3.4.0) and cowplot (v1.1.1). For Experiment 1 we first checked whether the experienced probability for a given stimulus correlated with the ascribed probability. To test whether crows could make statistical inferences, we recorded the percentage of trials in which crows made an optimal choice (selected the higher probability stimulus) during the forced choice paradigms. Terminated trials in which the crows left the light barrier prematurely or did not make a response were excluded from analyses. We used exact binomial tests to determine whether optimal choices occurred significantly more than chance. Additionally, for Experiment 1, we used a Welch t-test and a repeated measures one-way analysis of variance (ANOVA) to determine whether there was a difference in performance between the crows or session, respectively. Additionally, we used a Pearson's correlation to determine whether performance increased across the first 20 presentations of each stimulus. Crow performance was further analyzed with binomial general linear models with the logit link function to ascertain whether absolute or relative differences between the two reward probabilities is predictive of optimal choices. Crow and session number were included as random effects to account for the data coming from different birds and daily motivational fluctuations, respectively.