Full Title

Learning to perform auditory discriminations from observation is efficient but less robust than learning from experience

Authors:

Gagan Narula[1,2], Joshua Herbst[1,2], Richard H.R. Hahnloser[1,2*]

Affiliations:

[1] Institute of Neuroinformatics, University of Zurich and ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

[2] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland.

[*] Corresponding author: rich@ini.ethz.ch

Abstract

Social learning enables complex societies. However, it is largely unknown how insights obtained from observation compare with insights gained from trial-and-error, in particular in terms of their robustness. We use aversive reinforcement to train “experimenter” zebra finches to discriminate between auditory stimuli in the presence of an “observer” finch. We find that experimenters are

slow to successfully discriminate the stimuli but immediately generalize their ability to a new set of similar stimuli. By contrast, observers subjected to the same task instantly discriminate the initial stimulus set, but require more time for successful generalization. Drawing upon machine learning insights, we suggest that observer learning has evolved to rapidly absorb sensory statistics without pressure to minimize neural resources, whereas learning from experience is endowed with a form of regularization that enables robust inference.

Introduction

Humans and animals have the remarkable ability to generalize their acquired knowledge to new examples and situations (Pavlov 1927; Markman and Hutchinson 1984; Bass and Hull 1934; Spierings and Ten Cate 2016). For example, they can learn to discriminate threatening from harmless stimuli and they can generalize this knowledge to new instances of a threat. They are also capable of learning from few examples (Cherkin 1969; Bitterman et al. 1983), presumably because brains have evolved under the pressure of fatal consequences when threats are not immediately recognized. Two ethologically relevant learning metrics are thus the acquisition time and the transferability of acquired information. Which forms of learning focus more on the former and which more on the latter of these metrics?

We propose a comparative approach towards disentangling rapid learning from robust generalization, exploiting the fact that many animals are not only capable of learning from aversive or appetitive cues through trial-and-error type processes (Thorndyke 1905; Skinner 1953), but also from observing cues produced by conspecifics and other animals involved in learning or doing the same task (Galef 1988; Zentall 2006; Byrne 2003). In what way does the retained sensory information depend on whether the learning cue is experienced or observed, keeping all other parameters fixed?

We study cue modulation in an auditory stimulus discrimination task involving pairs of zebra finches (Okanoya and Dooling 1990; Sturdy et al. 1999; Tokarev and Tchernichovski 2014; Canopoli et al. 2014). Zebra finches are useful models for sensory learning thanks to their ability to detect subtle differences among highly stereotyped song renditions (Woolley and Doupe 2008). Their learning in stimulus playback experiments is dependent on cues such as behavioral context (Tchernichovski et al. 2001; Derégnaucourt et al. 2013).

Using aversive air-puffs, we trained one of the two birds in a pair to discriminate short from long renditions of a zebra finch song syllable (Go-NoGo avoidance conditioning, Fig. 1A; spectrograms of stimuli in Fig. 1B, durations in Fig 1C). We refer to these birds as "experimenters". Simultaneously, we allowed a paired zebra finch to observe the entire training phase of the experimenter, including the acoustic stimuli and the experimenter's actions. These latter birds are referred to as "observers"; they could engage in unrestricted visual and auditory interactions with experimenters, but did not perform the task until after experimenters completed their training phase.

The acoustic stimuli in such experiments are fully predictive of whether an air-puff is imminent or not. Experimenters reveal their ability to discriminate the stimuli by escaping from the perch before they get struck by the air-puff (Canopoli et al.2014). We refer to this form of learning as experience learning because birds learn to discriminate based on experience of the air-puffs. By contrast, observers could not learn from air-puff experiences, but they could learn from observing the air-puffs' direct and indirect effects on experimenters'. We refer to this form of learning as observation learning (which is not meant to imply that observers learn by imitating the actions of experimenters, which is commonly known as 'observational learning').

We expected that observers would be able to demonstrate their learned discrimination ability in a separate testing phase in which they were exposed to air-puffs. Here, we investigate the performance tradeoffs between experience learning and observation learning using two

measures taken from the machine learning community: learning speed and generalization performance.

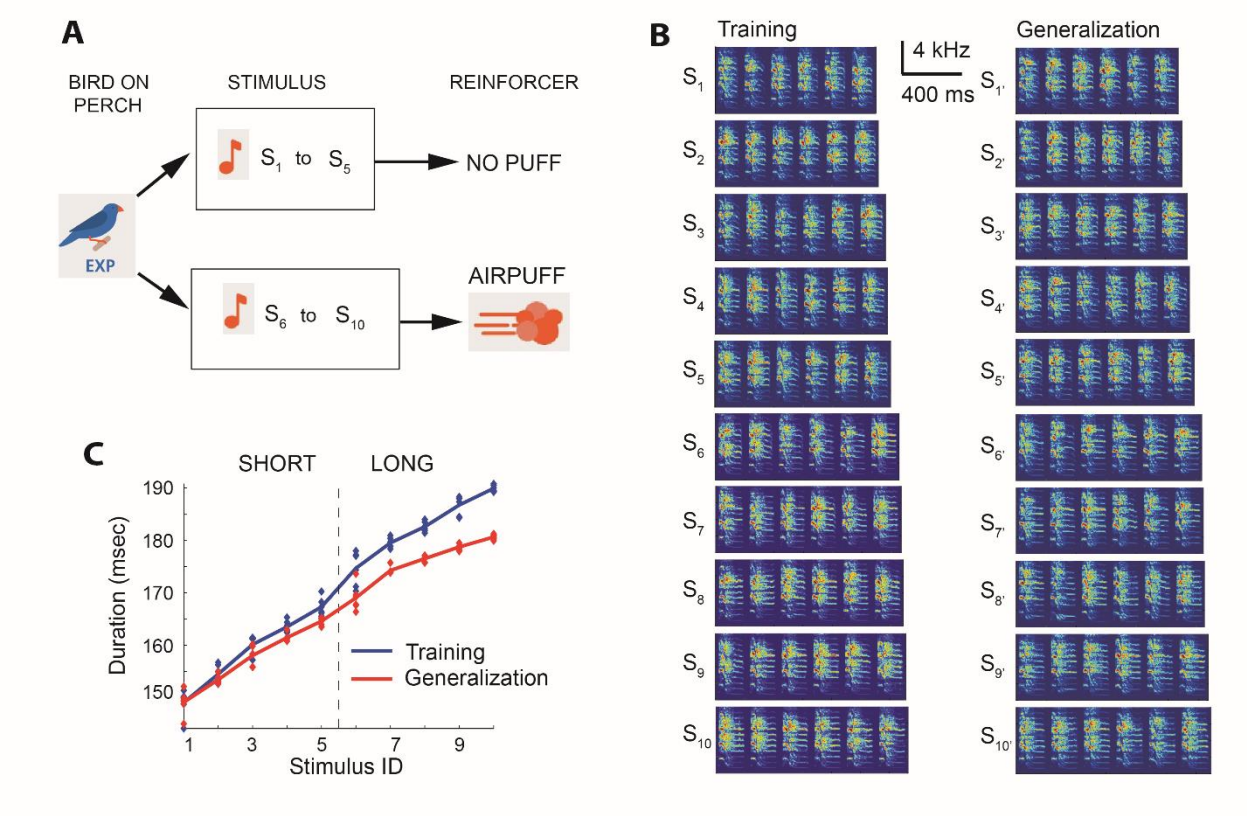

Fig. 1 A Go-NoGo auditory discrimination task. (**A**) When the experimenter (EXP) was on the perch continuously for 3.5 s, an acoustic stimulus S_i (i=1,..,10) was randomly chosen and played through a loudspeaker. The long stimuli S_6 to S_{10} were followed by an air-puff aimed at the experimenter. Experimenters were expected to learn to avoid the air-puffs by escaping the perch in puffed trials, and staying on the perch in unpuffed trials. (**B**) Log-power spectrograms of all ten stimuli in the training set (S_1 to S_{10}, left) and in the generalization set (S'_1 to S'_{10}, right). All stimuli were composed of a string of six renditions of a particular song syllable. (**C**) Syllable durations for the ten stimuli in the training set (blue) and generalization set (red, dots indicate individual syllable renditions). Either the long stimuli or short stimuli were followed by an air-puff.

Results

Observation learning induces rapid expression of auditory discrimination

During a pre-training phase, experimenters (EXP) were accustomed to air-puffs that followed one of two auditory stimuli of different duration. Then a training phase followed, during which we exposed EXP to the full training set of 10 auditory stimuli (Fig. 2A left panel and Supplementary methods). Gradually, EXP learned to escape from the perch more often in puffed trials; their escape probabilities (cumulated from trial onset to trial end) became larger on puffed trials than on unpuffed trials (Fig. 2B left and right). We quantified the birds' ability to discriminate stimulus class by the average difference in (cumulative) escape probabilities (dPesc) between puffed and unpuffed trials (Fig. 2, B, D and F). EXP attained a stringent performance criterion (see Methods and Fig. S1 D) after $4.8 \pm 2.9*10^3$ trials (mean ± std, n=10 birds). This criterion defined the end of the training phase, at which time EXP displayed a dPesc of 0.36 ± 0.06 (mean ± std, dPesc averaged over the last 3 blocks of training, including the criterion block). After the training phase (or observation phase for observers), the experimenter was replaced by the observer (OBS) and a naïve bird was placed in the observer's cage. Then we began testing the OBS using the same pre-training and training paradigms it was previously allowed to observe. We refer to the training of observers as testing, Fig. 2C right panel.

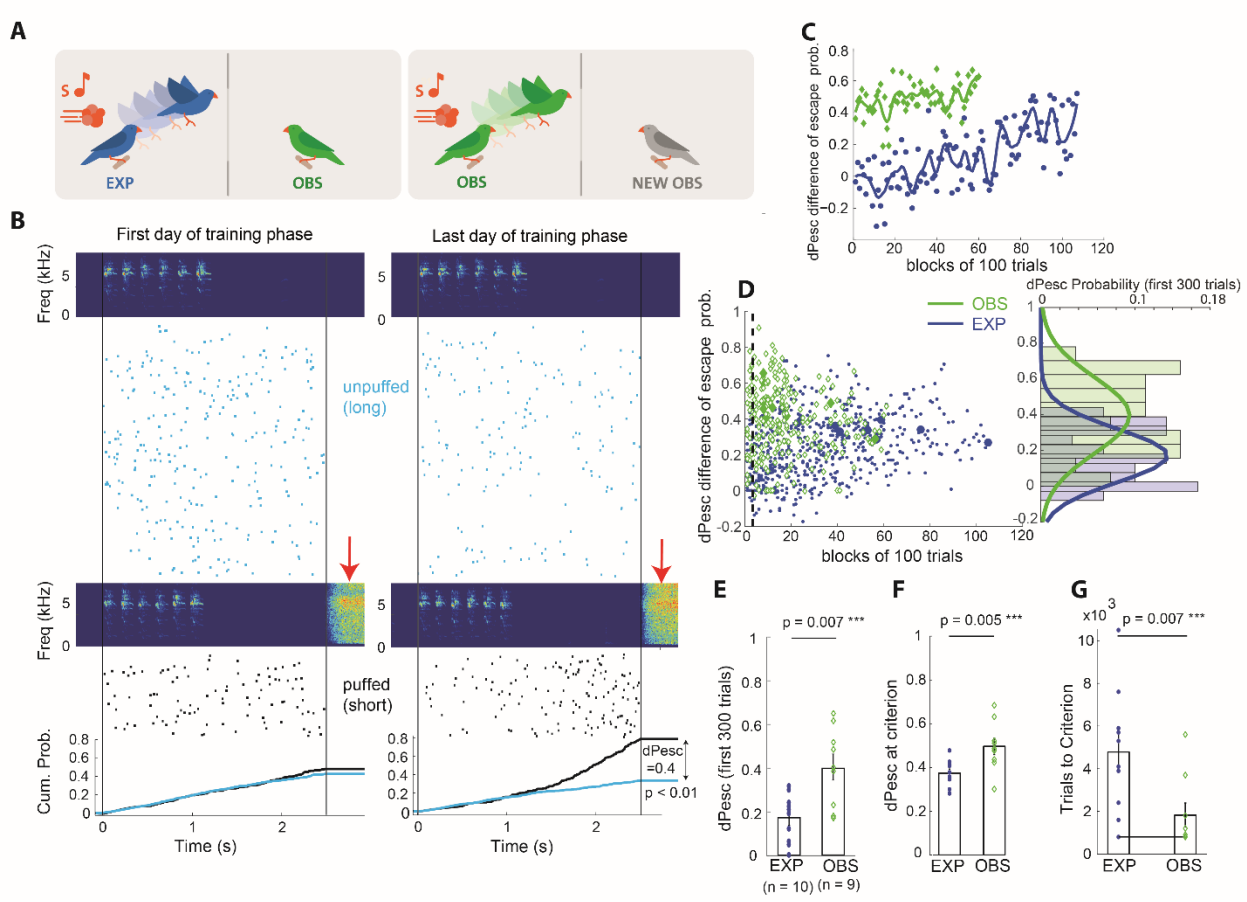

Fig. 2 Rapid learning in observers. (**A**) Experimental design: Observers (OBS) were separated from experimenters (EXP) by a screen that restricted visual interactions to the perch equipped with the air-puff delivery mechanism. During a training phase (left panel), OBS watched EXP perform the task. Thereafter, the knowledge learned by OBS was tested in the same paradigm with a new (naïve) OBS (right panel). (**B**) On the first training day (left), this example EXP showed roughly equal densities of escapes (rasters) for unpuffed (light blue) and puffed trials (black). The air-puff sounds are visible in spectrograms of microphone recordings (red arrows). On the last day of training (right), the cumulative escape probabilities (black and blue lines, bottom) discriminate puffed from unpuffed trials (z-test of individual proportions, $p<0.01$). (**C**) Difference of escape probabilities (dPesc) during training of an EXP (blue) and during testing of its OBS (green). Solid curves are smoothing spline fits (parameter: 0.3). (**D, left**) Scatter plot of dPesc as a function of trial block number (n=10 EXP, blue dots; n=9 OBS, green diamonds). dPesc at the criterion is depicted with a larger, solid symbol. (**D, right**) Probability distributions of dPesc for EXP (blue, n=10) and OBS (green, n=9) in the first 3 blocks (300 trials, up to dashed line),

fitted with Gaussian functions. (**E**) Bar plot of dPesc, showing that OBS (green diamonds, n=9) discriminate significantly better in the first 3 blocks of their testing than EXP (blue circles, n=10) in the first 3 blocks of their training. (**F**) Upon reaching the learning criterion, the average dPesc (3-block average) in OBS is significantly larger than in EXP. (**G**) OBS reach the learning criterion in fewer trials than EXP. The bars indicate averages across single birds, the black line indicates the lower bound (800 trials).

At the beginning of the testing phase (first 3 testing blocks), OBS displayed a significantly higher discrimination performance than EXP at the beginning of their training phase, Fig. 2E. Surprisingly, OBS' initial performance was no worse than that of EXP who had reached the learning criterion (average initial dPesc=0.40 in n=9 OBS vs average final dPesc=0.36 in n=10 EXP). OBS reached the performance criterion very rapidly, in only $1.82 \pm 1.7*10^3$ trials (mean ± std, n=9 birds), less than a third of the trials required by EXP (single-sided Wilcoxon rank sum test with alternative hypothesis: EXP > OBS, p = 0.0075 (not exact), test statistic = 75; Effect size: Cohen's d = 1.224, 95% CI not computed because of ties), Fig. 2G. After reaching the criterion, OBS showed a significantly higher discrimination performance than EXP (average dPesc in OBS: 0.47 ± 0.11; average dPesc in EXP: 0.36 ± 0.06 Wilcoxon rank sum test, p = 0.0056, test statistic = 12; Effect size: Cohen's d = 1.38, 95% CI = [-0.2 -0.031]), Fig. 2F.

Observers generalize poorly compared to experimenters

Experimenters could rapidly generalize their learned knowledge to novel instances of the stimuli (generalization set, spectrograms in Fig. 1B) but observers were not able to do so. To compare generalization in experimenters and observers, first, we allowed generalization observers (GENOBS) to watch generalization experimenters (GENEXP) learn to discriminate the stimuli in the training set, Fig 3A top. After training completion, GENEXP and GENOBS were separated

and placed in different experimental chambers, each paired with a naïve observer. GENEXP were then exposed to the generalization set of stimuli, Fig. 3A bottom left. GENOBS were exposed to a pre-training phase with two stimuli from the training set, followed by a testing phase comprising the ten stimuli from the generalization set, Fig. 3A bottom right. Contrary to our findings on the training set, GENOBS initially showed significantly poorer discrimination on the generalization set (average dPesc over the first 3 blocks in GENEXP: 0.42 ± 0.08 and in GENOBS: 0.2 ± 0.18, $p = 0.024$, test statistic = 66, two tailed Wilcoxon rank sum test, Effect size: Cohen's d= 1.11, 95% CI = [0.016, 0.36]), Fig 3 B, C. GENOBS also took more time than GENEXP to reach criterion ($4.9 \pm 4.0*10^3$ trials in n=9 GENOBS versus $1.1 \pm 0.5* 10^3$ trials in n=9 GENEXP, $p = 0.0064$ (not exact), test statistic = 10, two tailed Wilcoxon rank sum test; Effect size: Cohen's d = 1.3, 95% CI not computed because of ties), Fig 3D.

GENOBS needed significantly more trials to reach the learning criterion than did OBS ($p = 0.044$, test statistic = 17.5, Wilcoxon rank sum test), demonstrating that observers reacted to small differences between stimuli from the training and generalization sets. However, after reaching the criterion, OBS and GENOBS discriminated the stimuli equally well (similar dPesc at criterion, $p = 0.077$, test statistic = 20, Wilcoxon rank sum test). Thus, overall, observers seemed to associate the perch-escape behaviors by experimenters much more exclusively with the presented auditory stimuli than did the experimenters themselves, who associated the air puffs more inclusively with the stimuli (to include similar stimuli from the generalization set).

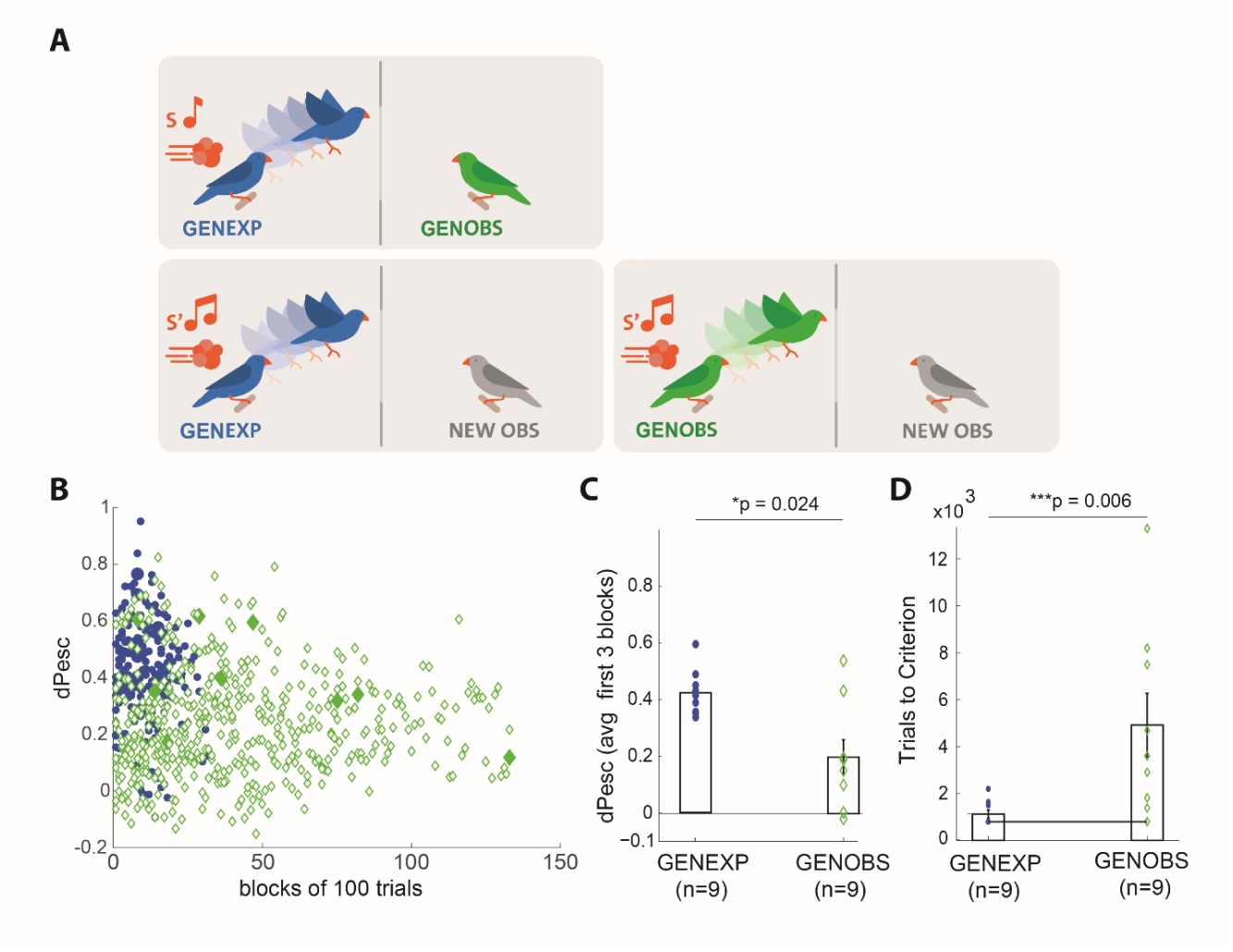

Fig. 3 Observers are poor generalizers. (**A**) Generalization experimenters (GENEXP) undergo the same training phase as EXP (training set of stimuli, S, top), after which they are exposed to the generalization set of stimuli S' during the testing phase (bottom left). Generalization observers (GENOBS) first observe the training set of stimuli (S, top) and then are tested on the generalization phase (S', bottom right). (**B**) Scatter plot of dPesc on the generalization set as a function of block number (100 trials per block) in all birds (n=9 GENEXP, blue dots; n=9 GENOBS, green diamonds), the criterion block is represented by larger solid symbol. (**C**) GENEXP discriminated stimuli in the generalization set better than GENOBS. Symbols indicate dPesc averaged across the first 3 blocks of the testing phase, bars represent averages across animals. (**D**) GENEXP reached the criterion faster than GENOBS. Symbols indicate trials to criterion, bars represent averages.

We inspected the escape behaviors of observers and experimenters. We found that after reaching the learning criterion, EXP and OBS displayed similar perch escape strategies. That is, they tended to abruptly increase their perch escape rates just before air-puff onsets, Fig. S3A, B. This similarity of behavior suggests that observers might learn from experimenters' actions.

Observers do not learn through passive perceptual processes

We set out to characterize the requirements for observation learning. We hypothesized that OBS learned from experimenters' actions in response to the air-puffs. To test whether observers indeed learned from experimenters' actions, we allowed experimenter and observer pairs to experience several thousand ($7.5 \pm 3.6*10^3$) stimulus playbacks including the sound of air-puffs, but not the tactile sensation of the puffs. We realized this perceptual paradigm by directing the air outlet away from the experimenters, Fig. 4 A. Consequently, experimenters never experienced the air-puff as a force against their body. We refer to observers in such pairs as perceptual learners (PLs), because they could potentially learn from the pairing of stimuli with air-puff sounds.

Experimenters in this perceptual paradigm never produced dPesc values different from 0 (average dPesc after 5000 training trials in 3 experimenters: [-0.065, -0.002, 0.007], p=0.81, p=0.25, p=0.64, respectively; z-test of individual proportions), hence they did not show the discriminative behavior that we suspected would drive learning in observers. When we tested PLs (n=7 birds) with air-puffs directed at them, they needed significantly more trials to reach criterion than OBS ($6.32 \pm 6.3*10^3$ trials in PLs versus $1.82 \pm 1.7*10^3$ in OBS; single-sided Wilcoxon rank sum test of alternative hypothesis PL > OBS, p = 0.008 (not exact), test statistic = 8.5; Effect size: Cohen's d = 1.036, 95% CI not computed because of ties), Fig. 4E. PLs were

slower than OBS even after removing an outlier bird (trials to criterion = 20300) in the PL group ($p = 0.016$, Cohen's d: 1.29). PL performance at criterion was comparable to OBS performance (0.32 ± 0.2 in PL versus 0.47 ± 0.11 in OBS, $p = 0.142$, test statistic = 46, Wilcoxon rank sum test, Cohen's $d = 0.93$, 95% CI = [-0.077 0.338]). The absence of rapid learning in PLs suggests that learning in OBS required an experimenter engaged in the task and responding to air puffs.

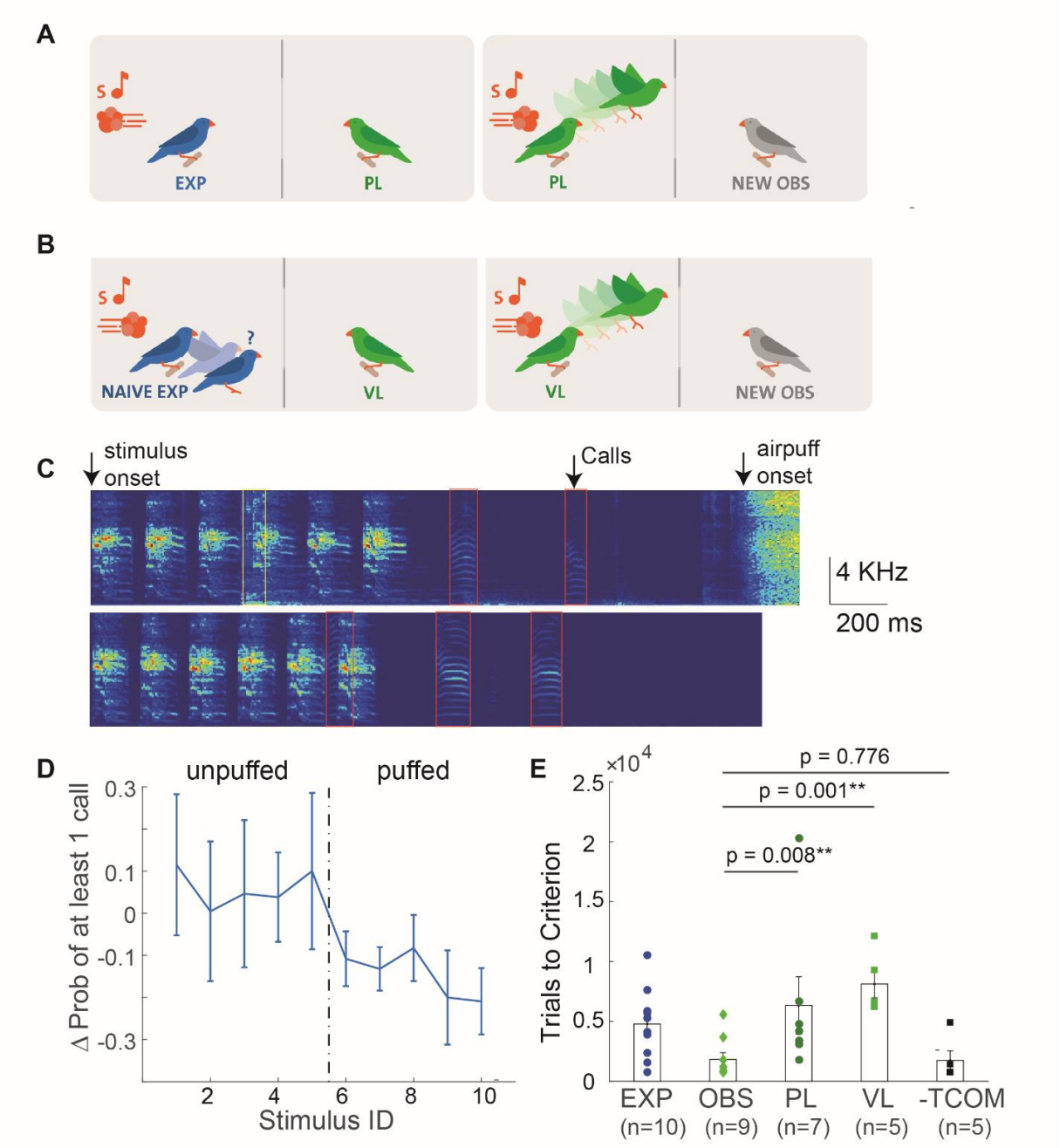

Fig. 4 Observers learn from behaving, expert experimenters, even in the absence of vocal interactions. (**A**) Perceptual Learners (PLs, n=7 birds) first observed a naïve experimenter trigger several thousand trials in which the air-puff was directed away from the experimenter's body (left). Thereafter, they were tested using air puffs (right). (**B**) Valence Learners (VLs, n=5 birds) observed experimenters that never reached the criterion (left). Additionally, three VLs were exposed to stimulus-contingent air puffs prior to observation. Thereafter, VLs were tested just like OBS (right). **(C)**

Spectrograms of microphone recordings of puffed (top) and unpuffed (bottom) trials. Vocal exchanges (calls, red rectangles) frequently occurred during the task. Wing flaps are also audible (yellow rectangle). **(D)** Difference in the probability of observing a call in the delay period and stimulus period (Delay – Stimulus) for the ten stimuli (n = 6 observers). Here, S1 – S5 are unpuffed. **(G)** Both PL (circles) and VL (squares) required significantly more trials than observers (light green diamonds) to reach criterion during the testing phase (OBS < PL, p = 0.008; OBS < VL, p = 0.001). Observers deprived of acoustic communication with experimenters during trial times are as quick as OBS (-TCOM = OBS, p = 0.776, Wilcoxon rank sum test).

Experimenters needed to be expert models to induce learning in observers

We expected observation learning to be most effective when information is provided by an expert. To probe for sensitivity on experimenter performance, we tested a group of Valence Learners (VLs, n=5) that observed naïve experimenters who did not reach the performance criterion within (on average) $6.9 \pm 3.0*10^3$ trials. These naïve experimenters were hit by air puffs on average 539 times out of 1000 puffed trials, and escaped in unpuffed trials on average on 400/1000 trials. In addition, to give VLs direct experience of the reinforcer (its valence), 3/5 VLs were initially exposed to air puffs (approximately 500 strong 1-s air puffs, see Methods). When tested, VLs were much slower than OBS to reach the learning criterion (average number of trials to criterion in VL [n = 5]: $8.0 \pm 2.5 *10^3$ versus OBS [n = 9]: $1.82 \pm 1.7*10^3$, single sided Wilcoxon rank sum test of alternative hypothesis VL > OBS, p = 0.001 (not exact), test statistic = 0, Cohen's d = 3.11, 95% CI not computed), Fig. 4E. The performance of VLs at criterion was lower than the performance of OBS (average dPesc for VL [n = 5]: 0.25 ± 0.11 versus for OBS [n = 9]: 0.47 ± 0.11, p= 0.007, test statistic = 42, Wilcoxon rank sum test, Cohen's d = 1.87, 95% CI = [0.034 0.355]). The poor testing results in VLs suggest that OBS did not learn by predicting the reward value and by converting this prediction into an optimal action during testing. Instead,

VL behavior suggests that OBS focus on experimenters' discriminative actions, which must necessarily contain the information required for observation learning.

Overall, PL and VL behavior was closer to EXP than to OBS. That is, upon reaching the learning criterion, there was no statistically significant difference in performance (dPesc) between EXP and PLs ($p = 0.41$, test statistic = 26, 95% C.I = [-0.2 0.2], Wilcoxon rank sum test; Cohen's d: 0.21) and a trend of higher performance in VL compared to EXP ($p = 0.07$, test statistic = 10, 95% C.I = [-0.21 0.05], Wilcoxon rank sum test; Cohen's d: 1.14). In combination, PLs and VLs emphasize the importance of experimenters' discriminative actions for observation learning.

Vocal exchanges are not required for observation learning

Given the importance of experimenter actions, we speculated that rapid learning in OBS was based on vocal exchanges between EXP and OBS through calls occurring during experimental trials, Fig 4 C. Indeed, on the last day of the training phase, when EXP had reached the learning criterion, we found a difference in calling behavior between puffed and unpuffed trials. In six OBS (on one day each), we inspected calling rates (defined as the probability of observing at least one call) in the stimulus period (from stimulus onset to stimulus offset) and in the delay period (defined from stimulus offset to air-puff onset), Fig. 4 D. In puffed trials, the calling rate was significantly lower in the delay period than in the stimulus period, whereas no such difference was seen in unpuffed trials (puffed trials: stim call probability 0.46 ± 0.26 vs delay call probability 0.27 ± 0.15, difference -0.19; unpuffed trials: stim call probability 0.39 ± 0.22 vs delay call probability 0.42 ± 0.26, difference 0.03; n=6 birds and 10 stimuli divided into two groups, $p = 10^{-6}$, two t-sample t-test, t statistic = 6.29, df = 58). Hence, call reduction could signal the imminent arrival of an air puff.

To test whether observers used calls as a learning cue, we housed experimenters and observers (n=5 pairs) in separate soundproof boxes and gave them visual access to each other by virtue of two adjacent windows. Moreover, to trigger social interest, we allowed birds to vocally interact with each other using a custom digital communication system composed of two microphones and loudspeakers and an echo cancellation filter (Supplementary methods). We suppressed vocal exchanges during stimulus presentation by interrupting the communication system from stimulus onset to air-puff offset. We termed the observers in this paradigm no-trial-communication learners (-TCOM). Despite elimination of vocal interactions during the discrimination task, we found that -TCOM acquired stimulus-discriminative information in amounts comparable to OBS (trials to criterion: -TCOM (n = 5): $1.74 \pm 1.78*10^3$; OBS (n = 9): $1.82 \pm 1.7*10^3$, p = 0.776, test statistic = 25, Wilcoxon rank sum test, 95% CI = [-1200 2300], Cohen's d: 0.05), Fig. 4G. Hence, it follows that OBS did not require immediate vocal interactions. They could learn from visual displays only or from vocal exchanges following trials.

Observation learning was not a simple form of stimulus enhancement

We did not find evidence that simple stimulus enhancement (Zentall 2006; Byrne 2003) could account for observers' rapid discrimination learning. Here, stimulus enhancement is defined as learning in an observing animal through the increased interaction with a particular stimulus after a demonstrating animal has directed its attention to the stimulus. Hoppitt and Laland (Hoppitt and Laland 2013) provide a necessary condition for stimulus enhancement: observers must exhibit higher response rates to enhanced stimuli, which implies that enhancement could reveal itself as an increase in stimulus-contingent escape behavior even when there is no reinforcement.

To probe for stimulus enhancement, we quantified escape behavior during the first 100 pre-testing trials during which two auditory stimuli were paired with air-puffs that were too weak to displace a bird from the perch. During these trials, OBS exhibited a dPesc of 0.09 ± 0.3 that was not significantly different from zero ($p = 0.44$, test statistic = 24, 95% CI = [-0.17 0.44]; Wilcoxon signed rank test), Fig. S4B. Hence, OBS were not initially drawn to either puffed or unpuffed stimuli. Rather, they expressed discriminative behavior only during later trials of the pre-testing phase, after we increased the strength of air-puffs. Thus, it seems that in addition to the stimuli and the actions of experimenters, observers needed also the aversive experience of air puffs to express their learned knowledge.

Logistic regression with L1 norm regularization differentiates observers and experimenters

We sought to identify possible explanations for the difference in generalization abilities between observers and experimenters. In the following, we draw upon insights from machine learning and statistical learning theory to suggest a simple mechanism that can account for the learning characteristics in experimenters and observers. In machine learning, the typical goal is to maximize generalization performance but not learning speed, which is why many state-of-art methods tend to learn slowly (LeCun et al.2015). Typically, undesired overfitting arises when few training examples are classified using too many parameters. Alternatively, performance can be poor on both training and testing data when many examples are classified using too few parameters. These two performance shortcomings imply that there is a tradeoff between learning and generalization known as the 'Bias-Variance dilemma' (Geman et al. 1992). This tradeoff has led researchers to develop regularization methods such as lasso and ridge regression (Tibshirani 1996), maximal margin (Cortes and Vapnik 1995), and dropout (Srivastava et al. 2014). Essentially, regularization methods improve generalization performance

by dynamically regulating the use of parameters and of training data (Srivastava et al. 2014; Tibshirani 1996).

In the context of our findings, these theoretical insights from statistical learning theory suggest that experience is associated with regularization whereas observation is not. We tested the hypothesis that regularization could set the divide between experimenter and observer performances, by training a simple artificial neuron with a logistic activation function to discriminate between the two stimulus sets, Fig. 5A. The neuron received input from a group of at least 22 input neurons tuned to diverse sound features such as amplitude, pitch, duration, and Wiener Entropy, collectively defining the feature set used in Sound Analysis Pro (SAP), a popular birdsong analysis software (Tchernichovski et al. 2000). To model observers, we trained the neuron to fire during puffed stimuli and to remain silent during unpuffed stimuli. We used a gradient descent learning rule that maximizes the likelihood of correct discrimination (Methods). We found that the discriminative performance of the 'observer' neuron increased rapidly to the theoretical limit on the training set, but when we interrupted the training at any time and evaluated the neuron's performance on the testing set, we found poor generalization, Fig. 5B. The reason for poor generalization was that the neuron based its classification on exceedingly many sound features that by chance were slightly informative about the reinforcing air-puff, Fig 5D.

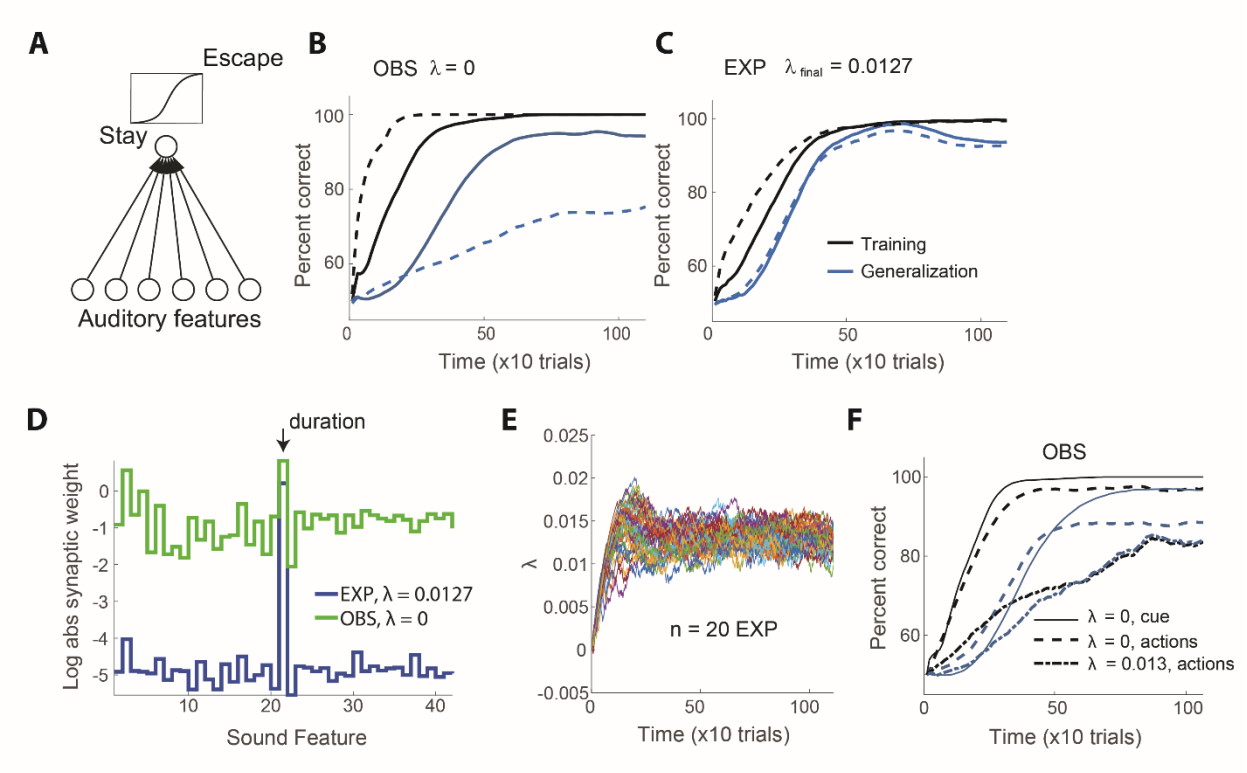

Fig. 5 Regularization can explain the performance differences between experimenters and observers. (**A**) The model neuron triggers escapes from the perch based on the logistic response to a set (here 6) of auditory features. (**B**) When an 'observer' neuron is modeled without L1 regularization (λ=0), the percent correct classification (PCC) on the training set (black line) increases rapidly but the PCC on the generalization set (blue line) increases much more slowly. Adding an extra 100 auditory features of frozen white noise (dashed lines) accentuates the contrast between fast learning and slow generalization. (**C**) When the 'experimenter' neuron is trained with L1 regularization (with dynamic estimation of λ, the final value of λ (on average) = 0.0127), the curves reporting PCC on training and generalization sets increase slowly but at roughly matched rates. Increasing the number of frozen noise inputs has almost no effect on PCC curves (dashed lines). (**D**) In observer neurons, the log absolute synaptic weights are roughly uniformly distributed. In experimenter neurons, the synaptic weights (black) are all near zero except the weight corresponding to syllable duration (auditory feature 21, black arrow). Thus, the experimenter neuron turns into a duration detector. Curves show averages across 50 simulation runs. **(E)** The dynamics of the regularization penalty λ under the reward prediction error rule (each color is one simulation run, n = 20 simulated birds). **(F)** Learning curves in observers are plausible

both when they learn from the auditory cues of air puffs (full lines) or from experimenter actions (cues affected by 30% random 'label noise', dashed lines). However, training and generalization performances get too close to be realistic when observers learn with L1-regularized learning including 30% label noise (dash-dotted line).

We then modeled experimenters by endowing the learning rule with L1 regularization. L1 regularization implements a conjunctive minimization of summed absolute synaptic weights (Tibshirani 1996). To explain why experimenters would have a non-zero positive value of the regularization penalty, we constructed a simple learning rule that dynamically regulates the regularization parameter λ in proportional to reward prediction error (Methods), which is known to be signaled by a class of dopaminergic neurons in the vertebrate brain (Schultz et al.1997; Hollerman et al. 1998; Gadagkar et al. 2016). In this way, regularization increases when the bird makes mistakes, whereas if the bird reaches a high rate of success, the reward prediction error reaches zero in expectation, which settles the value of λ. The observers' brain would not modulate λ because observers do not directly experience rewards and punishments during the experimenter training phase.

We found that interrupting the training process of the regularized neuron at any time resulted in roughly equal performances on both training and testing stimulus sets, Fig. 5C, similar to experimenters' behavior. However, the excellent generalization performance came at a cost: Because we implemented the L1 regularization as a small reduction of synaptic weights (Ng 2004), the synaptic weights of the experimenter neuron and with it the performance on the training set grew only slowly. The main effect of regularization was to concentrate the final synaptic weights on the duration feature, corresponding with our design of stimulus class, Fig. 5D.

To achieve robust generalization, the value of λ had to grow at a rate slower than the synaptic learning rate of the logistic neuron ($\alpha \ll \eta$). In numerical simulations, we found that the value of λ converged to a positive value, Fig 5 E. Training and generalization performance for the experimenter and observer were similar when λ was pre-fixed (to 0.013) or when it was dynamically altered.

We were unable to assert from our simulations whether the apparent learning cue in observers was the air puff's auditory cue (e.g., the experimenter's actions drew attention to the puff first, followed by a complex form of stimulus enhancement) or the experimenters' escape behaviors (as in action imitation). That is, the simulation results matched well with experimental data both when the learning cues for the observer neuron were the air puff sounds (Fig 5F, solid lines) and when the cues were the modelled escape events (we modeled the escapes as binary random variables with a 30% chance of not representing the true class label, corresponding with the average false positive and false negative rates of EXP of about 30%, Fig. 5F dashed lines).

Most importantly, regularization was necessary to achieve a good match with experimental findings. Namely, when we endowed the observer neuron with the same regularization constant λ as the experimenter neuron, but let the neuron learn not from presence/absence of air puff sounds but from the experimenter's noisy actions (see Methods), then training and generalization curves in the observer neuron were very similar and thus not representative of the data, (Fig. 5F, dot-dashed lines).

Discussion

We found that zebra finches can learn to discriminate auditory stimuli by observing expert discriminators. Experimenter and observers' learning performance was subject to a tradeoff that depended on whether the learning cue was experienced or observed. We inferred this cue

dependence thanks to our experiment design in which the stream of auditory stimuli was identical for experimenters and observers. Therefore, any differences in their abilities to learn and to generalize must have been entirely due to the learning cue, which was an aversive air-puff for experimenters and an observable action for observers. Our findings suggest that an experienced cue favors robust generalization, whereas an observed cue favors rapid learning.

Part of our findings are in line with social learning theories which suggest that to learn from others is a successful strategy with high payoff under a wide range of conditions (Axelrod and Hamilton 1981; Rendell et al. 2010). However, our findings also suggest a limitation to the ubiquitous success of social learning strategies. Namely, we find that social learning can lack robustness when environmental conditions even slightly change. In a sense, our findings evidence a sensory analogue to the common view that the best means to learn a (motor) skill is rigorous practice. As in the case of children who perform poorly in exams after neglecting their homework, insights gained through observation seem not to transfer well to new task instances. Currently, there is no reason to think that all forms of observation learning will be subject to lack of robustness. But our work raises the question as to whether there exist some forms of observation learning that promote robust transfer to new task instances.

Our work raises many interesting questions on the behavioral and neurobiological mechanisms used by observers to acquire stimulus-discriminative information. Behaviorally, observers could learn through social mechanisms of action imitation, of observational conditioning, and of stimulus enhancement, or a combination of these. Note that the definitions of these mechanisms are not strict enough to allow a discrete categorization of social learning in any one study (Hoppitt and Laland 2013). Our findings de-emphasize some known social learning mechanisms such as perceptual learning (evidenced by PL learners) and simple stimulus enhancement (evidenced by lack of discriminative behavior during pre-testing). Our experiments also de-emphasize vocal communication as a mechanism but reveal the importance of vision (-

TCOM learners). Overall, the importance of a demonstrating expert suggests that experimenters signal statistical differences between puffed and unpuffed stimuli via their perching behavior such as their rates of leaving the perch. Possibly, observers focused their attention more on the actions of experimenters rather than the stimuli that elicited the actions, which is why they apparently failed to identify the simplest environmental signal that can explain experimenters' behavior, which in our case was syllable duration.

Similar speed-robustness learning tradeoffs as we find exist in rapidly evolving artificial systems, in which high discrimination performance tends to be associated with slow learning as an unwanted side effect (Dauphin et al. 2014). The tradeoff we find between robustness in one learning paradigm and speed in another is most closely paralleled by regularization methods that control inference in artificial classifiers. Excellent generalization of experimenters agrees with strongly regularized classifiers whereas fast learning in observers agrees with weakly regularized classifiers. Our work suggests that the benefits of regularization may be inherent to experimenting but not to observing. Furthermore, subtractive weight depression through heterosynaptic competition has been observed in the amygdala (Royer and Paré 2003), which provides biological plausibility to our notion of L1 regularization. We hypothesize that such a form of weight subtraction is also seen in zebra finches when they are experimenting, but not when they are observing.

A common problem in machine learning is to set the degree of L1 penalty defined by the regularization parameter λ. This parameter is most often selected using grid search or random search methods, to localize the value that minimizes a cross-validation or held-out validation set error (Bergstra and Yoshua Bengio 2012). More sophisticated techniques, such as estimating a Gaussian process regression model between the hyperparameter (such as λ) and the validation error have recently been developed (Snoek et al. 2012). However, all these techniques require an evaluation of the validation error for optimization, for which there is currently no support in

animals and their brains. Therefore, it is far from clear how a brain could implement dynamic regularization. Our speculative proposal is that the balance between learning and regularizing is controlled by a neuromodulatory signal. Such signals are ubiquitous in the animal kingdom and are well suited to convey the amount of regularization, given that they respond sensitively to external reinforcements and their prediction errors (Pawlak and Kerr 2008; Hangya et al. 2015; Yu and Dayan 2005; Iglesias et al. 2013; Wolfram Schultz 1998). One possibility is that air-puff reinforcers drive changes in regularization via experimenters' escape actions, which is supported by the representation of action-specific reward values in brain areas innervated by neuromdulatory neurons (Samejima et al. 2005). This proposal delineates a possible neural system for comparative studies of learning from experience and from observation. It has been shown that reward prediction error and reinforcement learning algorithms in general, may be utilized by humans in order to understand the social value of others' behavior (Behrens et al. 2008; Joiner et al. 2017), to feel vicarious rewards from their success or failure (Mobbs et al. 2009) or from their approval (Izuma et al. 2008). We believe that the computational role of reward prediction error can be extended to that of regularization of learning, mediated by neuromodulator systems such as acetylcholine.

The speculative implications of our simulations are that a prerequisite for the evolution of observation learning was a sufficiently large brain capacity that provided rich sensory representations and put few constraints on usable neural resources for sensory processing. Evolution might have chosen traits in observers that are complementary to those associated with experimenting, explaining the apparent differences in what these learning strategies extract from the sensory environment.

Materials and Methods

Experimental animals

We used adult (older than 90 days post hatch, dph) female zebra finches (*Taeniopygia guttata,* N = 46 females) raised in our colony. Reasons for choosing females are outlined in Supplemental Information. All experiments were licensed by the Veterinary Office of the Kanton of Zurich.

Experimental setup

We adapted an operant conditioning paradigm using social reinforcement (Tokarev and Tchernichovski 2014; Canopoli, Herbst, and Hahnloser 2014). An experimenter and observer pair were placed adjacent to each other in separate cages. The birds could interact from a restricted window in one corner of the cage, forcing them to sit on their respective perches, Fig 1A. The experimenter perch had a sensor to detect presence or absence and trigger stimulus playback through a speaker. We used strong air-puffs directed toward the experimenter as an aversive reinforcement agent. The puffs motivated the experimenter to escape from its perch during 'puffed' class trials. Details of housing, perch and nutritional requirements are detailed in the Supplemental Information.

Stimuli

We created a set of 10 stimuli from the songs of an adult male zebra finch (o7r14) from our colony, Fig 1C. We computed syllable durations via thresholding of sound amplitude traces. Each stimulus S_i in this set (i=1,2, …,10) was made of a string of six syllable renditions, wherein each rendition was longer than the six renditions in stimulus S_{i-1}, Fig. 1E. Based on the ten

stimuli we defined two stimulus classes: the class 'short' was formed by stimuli S_1 to S_5, and the class 'long' was formed by stimuli S_6 to S_{10}. We use the terms 'puffed' and 'unpuffed' as class labels, irrespective of whether short or long stimuli were reinforced. We refer to the stimulus set $\{S_1,\ldots, S_{10}\}$ as the *training set*. To create a *generalization set* we formed another set of 10 stimuli $\{S'_1,\ldots,S'_{10}\}$ from renditions of the same syllable recorded on the very next day, Fig. 1E.

Bird groups and experiment hypothesis:

We used six different groups of experimenters and observers, as follows:

1. Experimenters (**EXP**, n=10 birds): These birds were trained to escape from the perch prior to arrival of air-puffs. The birds first underwent a pre-training phase in which they were accustomed to the setup, followed by a training phase (see *Procedure* in the Supplementary Information). Three out of nine experimenters were also tested on a generalization set of stimuli (Generalization phase) once the training phase was completed, Fig. 2A. Each phase ended when the bird's performance reached a set criterion (see *Performance measures and Statistical Criterion*).

2. Observers (**OBS**, n = 9 birds): Observers were subjected to three phases: an observation phase in which they observed the entire pre-training and training phases of an experimenter, a pre-testing phase (identical to the experimenter's pre-training phase), and a testing phase (identical to the experimenter's training phase), Fig. 2C.

3. Generalizing experimenters (**GENEXP**, n = 9 birds): these birds were tested on the generalization set of stimuli after they had finished the pre-training and training phases on the training set, Fig. 3A. Note: During the training phase, the experimental group GENEXP

is a biological replicate of the EXP group. As expected, there was no difference between EXP and GENEXP in learning time or discrimination accuracy on the training set (Trials to criterion: EXP = 4.8 ± 2.9*10^3, GENEXP = 6.5 ± 4.7*10^3, p = 0.45, test statistic = 90.5, Wilcoxon rank sum test; dPesc at criterion: EXP = 0.36 ± 0.06, GENEXP = 0.37 ± 0.09, p = 0.37, test statistic = 88.5, Wilcoxon rank sum test).

4. Generalizing observers (**GENOBS**, n = 9 birds): These birds underwent the same observation phase and pre-testing phase as OBS. Thereafter, during the testing phase, GENOBS were tested on the full generalization set, Fig. 3B.

5. Perceptual learners (**PL**, n = 7 birds): First, PL could watch an experimenter trigger several thousand stimuli and air-puffs. However, in their case the air-puffs were directed away from the experimenter (oriented downwards outside the cage, Fig 4 A) so PL never experienced or saw the effect of an air-puff against a bird prior to entering the pre-testing phase. Following this, PL then underwent the same pre-test and test phases as OBS.

6. Valence learners (**VL**, n = 5 birds): To test for sensitivity on demonstrator performance, we allowed n=5 VL birds to observe naive experimenters prior to their pre-testing phase. In addition, 3 of those VL were given several hundred stimulus-puff pairings (same protocol as training phase in EXP) prior to observing a naive experimenter. After completion of the pre-testing phase, VL were subjected to the testing phase, Fig. 4B.
7. No Trial Communication learners (-TCOM, n = 5 birds): To test against learning in observers from vocal cues (or the lack thereof, Fig 4 B, C) we separated five observers from their experimenters (pre-training and training phases) into an adjacent, acoustically isolated box. These observers (-TCOM) could view the EXP bird through a window and communicate

vocally through a custom (software controlled) communication channel (see Supplementary Methods) except during trial periods defined from stimulus onset to air-puff offset. During trials periods, the observers could only hear the stimuli and the sounds of air-puffs, but no sounds triggered by the experimenter. After completion of the EXP training phase, -TCOM observers were subjected to the pre-testing and testing phases as for OBS.

Performance measures and Statistical Criterion

For each bird, we partitioned the trials into non-overlapping bins of 100 trials. In each bin we computed the True Positive Rate (P_T) as the probability of escaping on puffed trials and the False Positive Rate (P_F) as the probability of escape during unpuffed trials. Our single measure of performance in each bin is the difference in escape probabilities $dPesc = P_T - P_F$. Within a bin, to decide whether a bird escaped significantly more on puffed trials than on unpuffed trials, we performed a z-test of independent proportions of the following null hypothesis H_0 and alternative hypothesis H_a:

$$H_0 : P_T = P_F, \quad H_a : P_T \neq P_F.$$

For the z-test of independent proportions, we computed in each bin the z-test statistics as follows (applying Yates' continuity correction):

$$z_{stat} = \frac{|P_T - P_F| - (\frac{n_T + n_F}{2 n_T n_F})}{\sigma_\epsilon},$$

$$\sigma_\epsilon = \sqrt{\hat{p}\hat{q}(\frac{n_T + n_F}{n_T n_F})}$$

$$\hat{p} = \frac{(n_T P_T + n_F P_F)}{n_T + n_F}, \hat{q} = 1 - \hat{p}$$

where *n* is the number of puffed trials in that bin. The p-value $\Pr\left[z > z_{stat}\right]$ was computed with the *normcdf* function in MATLAB (Mathworks Inc); a bin was "statistically significant" if the p-value in that bin was smaller than 0.01 (two-sided test).

Criterion and Trials to Criterion

We used two statistical measures to analyze performance: the first measure was used during the experiment to switch the experimental phase of the birds, the second was used to estimate, on a much finer scale, when a bird achieved high and stable discrimination accuracy.

Our criterion for determining when the pre-training/training phases of an EXP bird ends (pre-testing/testing for OBS) was the following: we performed the z-test based significance test (with significance at $p < 0.01$) on $dPesc$ over an entire day. If daily $dPesc$ was significantly greater than zero for two consecutive days, we switched the phase. This "coarse" criterion allowed us to check performance in a logistically tractable manner and provided high power to the test because the sample size was large (average daily number of trials for n=10 EXP: 722.14 ± 317.5).

For analyzing the data post-hoc, we used the criterion mentioned in the article: z-tests in 8 (x100 trial) bin windows and checking for 7/8 bins significant. This latter criterion was used because we wanted to analyze the data on a finer temporal scale and still make sure that the performance was stable. Accordingly, we computed the fraction of 100 trial bins with significant $dPesc$ in a sliding window of 8 bins, Fig. S1D. When this fraction crossed 90% (=0.875), we took the last bin in the window as the bin at which the performance criterion was reached ("criterion bin"). 'Trials to criterion' is then simply the number of all trials performed by the bird up to and including the criterion bin. Our conclusions of fast learning and poor generalization in observers were robust to changes in the definition of the learning criterion: For example, results were

unchanged when we changed the criterion from 7/8 significant bins to 3/4 bins, or when we computed the criterion in 200-trial bins instead of 100-trial bins (supplementary methods, robustness of statistics).

Group level statistical Tests

No explicit power analysis was used for this study. Our main experimental groups (EXP vs OBS in the article) have a sample size of n = 9 (one extra EXP bird was included because its observer partner was not tested). We believed this to be an appropriate sample size considering the statistical test we planned on using (non-parametric Wilcoxon test, which is conservative but has higher power for low sample sizes than a parametric test) and the time it took for preliminary experiments to finish.

To compare birds between two groups, we used Wilcoxon rank sum tests (wilcox.test() in R), either one tailed (when there was a concrete alternative hypothesis, e.g trials to criterion in OBS vs EXP, alternative: EXP > OBS) or two tailed (when there was no concrete alternative hypothesis, e.g. trials to criterion in GENOBS vs GENEXP). We first checked (for all bird groups) whether the trials to criterion were significantly non-Gaussian using the Shapiro Wilk test of normality (shapiro.test in R). Because only the EXP and VL trials to criterion were sufficiently Gaussian, we chose to perform non-parametric Wilcoxon tests instead of t-tests. All group level statistical tests and effect size calculations were performed using the R package (R Studio, https://www.R-project.org/).

Logistic regression with L1 regularization

We modeled experimenter and observer behaviors using logistic regression, which is a simple machine learning classifier that learns linear decision boundaries. In this model, the bird's

behavior (leave or stay on the perch) is computed from the input to the logistic neuron, formed by 21 syllable features provided by Sound Analysis Pro (SAP) (Tchernichovski et al. 2000), a popular software tool for characterizing birdsong and its development (SAP features include mean Wiener Entropy, mean pitch goodness, mean frequency modulation, pitch variance, etc., where mean and variance are computed across syllable duration). Syllable duration formed the 21st feature. Feature 22 was formed by a vector of 1's, endowing the logistic neuron with a bias term. Features 23 and beyond were formed by frozen noise that was randomly drawn from a Gaussian distribution and held fixed for a given syllable. In combination, the total dimensionality of sound-feature vectors $\mathbf{x}$ was n=22+nr, where nr is the dimensionality of frozen noise. The auditory input $\mathbf{z}_i$ to the logistic neuron associated with syllable rendition i presented during trial t was the z-transformed feature vector $\mathbf{z}_i = \frac{\mathbf{x}_i - \mathbf{m}_{t-1}}{\sqrt{\mathbf{v}_{t-1}}}$, where $\mathbf{m}_t = (1-\varepsilon)\mathbf{m}_{t-1} + \varepsilon \left\langle \mathbf{x}_i \right\rangle_{i=1\ldots6}$ is the running mean feature vector and $\mathbf{v}_t \to (1-\varepsilon)\mathbf{v}_{t-1} + \varepsilon \left\langle (\mathbf{x}_i - \mathbf{m}_t)^2 \right\rangle_{i=1\ldots6}$ is the running variance vector. Both the running vectors were updated after each trial (here ε is a small integration rate constant and $\langle . \rangle$ denotes averaging over the six syllables in a given trial).

The partial output of the logistic neuron in response to syllable i signals the probability $f(\mathbf{z}_i)$ of an imminent air-puff, given by $f(\mathbf{z}_i) = \frac{1}{1+\exp(-\mathbf{w}\mathbf{z}_i)}$, where $\mathbf{w}$ is the synaptic weight vector that forms a scalar product with the auditory input. The bird decides to leave the perch (or to not return to it) if $\left(\sum_{i=1}^{6} f(\mathbf{z}_i) \right) > 3$ (majority vote).

The probability that all six syllables correctly (and independently) predict arrival $(u=1)$ or absence $(u=0)$ of an air puff (under a Binomial model) is given by

$P_{\text{correct}} = \prod_i f(\mathbf{z}_i)^u (1 - f(\mathbf{z}_i))^{1-u}$. We train the synaptic weights by maximizing $\log P_{\text{correct}}$ using gradient ascent (maximum likelihood), $\Delta\mathbf{w} = \eta \nabla_{\mathbf{w}} \left(\log P_{\text{correct}} \right)$, where η is a small learning rate. Replacing the definition of $f(\mathbf{z}_i)$ into this expression, we find for observers the simple perceptron-like learning rule that enforces after each trial the weight update $\Delta\mathbf{w}_{\text{OBS}} = \eta \sum_i (u - f(\mathbf{z}_i))\mathbf{z}_i$. In simulations, we randomly picked a stimulus at each trial followed by the synaptic weight change. The only two parameters in this model are the integration rate ε and the learning rate η.

To model experimenters, we applied an additional constant weight subtraction $\Delta\mathbf{w}_{\text{EXP}} = \Delta\mathbf{w}_{\text{OBS}} - \lambda$ on successful trials (leave if puffed and stay if non-puffed), provided the individual synaptic weight was of sufficient magnitude, $|\mathrm{w}| > \lambda$ (to prevent small synaptic weights from changing sign).

We dynamically regulated λ in the following manner:

$$\lambda_t = \max[0, \lambda_{t-1} + \alpha(r_t - \bar{r}_{t-1})],$$

where the reward signal r_t was given by $r_t = \begin{cases} +1 & \text{if the bird's decision was correct in trial t} \\ -1 & \text{otherwise} \end{cases}$

and where $\bar{r}_t = \gamma r_t + (1 - \gamma)\bar{r}_{t-1}$ is a running average estimate of past rewards obtained by the bird. Decisions are correct when birds leave the perch on puffed trials and stay on the perch on unpuffed trials. We set the learning rate $\alpha = 0.00025$, $\gamma = 0.99$ and the initial value $\lambda_{t=0} = 0.0005$.

For the experimenter, the learning cue *u* is the occurrence (*u* = 1) or absence of an air-puff (*u* = 0). For the observer, we hypothesized that u could be either a) the air-puff cue (*u* = 1 during air-puffs and u=0 otherwise) to which the observer's attention is drawn through the experimenter's

behavior (Fig 5 B and Fig 5F solid lines), or b) the action of the experimenter (u = 1 during escapes and u = 0 otherwise). In this latter scenario, the observer is provided a noisy supervisory signal due to false positive and false negative decisions of the experimenter (average EXP false positive rate ~ 30%, average false negative rate ~ 35%, n= 10 EXP birds). To test hypothesis b, we simulated an observer neuron that on randomly chosen 30% of learning (Training set) trials was driven by erroneous learning cues (i.e. escapes on unpuffed trials and no-escapes on puffed trials), with and without regularization (Fig 5F dashed and dot-dashed curves, respectively).

Acknowledgements

We thank Rodney Douglas, Fatih Yanik, Andreas Nieder, and Klaas-Enno Stephan for help with the manuscript. We also acknowledge help with experiments from Heiko Hörster and Aleksander Jovalekic. This work was supported by the Swiss National Science Foundation (grant 31003A_127024) and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 / ERC Grant AdG 268911) to R.H.R

Author contributions

G.N. designed the experiment, carried out the experiment, analyzed the data and wrote the manuscript. J.H contributed to experiment by developing software. R.H designed the experiment and wrote the manuscript.

Competing interests

The authors declare no financial or non-financial competing interests.

References

Axelrod, Robert, and William D Hamilton. 1981. “The Evolution of Cooperation.” *Science (New York, N.Y.)* 211 (4489): 1390–96.

Bass, M. J., and C. L. Hull. 1934. “The Irradiation of a Tactile Conditioned Reflex in Man.” *Journal of Comparative Psychology* 17 (1): 47–65. doi:10.1037/h0074699.

Behrens, Timothy E J, Laurence T Hunt, Mark W Woolrich, and Matthew F S Rushworth. 2008. “Associative Learning of Social Value.” *Nature* 456 (7219): 245–49. doi:10.1038/nature07538.

Bergstra, James, and Yoshua Bengio. 2012. “Random Search for Hyper-Parameter Optimization.” *Journal of Machine Learning Research* 13: 281–305. doi:10.1162/153244303322533223.

Bitterman, M E, R Menzel, Andrea Fietz, and Sabine Schäfer. 1983. “Classical Conditioning of Proboscis Extension in Honeybees (Apis Mellifera).” *Journal of Comparative Psychology* 97 (2): 107–19. doi:10.1037/0735-7036.97.2.107.

Byrne, R W. 2003. “Imitation as Behaviour Parsing.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 358 (1431): 529–36. doi:10.1098/rstb.2002.1219.

Canopoli, A., J. Herbst, and R.H.R. Hahnloser. 2014. “A Higher Sensory Brain Region Is Involved in Reversing Reinforcement-Induced Vocal Changes in a Songbird.” *Journal of Neuroscience* 34 (20): 7018–26.

Cherkin, Arthur. 1969. “Kinetics of Memory Consolidation: Role of Amnesic Treatment Parameters.” *Proceedings of the National Academy of Sciences of the United States of America* 63 (4): 1094–1101. doi:10.1073/pnas.63.4.1094.

Cortes, C., and V. Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20 (3): 273–97.

doi:10.1007/BF00994018.

Dauphin, Yann, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio. 2014. “Identifying and Attacking the Saddle Point Problem in High-Dimensional Non-Convex Optimization.” *arXiv*, 1–14. http://arxiv.org/abs/1406.2572.

Derégnaucourt, Sébastien, Colline Poirier, Anne Van Der Kant, and Annemie Van Der Linden. 2013. “Comparisons of Different Methods to Train a Young Zebra Finch (Taeniopygia Guttata) to Learn a Song.” *Journal of Physiology Paris* 107: 210–18. doi:10.1016/j.jphysparis.2012.08.003.

Gadagkar, Vikram, Pavel A. Puzerey, Ruidong Chen, Eliza Baird-Daniel, Alexander R. Farhang, and Jesse H. Goldberg. 2016. “Dopamine Neurons Encode Performance Error in Singing Birds.” *Science (New York, N.Y.)* 354 (6317): 1278–82. doi:10.1126/science.aah6837.

Galef Jr., Bennett G. 1988. “Imitation in Animals: History, Definition, and Interpretation of Data From the Psychological Laboratory.” In *Social Learning: Psychological and Biological Perspectives*, 3–28.

Geman, Stuart, Elie Bienenstock, and René Doursat. 1992. “Neural Networks and the Bias/Variance Dilemma.” *Neural Computation*. doi:10.1162/neco.1992.4.1.1.

Hangya, Balázs, Sachin P. Ranade, Maja Lorenc, and Adam Kepecs. 2015. “Central Cholinergic Neurons Are Rapidly Recruited by Reinforcement Feedback.” *Cell* 162 (5): 1155–68. doi:10.1016/j.cell.2015.07.057.

Hollerman, J R, L Tremblay, and W Schultz. 1998. “Influence of Reward Expectation on Behavior-Related Neuronal Activity in Primate Striatum.” *J Neurophysiol* 80 (2): 947–63. doi:10.1016/S0531-5131(03)00188-2.

Hoppitt, William, and Kevin N. Laland. 2013. *Social Learning: An Introduction to Mechanisms,*

Methods, and Models. *Princeton University Press*. http://www.scopus.com/inward/record.url?eid=2-s2.0-84883986293&partnerID=tZOtx3y1.

Iglesias, Sandra, Christoph Mathys, Kay H. Brodersen, Lars Kasper, Marco Piccirelli, Hanneke E M denOuden, and Klaas E. Stephan. 2013. "Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning." *Neuron* 80 (2). Elsevier Inc.: 519–30. doi:10.1016/j.neuron.2013.09.009.

Izuma, Keise, Daisuke N. Saito, and Norihiro Sadato. 2008. "Processing of Social and Monetary Rewards in the Human Striatum." *Neuron* 58 (2): 284–94. doi:10.1016/j.neuron.2008.03.020.

Joiner, Jessica, Matthew Piva, Courtney Turrin, and Steve W. C. Chang. 2017. "Social Learning through Prediction Error in the Brain." *Npj Science of Learning* 2 (1). Springer US: 8. doi:10.1038/s41539-017-0009-2.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. doi:10.1038/nature14539.

Markman, Ellen M., and Jean E. Hutchinson. 1984. "Children's Sensitivity to Constraints on Word Meaning: Taxonomic versus Thematic Relations." *Cognitive Psychology* 16 (1): 1–27. doi:10.1016/0010-0285(84)90002-1.

Mobbs, D., R. Yu, M. Meyer, L. Passamonti, B. Seymour, A. J. Calder, S. Schweizer, C. D. Frith, and T. Dalgleish. 2009. "A Key Role for Similarity in Vicarious Reward." *Science* 324 (5929): 900–900. doi:10.1126/science.1170539.

Ng, Andrew Y. 2004. "Feature Selection, L 1 vs. L 2 Regularization, and Rotational Invariance BT - Proceedings of the Twenty-First International Conference on Machine Learning," 379–87.

Okanoya, K, and R J Dooling. 1990. “Temporal Integration in Zebra Finches (Poephila Guttata).” *J Acoust Soc Am* 87 (6): 2782–84. doi:10.1121/1.399069.

Pavlov, I. P. 1927. *Conditioned Reflexes*. *Oxford University Press*. Vol. 17. doi:10.2307/1134737.

Pawlak, V., and J. N. D. Kerr. 2008. “Dopamine Receptor Activation Is Required for Corticostriatal Spike-Timing-Dependent Plasticity.” *Journal of Neuroscience* 28 (10): 2435–46. doi:10.1523/JNEUROSCI.4402-07.2008.

Rendell, Luke E., Robert Boyd, D Cownden, M Enquist, K Eriksson, M W Feldman, L Fogarty, S Ghirlanda, T Lillicrap, and Kevin N. Laland. 2010. “Why Copy Others? Insights from the Social Learning Strategies Tournament.” *Science (New York, N.Y.)* 328 (5975): 208–13. doi:10.1126/science.1184719.

Royer, Sébastien, and Denis Paré. 2003. “Conservation of Total Syanptic Wight through Balanced Synaptic Depression and Potentiation.” *Nature* 422 (April): 518–22. doi:10.1038/nature01532.1.

Samejima, K, K Doya, Yasumasa Ueda, and Minoru Kimura. 2005. “Representation of Action-Specific Reward Values in the Striatum.” *Science (New York, N.Y.)* 310 (5752): 1337–40. doi:10.1126/science.1115270.

Schultz, W, P Dayan, and P R Montague. 1997. “A Neural Substrate of Prediction and Reward.” *Science* 275 (June 1994): 1593–99. doi:10.1126/science.275.5306.1593.

Schultz, Wolfram. 1998. “Predictive Reward Signal of Dopamine Neurons.” *Journal of Neurophysiology* 80 (1): 1–27.

Skinner, B.F. 1953. “Science And Human Behavior,” 461. doi:10.1037/h0052427.

Snoek, Jasper, Hugo Larochelle, and Rp Adams. 2012. “Practical Bayesian Optimization of

Machine Learning Algorithms." *Nips*, 1–9. doi:2012arXiv1206.2944S.

Spierings, Michelle J, and Carel Ten Cate. 2016. "Budgerigars and Zebra Finches Differ in How They Generalize in an Artificial Grammar Learning Experiment." *Proc Natl Acad Sci U S A* 113 (27): 3977–84. doi:10.1073/pnas.1600483113.

Srivastava, N, G Hinton, and A Krizhevsky. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *The Journal of Machine*.

Sturdy, Christopher B, L S Phillmore, J L Price, and R G Weisman. 1999. "Song-Note Discriminations in Zebra Finches (Taeniopygia Guttata): Categories and Pseudocategories." *Journal of Comparative Psychology*. Journal of comparative Psychology. doi:10.1037/0735-7036.113.2.204.

Tchernichovski, O, P P Mitra, T Lints, and F Nottebohm. 2001. "Dynamics of the Vocal Imitation Process: How a Zebra Finch Learns Its Song." *Science (New York, N.Y.)* 291 (5513): 2564–69. doi:10.1126/science.1058522.

Tchernichovski, O, F Nottebohm, Ce Ho, B Pesaran, and Pp Mitra. 2000. "A Procedure for an Automated Measurement of Song Similarity." *Animal Behaviour* 59 (6): 1167–76. doi:10.1006/anbe.1999.1416.

Thorndyke, E. 1905. *The Elements of Psychology: The "law of Effect."* New York, NY, US: A G Seiler. doi:10.1057/9780230203815.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society* 58 (1): 267–88.

Tokarev, K., and O. Tchernichovski. 2014. "A Novel Paradigm for Auditory Discrimination Training with Social Reinforcement in Songbirds." *bioRxiv*. Cold Spring Harbor Labs Journals. http://biorxiv.org/content/early/2014/04/12/004176.abstract.

Woolley, Sarah C, and Allison J Doupe. 2008. “Social Context-Induced Song Variation Affects Female Behavior and Gene Expression.” *PLoS Biology* 6 (3): e62. doi:10.1371/journal.pbio.0060062.

Yu, Angela J., and Peter Dayan. 2005. “Uncertainty, Neuromodulation, and Attention.” *Neuron* 46 (4): 681–92. doi:10.1016/j.neuron.2005.04.026.

Zentall, Thomas R. 2006. “Imitation: Definitions, Evidence, and Mechanisms.” *Animal Cognition* 9 (4): 335–53. doi:10.1007/s10071-006-0039-2.

SUPPLEMENTARY MATERIALS

Supplementary File (pdf) contents: Supplementary information containing details on Materials and Methods as well as supplementary figures S1, S2, S3 and S4 (cited in main text).