Reward Probability and the Variability of Foraging Behavior in Rats

W. David Stahlman and Aaron P. Blaisdell
University of California, Los Angeles, U.S.A.

The connection between reduced reinforcement probability and increased behavioral variability has been well established in recent years (Gharib, Gade, & Roberts, 2004; Stahlman, Roberts, & Blaisdell, 2010). Researchers have hypothesized that this relationship is an adaptive one - it is beneficial for animals to increase behavioral variability in response to low likelihood of success, because this increase in variability potentially allows them to discover new behavioral options that are more highly rewarded. We conducted a study to investigate the relationship between behavioral variability and reward probability in an ecologically valid experimental task. We trained rats to search for hidden food in the presence of either of two landmarks, each that signaled a different likelihood of reward (i.e., HI and LO). Variation in locations searched was higher in the presence of the LO probability landmark. These results build on prior findings that reward expectation drives behavioral variability.

A classic and robust finding in the study of animal learning and behavior is the multifold effect of extinguishing a previously rewarded response. First, a response that is undergoing extinction will generally have a reduced probability of occurring, relative to post-acquisition responding. Second, there will be more variability in the form of the response while it is being extinguished (e.g., Antonitis, 1951; Neuringer, Kornell, & Olufs, 2001) or following a downshift in reward value (Flaherty, Powell, & Hamilton, 1979; Pecoraro, Timberlake, & Tinsley, 1999). While the reduction in previously reinforced responding in the face of extinction has been extremely well documented, only recently have researchers begun to fully examine and unpack the reasons for increased variability of behavior in extinction. Recent research has indicated that expectation of reward is the force that drives the amount of variability in behavior (e.g., Gharib, Derby, & Roberts, 2001; Gharib et al., 2004; Stahlman, Roberts, et al., 2010; Stahlman, Young, & Blaisdell, 2010).

Neuringer, Deiss, and Olson (2000) found that rats that were explicitly reinforced with food for variability in behavior tended to discover more efficient behavioral options (i.e., behaviors that pay off at a greater rate) quicker than control animals that received no explicit reinforcement for variability. This indicates that the modulation of behavioral variability may be important to organisms for the purpose of discovering new rewarded action. In a study in a phylogenetically distinct species, the Bengalese finch, Tumer and Brainard (2007) discovered that the very small amount of variability in "crystallized" bird song is

adaptive, such that the finches can modulate their song to avoid an aversive auditory stimulus. The authors suggest, "In other systems as well, subtle variation in performance may reflect continued experimentation to optimize behavior…" (p. 1244; see also Neuringer, 2004). Indeed, Gharib et al. (2004) proposed that high levels of behavioral variation in low-reinforcement conditions serve to increase the likelihood that an organism discovers and exploits more-highly rewarded behaviors.

Both early (e.g., Antonitis, 1951) and recent (e.g., Stahlman, Roberts, et al., 2010) research in the investigation of behavioral variability has primarily used conventional operant procedures conducted in Skinner boxes. While this apparatus allows for a high level of precision and experimental control, it is not a naturalistic context to explore the relationship between reinforcement probability and behavioral variability. Therefore, it is of interest to determine whether the relationship between reward probability and variability can be demonstrated in a more ecologically valid setting. We were particularly interested in investigating whether the relationship between reward probability and behavioral variability is maintained in a spatial navigation task. Spatial learning and navigation are governed by a number of processes, some of which are unique to spatial behavior (O'Keefe & Nadel, 1978; Poucet & Benhamou, 1997). Establishing the relationship of variability and expectation in a spatial task would be a strong demonstration of the generality of the rule.

Utilizing a within-subjects design, we conducted an experiment with rats to investigate the relationship between behavioral variability and reinforcement likelihood in a spatial task that engages the natural behavior of the rat—foraging for hidden food in an open field. Each of two landmarks signaled both the location and probability (100% vs. 20%) of food reward in an open field. We predicted that the low-reward landmark would instantiate a greater amount of variation in the rats' search behavior. In other words, we expected lower probability of food to shift the balance of search away from exploitation and towards exploration.

Method

Subjects

Ten female Long-Evans rats (*Rattus norvegicus)* obtained from Harlan (Indianapolis, IN) served as subjects. Subjects were pair-housed in translucent plastic tubs with a substrate of wood shavings in a vivarium maintained on a 12-hr dark/12-hr light cycle. All experimental manipulations were conducted during the dark portion of the cycle. A progressive food restriction schedule was imposed over the week prior to the beginning of the experiment, until each rat received approximately 15 g of food each day. All animals were handled daily for approximately 30 s during the five days prior to the initiation of the study.

Apparatus

The open field (see Fig. 1) consisted of a flat panel of wood (1.52 x 1.52 m), painted white and placed on a table 94 cm above the floor. Sixteen Tupperware[TM] cups were arranged in a 4 x 4 grid. The cups measured 6.35 cm in height and 7.62 cm in diameter. Adjacent cups were spaced 32 cm apart (center to center). The cups were recessed into the floor of the open field, with the top lip of each cup extending 2 cm above the surface of the open field. The cups were filled halfway with cedar wood shavings. Cocoa Puffs[TM], as a positive reinforcer, could be placed in the cups either on top of or buried underneath the wood shavings (see procedure; see also Stahlman & Blaisdell, 2009). The

apparatus was located in a classroom with numerous extra-maze cues present. Lab benches flanked the west and east walls of the room. A large whiteboard hung on the south wall, while doors were located in both the northwest corner as well as in the center of the north wall. A 40-watt incandescent vertical lamp in the northeast corner of the room and a small bank of fluorescent lights on the west side of the room provided dim, indirect illumination.

Two distinctly shaped and colored blocks of wood served as landmarks. A red rectangle and an orange triangle served as landmarks A and B, counterbalanced. Experimental landmarks were approximately 10.16 cm x 2.54 cm.
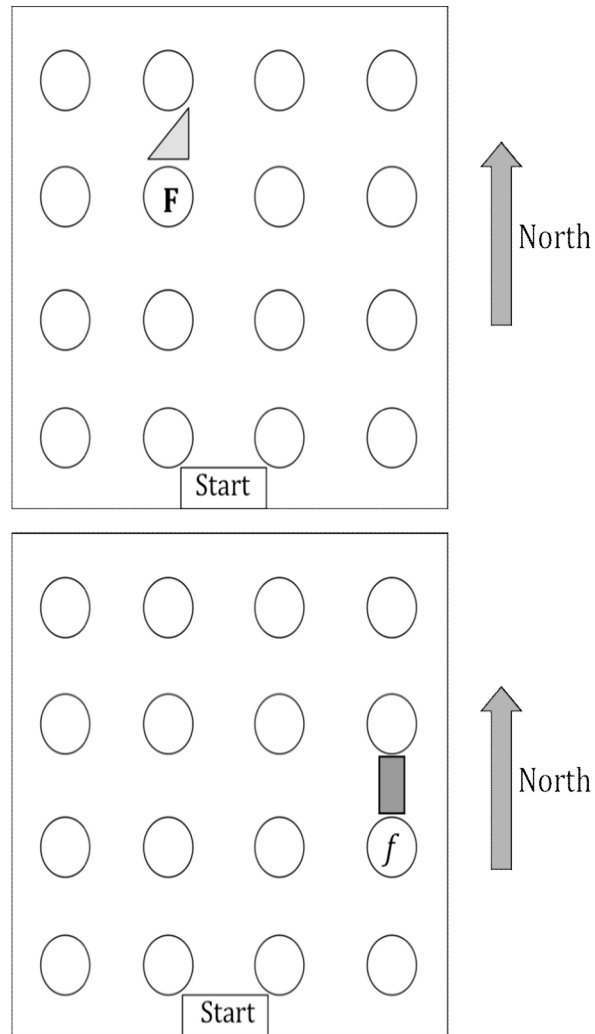
Figure 1. Diagram of the experimental apparatus. Top panel: Example of a HI probability trial (100% probability of reinforcement). The triangle is a landmark indicating both the likelihood of reinforcement and the location of food (F). Bottom panel: Example of a LO probability trial (20% probability of reinforcement). The rectangle signals the probability of reinforcement and possible location of food (*f*).

Procedure

Habituation Training: Sessions 1-2. Approximately 30 pieces of cereal were scattered across the open field prior to each trial. Rats were individually placed at the start location at the

southern end of the open field. They were each allowed to explore the open field for ten minutes before being removed from the apparatus.

Phase 1 Training: Sessions 3-10. Trials in this phase consisted of the placement of one of the 2 wooden landmarks (LMs) at a location on the open field. For two sessions, two pieces of cereal were placed inside the cup to the immediate south of the LM, with two additional pieces placed on the wooden table at locations directly adjacent to the goal cup. On the next two sessions, only two pieces of food were placed in the goal cup, with one being partially buried under the wood substrate (appx. 50% visible). Following this step, the pieces were buried completely under the substrate at a depth of approximately 1 cm. In all cases, the location of the food (and therefore, the LM) was randomized. A trial was terminated when the rat found and consumed all food, or after 3 min, whichever came first. Each session consisted of four trials (two HI trials and two LO trials). The trial order was counterbalanced across session.

Phase 2 Testing: Sessions 11-18. The LMs were placed on the open field in the same fashion as in Phase 1; however, the probability of finding food in the goal cup for the LO landmark was reduced to 20%. The probability of food in the goal cup on HI trials remained at 100%, as in Phase 1. Reinforcement on LO trials was randomly determined with probability of 20%. For the first session of Phase 2, there were four trials (2 of each type). For Sessions 12-18, there were six experimental trials, with 3 trials of each type. Trials were terminated when the rat searched the goal cup or at 3 minutes, whichever came first. As before, the location of the goal, and therefore the LMs, was randomized across trials.

Measures. A "search" was defined as the nose of the rat breaking the top plane of the goal cup. In real time, we recorded the cup location of each search response in sequence. We used a stopwatch to record the interresponse times (IRTs) between cup searches. Using these data, we were able to calculate the temporal distance between each response and a rat's median IRT; this served as our measure of temporal variability. We calculated interresponse distance (IRD), which served as a measure of spatial variability in responding. We analyzed the last eight sessions of data, which comprised the totality of Phase 2.

Results

A dependent-samples *t*-test revealed no effect of reward likelihood on pre-response times, $t(7) = 0.61$, $p = 0.28$, indicating no differences in response rate across the two trial types. Separate dependent-samples *t*-tests also found no effect of reward likelihood on the temporal variability of responding, nor of mean IRD of responses, $p > 0.05$.

Figure 2A is a depiction of the mean number of choices made per trial type across sessions; the rats performed a mean of 2.8 searches on LO trials and a mean of 2.1 searches on HI trials. Though not significant, a dependent samples *t*-test revealed a trend towards more searches occurring on LO trials, $t(7) = 1.79$, $p = 0.06$. On average, rats searched more locations on LO trials on the first seven test sessions (and 7 of 8 total; Fig. 2B). Figure 2B also depicts an apparent trend towards lower variability over test sessions. For each rat, on each trial type, we calculated the slope of the trend line of search frequency across individual test sessions; the slopes on HI vs. LO trials (means = -0.14 and -0.29, respectively) were not significantly different, $t(7) = 1.41$, $p = 0.2$. Since there was no difference in the trend across the trial types over session, we pooled the two trial types and conducted a paired-samples *t*-test on search frequency comparing the first and last test session; the mean number of searches on the first test session was 6.8, while the mean search frequency on the last session was 3.2. This difference was statistically significant, $t(7) = 3.31$, $p < 0.05$, indicating that rats performed fewer searches at the end of Phase 2.

A)

B)
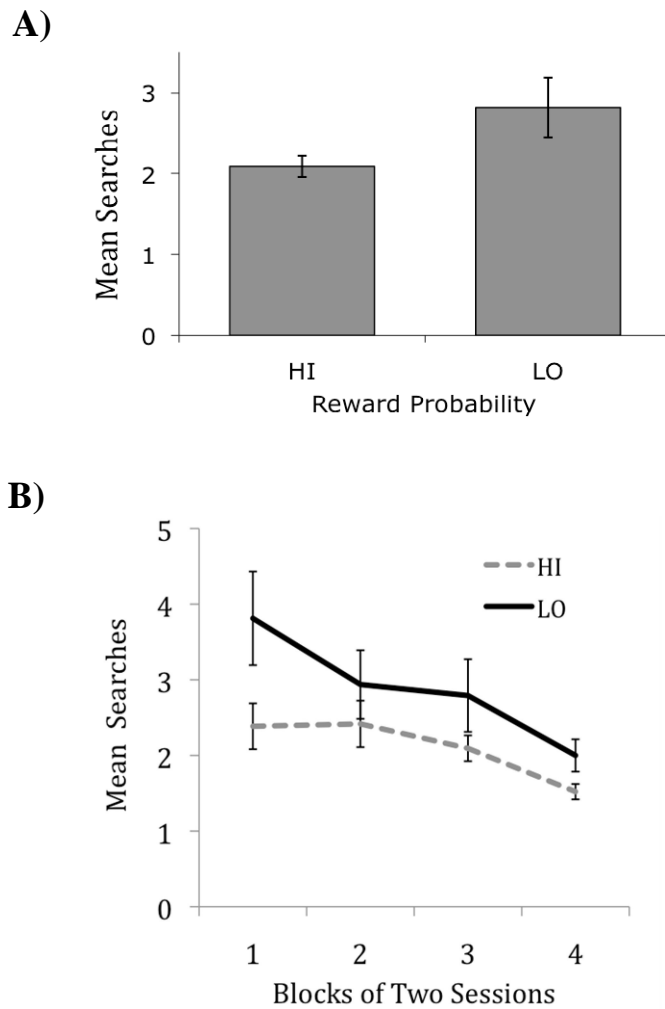
Figure 2. A) Mean number of searches in Phase 2 across the two probabilities of reward. B) Mean number of searches on each trial type across the eight Phase 2 experimental sessions, in two-session blocks.

To examine another measure of spatial variability of responding, we constructed a distribution of within-trial search frequency (Fig. 3). The distributions of searches on both trial types are strongly positively skewed. On HI trials, however, the rats never required more than seven searches to find the goal cup and terminate the trial; on LO trials, the rats occasionally took many more choices (up to 17). In fact, on 13 of the 184 trials (7.1%), the rats searched in more than seven locations prior to the termination of a trial. We calculated each rat's search range (i.e., max searches minus min searches) for each trial type, which serves as a measure of the width of the distribution. This measure was greater on LO trials (mean = 9.25) than on HI trials (mean = 4.5), $t(7) = 2.57$, $p < 0.05$. Thus, a landmark signaling a lower probability of finding food engendered greater skew indicating increased variation in searching. The distribution of search frequencies was significantly wider on LO trials than on HI trials, indicating that behavioral

variation was greater with reduced expectation or reward; this pattern of behavior was true of both the standard deviation (Fig. 4A), $t(7) = 2.68$, $p < 0.05$, and the coefficient of variation of search frequency (Fig. 4B), $t(7) = 3.59$, $p < 0.01$.
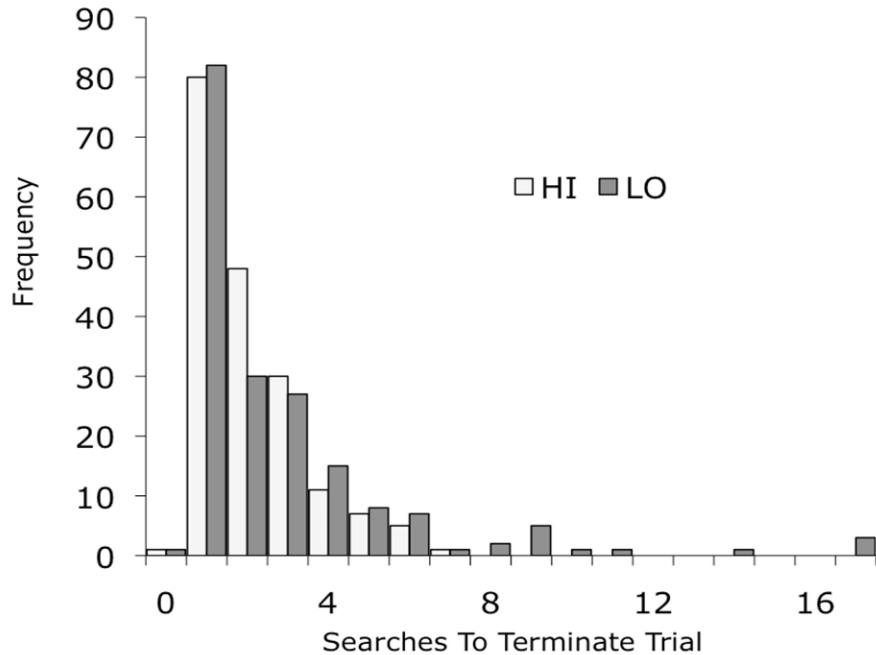

Figure 3. Frequency distribution of number of search locations per trial across the two trial types during Phase 2, across all trials and subjects. "Searches" refers to the total number of goal cups investigated prior to the termination of a trial.

A final noteworthy finding is the rats' probability of actually finding the goal cup during the trial – on HI trials, the rats only failed to search the goal cup on 2 of 184 trials (1.1%); however, on LO trials, the rats failed to search the goal cup on 13 of 184 trials (7.1%).
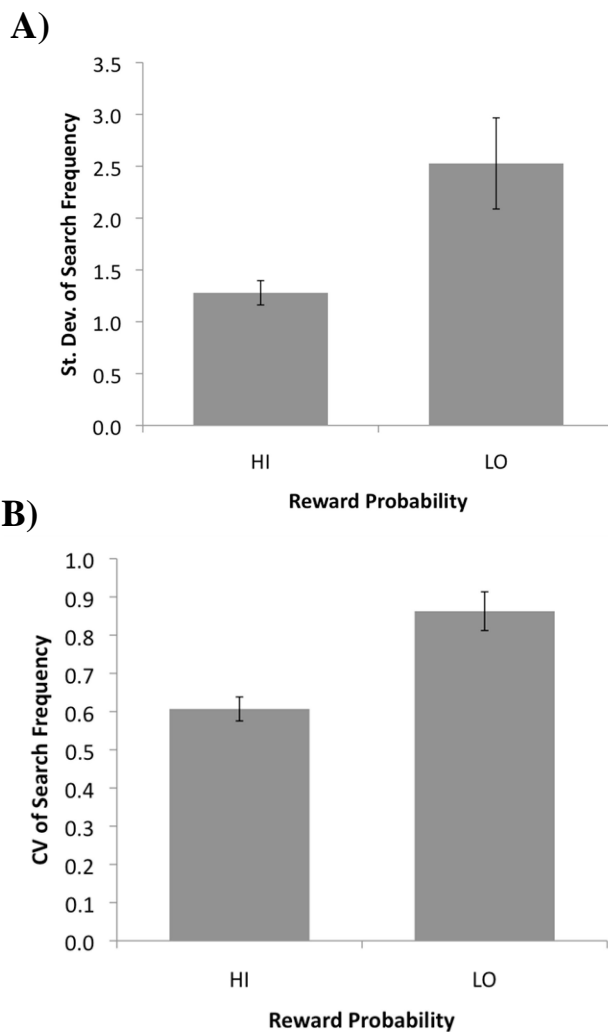
A)

B)

Figure 4. A) Mean standard deviation of rats' search frequency across the two levels of reward probability. B) Mean coefficient of variation (CV) of rats' search frequency across the two levels of reward probability.

Discussion

Foraging variability was higher with lower reward likelihood; rats tended to search in more potential goal locations, and exhibited more variation in the frequency of cups searched on LO trials as compared to HI trials. The effect on the mean number of searches per condition was marginally significant; however, examination of the variance of search frequency demonstrated that the frequency of the rats' search behaviors became more variable with a lower reward probability. Most importantly, as searches were non-normally distributed, a more sensitive measure of variation (i.e., comparison of animals' search ranges in each condition) showed reliably greater variation in search locations on LO than HI trials. The constellation of response measures indicates that spatial variability in foraging is modulated by the expectation of reward; however, there was no effect

of reward probability on the temporal variation of search behavior. This suggests that different measures of behavioral variability are dissociable – an increase in the variability of spatial responding does not necessarily correspond to an increase in temporal variability. This is interesting because a number of prior studies indicate that spatial and temporal variability tend to covary in this regard (e.g., Stahlman, Roberts, et al., 2010; Stahlman, Young, et al., 2010).

This experiment found evidence of the negative relationship between reward expectation and behavioral variability in a more ecologically valid experimental situation—an open field. This study adds to prior research demonstrating similar effects in instrumental nose-poking (Antonitis, 1951), bar-pressing duration (Gharib et al., 2001, 2004), and bar-press sequence variability (Cherot, Jones, & Neuringer, 1996) in rats, as well as instrumental (Eckerman & Lanson, 1969; Stahlman, Roberts, et al., 2010) and Pavlovian (Stahlman, Young, et al., 2010) key- or screen-pecking in pigeons. Our finding that spatial search behavior is more variable on trials with a lower probability of reward supports a behavior systems view of foraging behavior (e.g., Timberlake & Lucas, 1989): animals may engage less in consummatory mode behaviors and more in appetitive mode behaviors when the expectation of reward is low.

Our results are consistent with similar effects produced by a downshift in reward magnitude in foraging tasks with rats. Flaherty et al. (1979) found that rats showed increased levels of open rearing, locomotion, and search behaviors following a shift from 32% to 4% sucrose solution reward in an open field. Utilizing similar reward parameters, Pecoraro et al. (1999) reported an increase in rats' spatial search behavior in an elevated plus maze when reward was downshifted. The authors suggested that downshift in reward magnitude shifted rats from a consummatory mode of action to an appetitive (search-based) mode of action (see also Timberlake & Lucas, 1989).

Both our results and those from incentive-downshift experiments suggest that low-reinforcement conditions increase variability in behavior as an adaptive tool for the optimization of behavior (Gharib et al., 2004). The next step is to show direct evidence for the adaptive role of variability in the discovery of better-rewarded actions.

References

Antonitis, J. J. (1951). Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of Experimental Psychology, 42*, 273-281.

Cherot, C., Jones, A., & Neuringer, A. (1996). Reinforced variability decreases with approach to reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 497-508.

Eckerman, D. A., & Lanson, R. N. (1969). Variability of response location for pigeons responding under continuous reinforcement, intermittent reinforcement, and extinction. *Journal of the Experimental Analysis of Behavior, 12,* 73-80.

Flaherty, C. F., Powell, G., & Hamilton, L. W. (1979). Septal lesion, sex, and incentive shift effects on open field behavior of rats. *Physiology & Behavior, 22,* 903-909.

Gharib, A., Derby, S., & Roberts, S. (2001). Timing and the control of variation. *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 165-178.

Gharib, A., Gade, C., & Roberts, S. (2004). Control of variation by reward probability. *Journal of Experimental Psychology: Animal Behavior Processes, 30*, 271-282.

Neuringer, A. (2004). Reinforced variability in animals and people: Implications for adaptive action. *American Psychologist, 59*, 891-906.

Neuringer, A., Deiss, C., & Olson, G. (2000). Reinforced variability and operant learning. *Journal of Experimental Psychology: Animal Behavior Processes, 26*, 98-111.

Neuringer, A., Kornell, N., & Olufs, M. (2001). Stability and variability in extinction. *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 79-94.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, England: Clarendon Press.

Pecoraro, N. C., Timberlake, W. D., & Tinsley, M. (1999). Incentive downshifts evoke search repertoires in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 153-167.

Poucet, B., & Benhamou, S. (1997). The neuropsychology of spatial cognition in the rat. *Critical Reviews in Neurobiology, 11,* 101-120.

Stahlman, W. D., & Blaisdell, A. P. (2009). Blocking of spatial control by landmarks in rats. *Behavioural Processes, 81,* 114-118.

Stahlman, W. D., Roberts, S., & Blaisdell, A. P. (2010). Effect of reward probability on spatial and temporal variation. *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 77-91.

Stahlman, W. D., Young, M. E., & Blaisdell, A.P. (2010). Response variability in pigeons in a Pavlovian task. *Learning & Behavior, 38,* 111-118.

Timberlake, W. D., & Lucas, G. A. (1989). Behavior systems: An integrative approach. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary theories of learning: Instrumental conditioning theory and the impact of biological constraints on learning* (pp. 237-276). Hillsdale, NJ: Erlbaum

Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of crystallized adult birdsong. *Nature, 450*, 1240-1244.