



**The *Law of Expect* or a *Modified Law of Effect*?  
Outcome expectation and variation in learned behavior<sup>1, 2</sup>**

*Aaron P. Blaisdell*<sup>3</sup>  
Department of Psychology  
UCLA

*Alexandra Stolyarova*  
Department of Psychology  
UCLA

*W. David Stahlman*  
Department of Psychological Science  
University of Mary Washington

**Abstract:**

We review converging lines of evidence describing the negative relationship found between the signaled probability of food and response variation. This relationship is quite general: It is observed in both temporal and spatial behavioral dimensions, in both rats and pigeons, and in both the operant chamber and in open-field settings. Response variation is also greater under conditions involving smaller or delayed food rewards compared to larger or immediate rewards. These data support a new *Law of Expect* which states that outcome expectation is a determinant of response variation, with variation increasing or decreasing with a decrease or increase in outcome expectation, respectively. By comparison, we present simulations of a *Modified Law of Effect* extending the explanatory power of Thorndike's original *Law*. While this *Modified Law of Effect* has improved explanatory power, it fails to account for all of the empirical data. The *Law of Expect* also fails to account for all of the empirical phenomena. Thus, both *Laws* are necessary. Investigations into the neural basis of these laws may lead to new insights into the generation and regulation of creative behavior and to a better understanding of the neural systems involved in learned behavior.

Keywords: *Law of Effect, Law of Expect, Variation, Variability, Expectation, Learning*

**Resumen**

Se revisan líneas convergentes de evidencia y describimos la relación negativa encontrada entre la probabilidad señalizada de alimento y la variabilidad de respuesta. Esta relación es bastante general: se observa en las dimensiones temporal y espacial del comportamiento, tanto en ratas como palomas, y tanto en la cámara operante como en situaciones de campo abierto. La variabilidad de la respuesta también es mayor en condiciones que implican recompensas de comida más pequeñas o demoradas, comparadas con recompensas más grandes o más inmediatas. Estos datos apoyan una nueva Ley de Expectativa que establece que la expectativa de resultados es un determinante de la variabilidad de la respuesta, creciendo o

1 Reference to this article on the web is: <http://conductual.com/content/law-expect-or-modified-law-effect>

<sup>2</sup> Acknowledgments: Portions of this paper were presented at the biennial SINCA, in Mexico City, Mexico, 2015. Simulation work was supported by the UCLA Division of Life Sciences Recruitment and Retention Fund (A. Izquierdo). Requests for reprints should be addressed to Aaron P. Blaisdell, UCLA Department of Psychology, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563, USA; e-mail: [blaisdell@psych.ucla.edu](mailto:blaisdell@psych.ucla.edu).

<sup>3</sup> Corresponding Author: Aaron P. Blaisdell, UCLA Department of Psychology, 1285 Franz Hall, Los Angeles, CA 90095-1563. [Blaisdell@psych.ucla.edu](mailto:Blaisdell@psych.ucla.edu).



decreciendo en relación a una disminución o un aumento en la expectativa de resultados, respectivamente. Presentamos simulaciones de una Ley del Efecto Modificada que extiende el poder explicativo de la Ley original de Thorndike. Aunque que esta Ley del Efecto Modificada posee mayor poder explicativo, no da cuenta de todos los datos empíricos. La Ley de Expectativa tampoco explica todos los fenómenos empíricos. Por esta razón, ambas leyes son necesarias. La investigación sobre las bases neurales de estas leyes puede conducir a nuevas ideas sobre la generación y regulación de la conducta creativa y a una mejor comprensión de los sistemas neurales implicados en el comportamiento aprendido.

Palabras clave: *Ley del Efecto, Ley de la Expectativa, Variación, Variabilidad, Expectativa, Aprendizaje*

Variation is fundamental to Darwin's theory of evolution by natural selection (Darwin, 1859). The process of selection operates on the substrate of heritable phenotypic variation to shape the evolution of traits. As powerful as it is, there are limits to the scope of evolutionary processes. Most notably, evolution does not act on the individual, but only on populations of individuals. Thus, the forces of natural selection do not continue to shape an individual's phenotype.

Skinner famously wrote, "Where inherited behavior leaves off, the inherited modifiability of the process of conditioning takes over" (Skinner, 1953, p. 83). That is, organisms whose behavior is modified through learning processes are not limited to a fixed behavioral phenotype. The analogy between natural selection and learning through reinforcement suggests that processes of variation and selection that are prevalent in evolutionary analyses should be equally prevalent in analyses of learned behavior. Indeed, selection has been of central importance in learning theory. Processes of behavioral variation, however, have received much less attention (Epstein, 1991; Roberts, 2014). This is surprising given the large amount of work on phenotypic variation in evolutionary biology, with a range of mechanisms including genetic mutation and polymorphisms, genetic recombination during sexual reproduction, epigenetics and developmental processes, and other environmental regulatory factors. A comparable set of mechanisms by which behavioral variation is modified have not yet become of central interest in learning theory. Instead, it is often assumed that selection acts on a preexisting amount of behavioral variation without explicitly noting where this variation comes from, nor how it contributes to the shaping of learned behavior. Epstein's ingenious generativity theory has been proposed as a holistic account for the orderliness of behavioral variability (see Epstein, 2014, for a cogent account of the theory). The theory provides a formal account for how novel behavior emerges predictably from conditions in which behavior is ineffective or the conditions themselves are novel. Nevertheless, it is agnostic with respect to (or perhaps eschews accounts in terms of) psychological mechanisms that regulate behavioral variation.

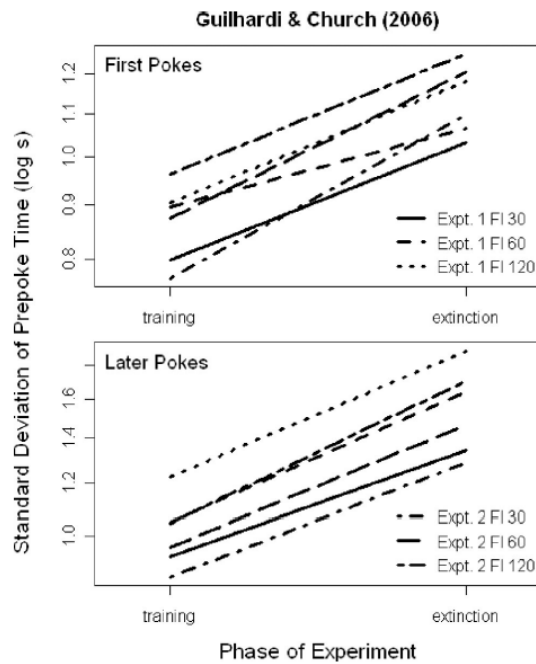
In the past, others and we have made the claim that expectation of the outcome plays a central mechanistic role in the modulation of behavioral variation. In this article, we provide an overview of our research on this topic. We then discuss whether outcome expectation is a necessary theoretical construct to explain these results.

### Empirical studies

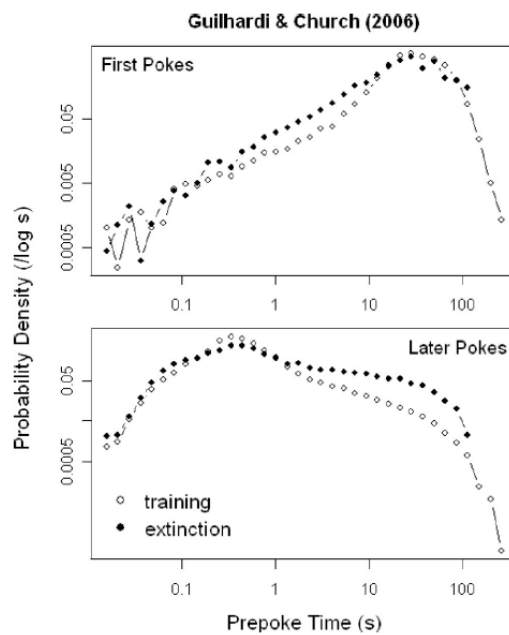
An early clue to the role of expectation in driving behavioral variation comes from the study of extinction of learned behavior. Historically, researchers reported that behavioral variation increases markedly during extinction of operant behavior. When a previously reinforced response (e.g., lever pressing) is no longer followed by reward, the action does not merely become less frequent; behavior also tends to become more variable in nature. This relationship has been well documented (Antonitis, 1951; Balsam, Deich, Ohyama, & Stokes, 1998; Eckerman & Lanson, 1969; Frick & Miller, 1951; Herrick &



Bromberger, 1965; Millenson & Hurwitz, 1961; Neuringer, Kornell, & Olufs, 2001; Notterman, 1959; Stebbins & Lanson, 1962).



**Figure 1.** Standard deviations of prepoke times as a function of peck order and experimental phase. Data from Guilhardi and Church (2006). The legend identifying the various lines is in two parts: half in the upper panel, half in the lower panel. Each point is a mean over 24 rats. Reprinted from Stahlman, Roberts, and Blaisdell (2010a) with permission.

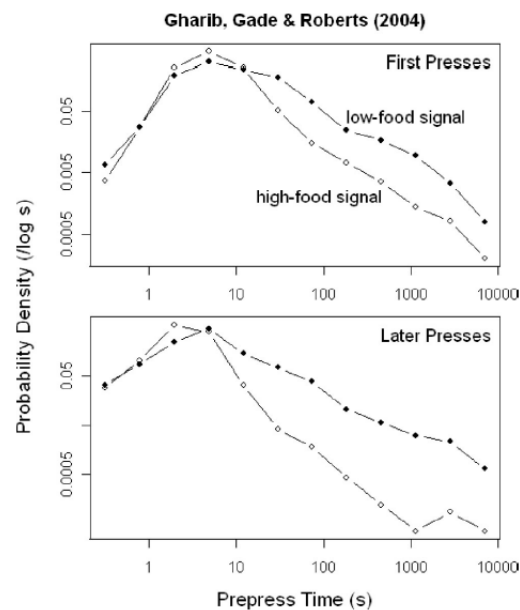


**Figure 2.** The distribution of prepoke times as a function of training condition. Data from Guilhardi and Church (2006). Each point is a mean over 24 rats, three fixed-interval values, and two experiments. Reprinted from Stahlman, Roberts, and Blaisdell (2010a) with permission.



We confirmed this in a reanalysis of the raw nosepoke data from a study by Guilhardi and Church (2006) reporting the extinction of operant behavior in the rat (Stahlman, Roberts, & Blaisdell, 2010a). Figure 1 shows that standard deviations of time to first poke and interpoke interval (prepoke times) for subsequent pokes increased during extinction after initial training on various fixed-interval (FI) schedules of food reinforcement. Importantly, Figure 2 shows that the distribution of prepoke times was not shifted, as predicted by a change in mean response rate, but broadened during extinction. Thus, during extinction when expectation of a food reward diminishes, variation in the timing of the previously-reinforced responses increased.

An interpretation of this relationship in terms of outcome expectation is, however, problematic. Many factors differ between training and extinction, however, confounding an interpretation purely in terms of outcome expectation. Extinction differs in non associative factors. The density of reward is lower during extinction than during acquisition. Response rates tend to be lower (after the extinction burst has subsided) during extinction as well. Associative factors also differ between extinction and established behavior late in acquisition. One of them is expectation of reward, which is lower during extinction than after acquisition. Context conditioning can also differ, being lower during extinction when the association between the context and the outcome is also undergoing extinction. The context may become a modulator of extinction learning, as is demonstrated in studies of renewal (Bouton & King, 1983).



**Figure 3.** Preresponse time as a function of response order. Both axes are logarithmic. Data from Gharib, Gade, and Roberts (2004), Experiment 2, 100%/25% phase. First = data from the first bar presses during a trial. Later = data from all later bar presses. Each point is a mean over 11 rats. Reprinted from Stahlman, Roberts, and Blaisdell (2010a) with permission.

These problems necessitate the adoption of a method other than extinction that avoids these confounding factors if one is to establish a causal link between outcome expectation and behavior variation. Gharib, Gade, and Roberts (2004) developed a steady-state operant procedure with which to cleanly study the role of outcome expectation in behavior variation. Rats were trained to press a bar for a food reward. There were two instrumental cues (discriminative stimuli) used in each session, with 50% of the trials presenting a high-food cue and the remaining trials presenting a low-food cue. On trials with the high-food cue, lever pressing was rewarded on every trial. Lever pressing was only reinforced on 25% of trials with the low-food cue. Figure 3 shows results from this experiment. The cue signaling a lower reward probability produced higher variation in lever-press times. In Gharib et al.'s procedure, trials were

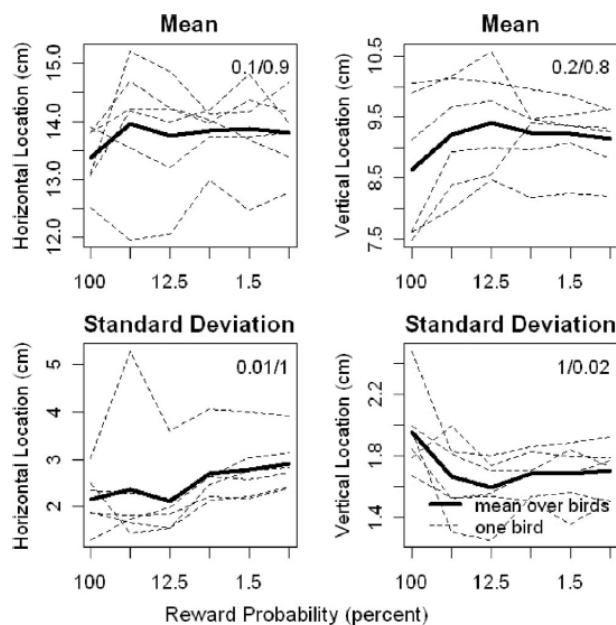


randomly mixed. Thus, most factors were equated for the two types of trials, including overall density of food in each session, time since most recent food, overall session response rate, time since most recent response, and context conditioning. This leaves reward expectancy signaled by each cue due to their respective histories of reinforcement as the most likely factor determining variation in lever press times.

These data suggest a general *Law of Expect*:

*Expectation of reward modulates response variation such that they have a negative relationship.*

**As expectation increases, variation decreases.** This reflects exploitation of a known resource. **As expectation decreases, variation increases.** This reflects exploration for other resources when the known reward is expected to be of low probability. Humans and pigeons have been shown to be sensitive to probabilities of signaled reward as a modulator of the exploration and exploitation decisions (Racey et al., 2011), thus, it is not surprising to find this in rats, too. Bolles (1972) went so far as to propose that psychological processes based on expectancy could completely replace those based on S-R habit as an explanation for the fundamental principles of reinforcement that has received so much attention in the behavior-analytic tradition of Thorndike and Skinner. For example, based on a substantive amount of research on expectancies beginning with the work of Tolman (1932), Bolles suggested that three postulates concerning expectancy could provide a complete account of reinforcement learning: a) the strength of the S-O expectancy (i.e., the Pavlovian association), b) the strength of the R-O expectancy (which underlies goal-directed behavior), and c) the value of S (i.e., incentive motivation). For a more recent review of these topics, see Blaisdell (2008).



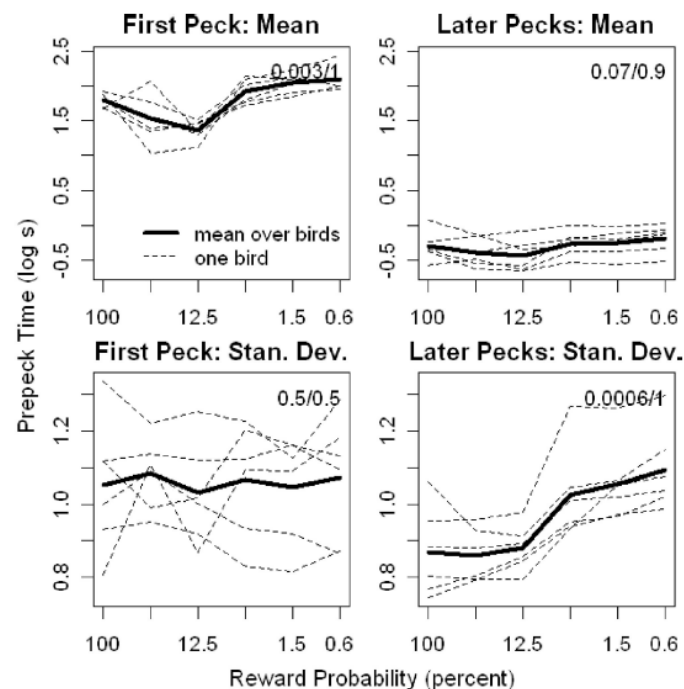
**Figure 4.** Peck location as a function of probability of reward of operant screen pecks. The left panels show horizontal position, the right panels, vertical position, both measured from the lower left-hand corner of the touchscreen. Each dotted line is from a different bird. The solid line shows the mean over birds. The two numbers in the upper right of each graph are p values from one-tailed t tests of the hypothesis that the values are increasing (first number) or decreasing (second number). Reprinted from Stahlman, Roberts, and Blaisdell (2010a) with permission.

If this relationship is general, then evidence should be found in a wide range of learning situations, test settings, response types, and species. We tested the generality of this rule in several ways: in



different species (pigeons and rats), in different dimensions of behavior (temporal and spatial), and in different types of experimental settings (operant chamber and open field).

First, we investigated the rule in operant screen pecking in pigeons (Stahlman et al., 2010a). Pigeons were presented with six types of trials in each session. On each trial, a discriminative stimulus (a colored circle) was presented on the center of the screen. The pigeon had to complete a random-ratio (RR) 5 requirement of pecking to the cue to end the trial. Upon the termination of a trial, the cue disappeared from the screen and the intertrial interval (ITI) began. There were 6 cues, each associated with a different probability of reward (2.8 s access to mixed grain from a hopper below the touchscreen) ranging from 100% to 0.6%. If reward was scheduled, it was delivered immediately upon the termination of a trial.



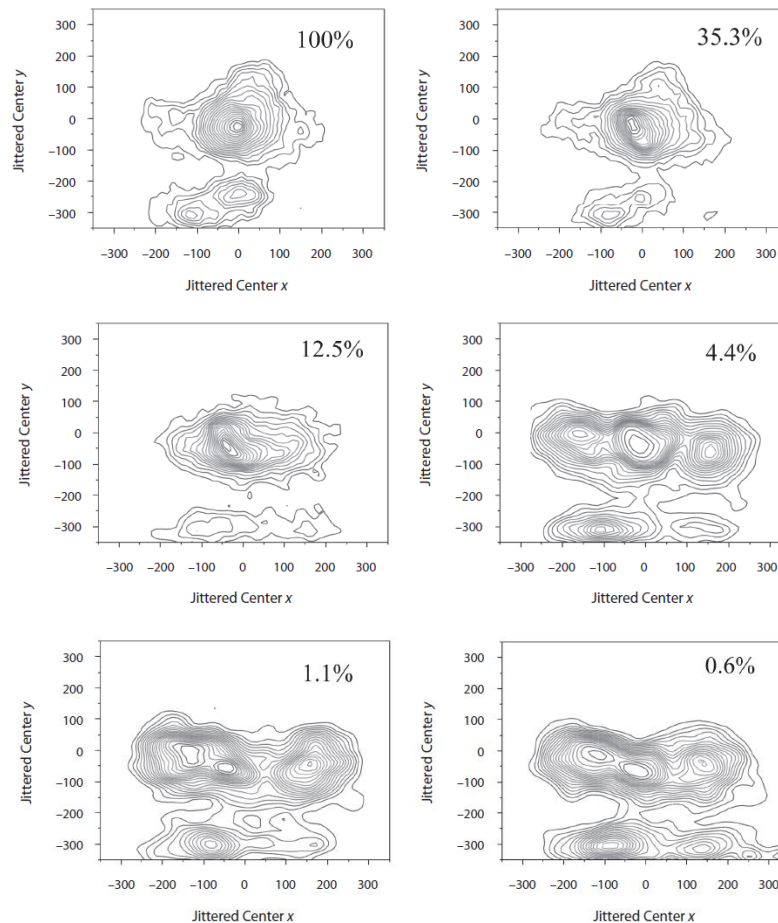
**Figure 5.** Prepeck times as a function of probability of reward of operant screen pecking. First pecks are the first pecks during a trial; later pecks are all later pecks during that trial. The two numbers in the upper right of each graph are p values from one-tailed t tests of the hypothesis that the values are increasing (first number) or decreasing (second number). Each dotted line is from a different bird. The solid line shows the mean over birds. Reprinted from Stahlman, Roberts, and Blaisdell (2010a) with permission.

We were interested in the effects of reward probability signaled by the cue on variation in two dimensions of the behavior: temporal and spatial. To assess temporal variation, we measured log interresponse times (IRT) between pecks. To assess spatial variation, we measured standard deviation of peck locations on the screen. Using this procedure, we have observed that response variation in the spatial (Figure 4, bottom panels) and temporal (Figure 5, bottom panels) dimensions increases as a decreasing function of probability of reward signaled by the cue. This relationship held for both peck location (i.e., spatial domain) and the interpeck interval (temporal domain). As with the results from rats (Gharib et al., 2004), variability was uncorrelated with mean location or interresponse times (IRTs; top panels of Figures 4 and 5), suggesting that the two aspects of behavior are the result of independent processes.

Our results from the pigeon operant pecking procedure appear to support the general principle of the *Law of Expect*: the greater the expectation of an appetitive reinforcer, the less variability would be present in the response. This was true for both temporal and spatial dimensions of behavior. But would

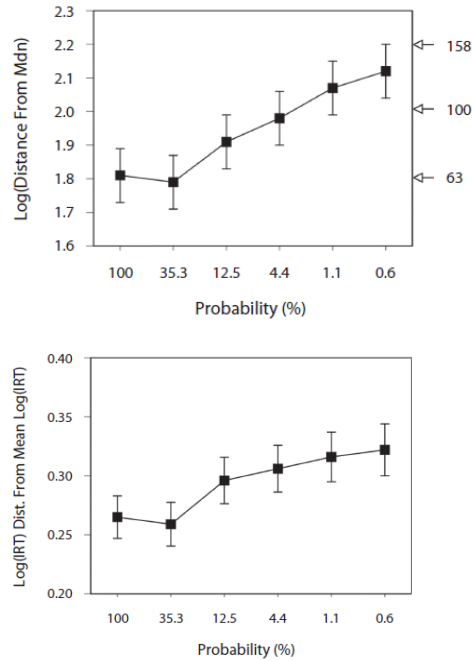


this same rule apply to a Pavlovian contingency in which delivery of the reward is not dependent on the subject's response? We conducted an autoshaping procedure (Brown & Jenkins, 1968; Williams & Williams, 1969) similar to the operant procedure described above (Stahlman, Young, & Blaisdell, 2010b). As in the operant procedure, each cue was associated with one of six probabilities of food US, except that each trial was fixed at 10 s in length, and if food was scheduled to be delivered, it was not dependent on screen pecks at the disk.

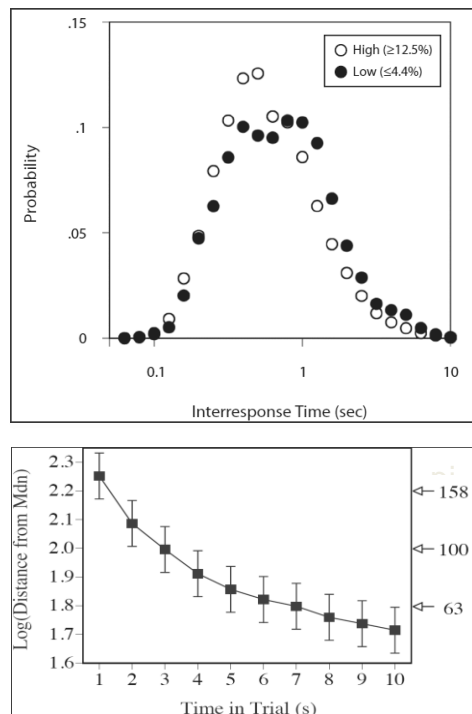


**Figure 6.** Peck density plots for each of the six Pavlovian cue trial types, collapsed across subject. The peck density plot models a smooth surface that describes how dense the data points are at each point in that surface and functions like a topographical map. The plot adds a set of contour lines showing the density in 5% intervals in which the densest areas are encompassed first. The JMP (Version 8, SAS Institute, Cary, NC) bivariate nonparametric density function was used to generate these plots. Reprinted from Stahlman, Young, and Blaisdell (2010b) with permission.

Peck density plots in Figure 6 show the spatial variation elicited by each of the Pavlovian cues. As with the operant procedure, screen pecks became more variable as probability of reward signaled by the Pavlovian cue went down. Figure 7 shows the quantification of variation in the spatial (top) and temporal (bottom) dimensions. As in the operant procedure, response variation in both dimensions was inversely related to the signal value of the Pavlovian cues. Thus, variation of behavior is inversely related to probability of reward for both instrumental and Pavlovian contingencies. The top panel of Figure 8 reveals the flatter, wider probability distribution of IRTs on the three lowest-rewarded trials (< 4.4%), relative to the three more frequently rewarded trials (> 12.5%). Thus, as with operant behavior, lower probabilities of reward produced a greater variety of IRTs both shorter and longer than the mean.



**Figure 7.** Top panel: Graph of the log of the mean distance (in pixels) from the individual bird median spatial location as a function of reward probability. The raw values of the dependent variable are located on the right-side vertical axis. Bottom panel: Log(IRT) of temporal distance (in seconds) from the mean log(IRT) for each subject as a function of reward probability. Error bars denote standard errors of the means. Reprinted from Stahlman, Young, and Blaisdell (2010b) with permission.



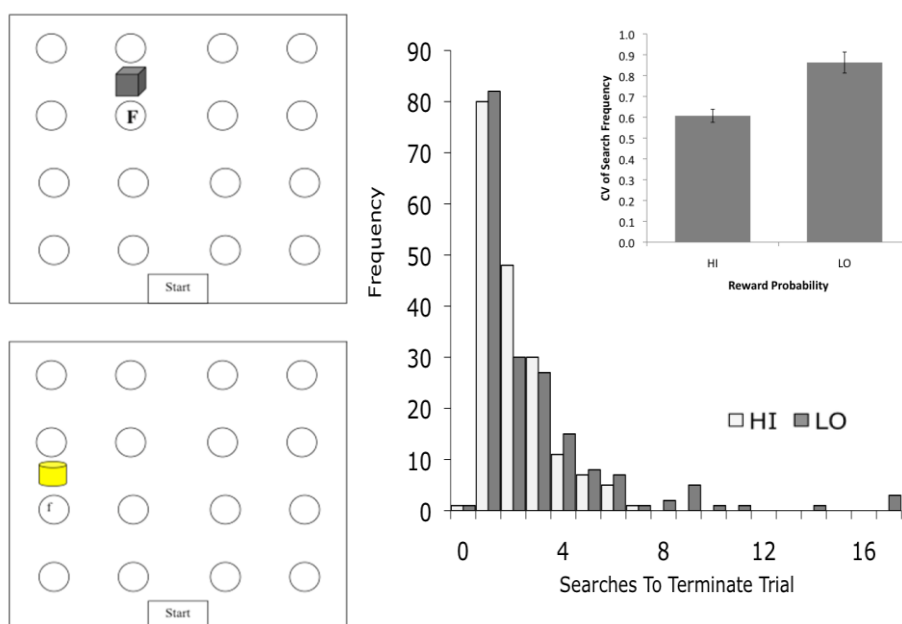
**Figure 8.** Top panel: Response probability as a function of IRT (in seconds) on a logarithmic scale for the lower and higher probability Pavlovian stimuli (cf. Gharib et al., 2004). Bottom panel: Graph of the log of the mean distance (in pixels) from the individual bird median spatial location as a function of trial time. The raw values of the dependent variable are located on the right-side vertical axis. Error bars denote standard errors of the means. Reprinted from Stahlman, Young, and Blaisdell (2010b) with permission.





This result is particularly interesting because variability of behavior is dependent on the likelihood of food delivery even when the response is entirely inconsequential. Despite its impotence, pecking was acquired and maintained throughout training. We found that Pavlovian responding was more variable in both spatial and temporal domains on trials signaling a low probability of reinforcement. This indicates that elicited behavior is an inverse function of Pavlovian expectation of positive outcomes.

One would expect that if variability decreases as an increasing function of reward expectation, then the closer the subject is within a trial to the time that a reward might be expected, the less variable should be the response. Gharib, Derby, and Roberts (2001) reported this in an instrumental peak-time procedure in rats. We also found this in our Pavlovian procedure with pigeons (Stahlman et al., 2010b). As shown in Figure 8 (bottom panel), variation in screen peck locations decreased as a function of time in the trial. Variability was lowest at the termination of the trial, when food would be delivered on rewarded trials.



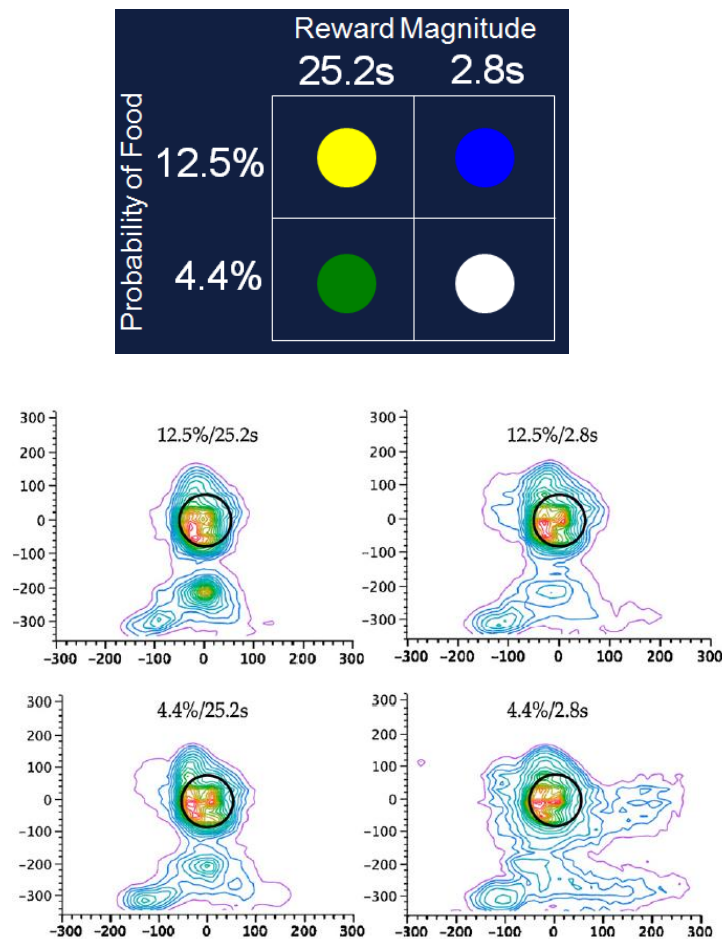
**Figure 9.** Left panels: diagram of the experimental apparatus. Top panel: Example of a HI probability trial (100% probability of reinforcement). The cube is a landmark indicating both the likelihood of reinforcement and the location of food (F). Bottom panel: Example of a LO probability trial (20% probability of reinforcement). The cylinder signals the probability of reinforcement and possible location of food (f). Right panel: 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 cups investigated prior to the termination of a trial. Inset: Mean (+/- SEM) coefficient of variation (CV) of rats’ search frequency across the two levels of reward probability. Reprinted from Stahlman and Blaisdell (2011a) with permission.

Of course, learning is not unique to contrived situations such as a conditioning chamber. As such, it is important to establish learning in more-naturalistic settings that may better reflect contingencies animals should experience in the wild. We sought evidence that the *Law of Expect* would manifest in a navigational task in an open field (Stahlman & Blaisdell, 2011a). In this experiment, rats learned to use two different landmarks (small wood blocks that differed in shape and brightness) to find a hidden food reward buried in sand in one of 16 cups on a large wooden board (Figure 9, left panel). The goal was always located to the South of the landmark. The goal was always baited with food in the presence of the high-food landmark, but only on 20% of the trials with the low-food landmark. If rats had a lower expectation of food on trials with the low-food landmark, then we expected to find greater variability on



these trials than on trials with the high-food landmark. Unlike in an operant chamber, we measured variability in total number of cups searched before the rat searched in the goal cup. Thus, if the rat had a lower expectation of food on low-food trials, they should explore more before looking in the goal compared to on high-food trials. This is what we found (Figure 9, right panel).

If behavioral variation in the learned response is inversely related to outcome expectation in a lawful way, then any manipulation that alters outcome expectation should correspondingly affect variation in the learned response. The experiments described above all manipulate the probability of reward. To further explore this hypothesis, we investigated two additional manipulations that affect reward expectation; reward magnitude and delay to reward (Stahlman & Blaisdell, 2011b).

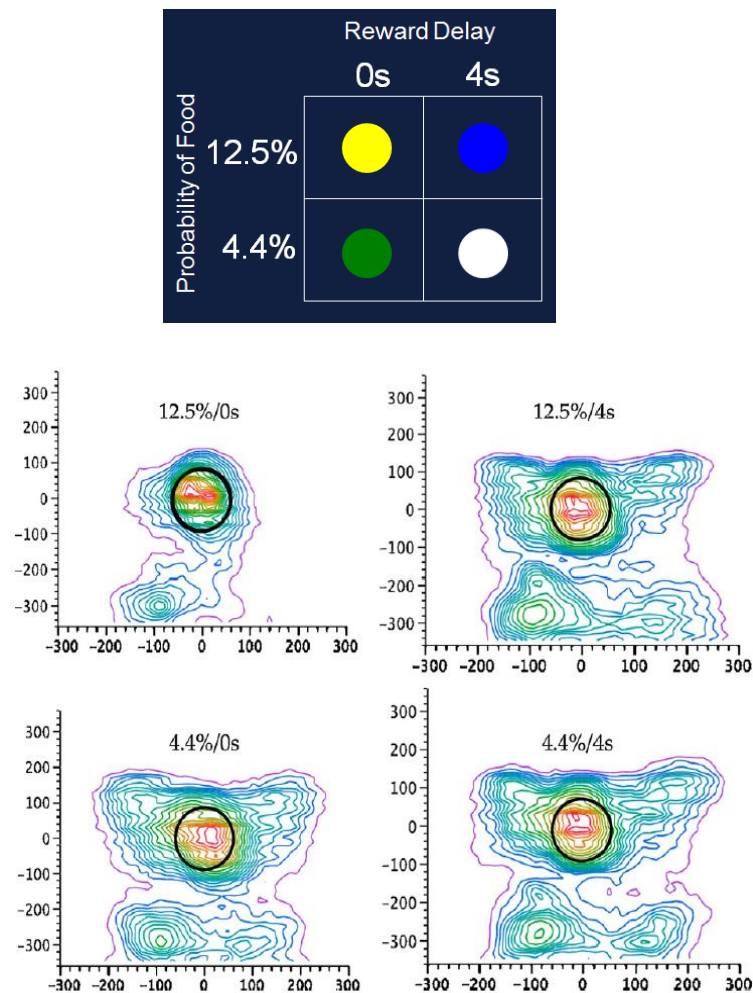


**Figure 10.** Top panel: Diagram of the 2x2 within-subject design. Each of four discriminative stimuli was associated with one of two probabilities of reinforcement and one of two magnitudes of reinforcement. Bottom panel: Nonparametric density plots illustrating the mean spatial location of pecks on the touchscreen for each trial type, collapsed across subject. The peck density plot models a smooth surface that describes how dense the data points are at each point in that surface and functions like a topographical map. The plot adds a set of contour lines showing the density in 5% intervals in which the densest areas are encompassed first. The black circle indicates the location and size of the stimulus disc, which is centered at (0,0). Units on both x- and y-axes are in pixels. Reprinted from Stahlman and Blaisdell (2011b) with permission.

We previously found that the greatest difference in behavioral variation tended to occur between cues signaling probabilities of food of 12.5% and 4.4% (e.g., Figure 6; Stahlman et al., 2010b). Thus, we chose to investigate the effect of changing the magnitude (Figure 10) or delay (Figure 11) of reward on variability produced at these two levels of probability. Reward size was manipulated by presenting either a



single 2.8-s delivery of grain (small reward) or 9 consecutive 2.8-s deliveries of grain for a total of 25.2 s of reward (large reward; Figure 10, top panel). We replicated the difference in spatial and temporal variation of pecking to the high probability (12.5%) and low probability (4.4%) cues at the small reward magnitude as used by Stahlman et al. (2010a). Would increasing reward magnitude reduce response variation on 4.4% probability trials, which typically show higher variation than do 12.5% trials? The screen peck density plots in the bottom panel of Figure 10 reveal that this was indeed the case. Figure 12 shows quantitatively that both spatial (top left) and temporal (bottom left) variation to the 4.4% probability cue was significantly reduced in the High magnitude relative to Low magnitude of reinforcement conditions.

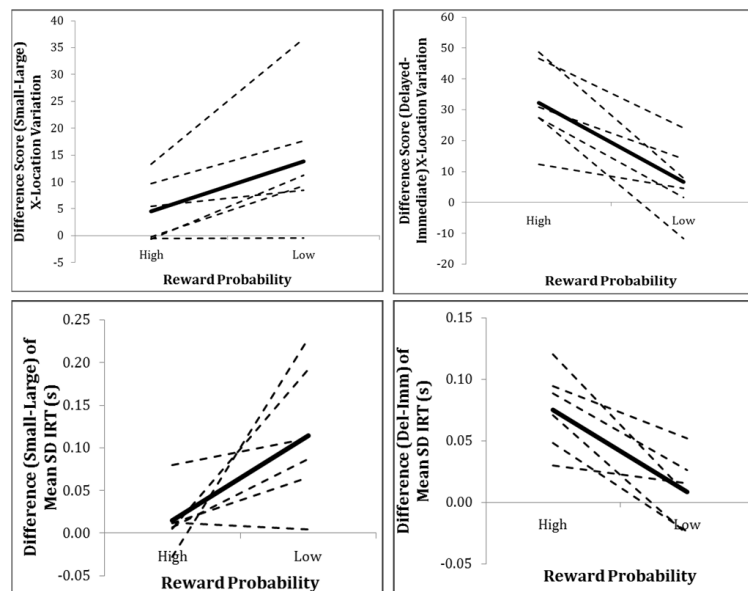


**Figure 11.** Top panel: Diagram of the 2x2 within-subject design. Each of four discriminative stimuli was associated with one of two probabilities of reinforcement and one of two delays to reinforcement. Bottom panel: Nonparametric density plots illustrating the spatial location of pecks on the touchscreen on each trial type for a representative individual subject in the previous experiment (see caption for Figure 10). Reprinted from Stahlman and Blaisdell (2011b) with permission.

A separate experiment replicated the difference in spatial and temporal variation in peck responses between the 12.5% and 4.4% probability cues when reward was delivered immediately at the end of the trial. Delaying the reward by 4-s after trial termination (Figure 11, top right panel), however, resulted in increased spatial variance in the screen peck density plots on the 12.5% probability trials. Figure 12 shows quantitatively that both spatial (top right) and temporal (bottom right) variation to the 12.5% probability cue significantly increased in the 4-s Delay relative to 0-s Delay to reinforcement



conditions. Thus, manipulations that increased or decreased the expectation of reward had the predicted systematic effects on response variability.



**Figure 12.** Top Left: Mean difference score of the mean standard deviation of x-axis response location as a function of reward probability (High = 12.5%, Low = 4.4%; Large = 25.2 s reward, Small = 2.8 s reward). Values were obtained by subtracting each bird's mean variation on Large trials from their variation on Small trials for each level of reward probability. The dashed lines indicate individual birds; the solid line indicates the mean across birds. Bottom Left: Mean difference score of the mean standard deviation of log IRT as a function of reward probability. Top Right: Mean difference score of the mean standard deviation of x-axis response location as a function of reward probability (High prob = 12.5%, Low prob = 4.4%; Imm = 0-s delay, Del = 4-s delay). Bottom Right: Mean difference score of the mean standard deviation of log IRT as a function of reward probability. Reprinted from Stahlman and Blaisdell (2011b) with permission.

### Functional analysis of the *Law of Expect*

To recap, we found evidence for the *Law of Expect*, expectation-controlled response variability, in both the conditioning chamber and the open field, in rats and pigeons, and involving operant and Pavlovian contingencies. Response variation was observed in both spatial and temporal dimensions of behavior, as well as in choice behavior in the open field. The *Law of Effect* describes the well-known role of reinforcement in selecting for learned responses. It operates by strengthening Stimulus-Response (S-R) associations for reinforced responses, but weakening S-R associations for non-reinforced responses. The strength of each S-R association determines the probability of the response. *The Law of Expect*, by contrast, describes the role of outcome (reward) expectation on increasing or decreasing variation in the response. It operates through the strengthening and weakening of Stimulus-Outcome (S-O) associations. The strength of each S-O association determines the degree of variation of the response. Stronger S-O associations bias the individual toward more stereotyped learned responses indicative of an exploitation behavioral mode. Weaker S-O associations bias the individual toward more variable responses indicative of an exploration behavior mode.

It makes a good deal of sense for an animal to behave with greater variability if reward is unlikely. Gharib et al. (2004) describe the adaptive functionality of the relationship between variability and reward expectation thusly:

If an animal's actions vary too little, it will not find better ways of doing things; if they vary too much, rewarded actions will not be repeated. So at any time there is an optimal amount of



variation, which changes as the costs and benefits of variation change. Animals that learn instrumentally would profit from a mechanism that regulates variation so that the actual amount is close to the optimal amount. (p. 271).

The *Law of Expect* should naturally occur under any situations where response variation is driven by outcome expectation. That is, organisms that are capable of learning through reward (instrumental and Pavlovian), should be equipped by evolution with a learning mechanism that instantiates the *Law of Expect*, just as it is equipped with a learning mechanism that instantiates the *Law of Effect*. Suppose a new region of space (away from the cue on the touchscreen or from the landmark in the open field) suddenly became available as a “secret cache” of reward. The likelihood that the subject would discover this new resource would be significantly higher on high-variation trials than on low-variation trials. Likewise if a new method of procuring a reward were made available, such as a change in the timing or sequence of responses, either in absolute terms or relative to prior rewarded responses (Neuringer, 2002; 2004). Greater variation in the form, timing, or sequence of the response entails a greater likelihood of discovering the new reward (see also Pecoraro, Timberlake, & Tinsley, 1999).

These suppositions elegantly map on to real world foraging situations. While an animal is foraging for food, depletion of the local food source should increase the probability of searching elsewhere. Likewise, if one initially successful method of extracting embedded food (mollusks buried under sand, grubs hiding in the bark of a tree, etc.) stops being successful, perhaps changing the form of the extraction response (e.g., digging more forcefully, twisting the beak into a crack in the bark a different way) may increase the probability of success.

Variability in behavior has been proposed to play a functional role in the continuous adaptation of behavior in many domains, including motor learning and rehabilitation (Bartlett, Wheat, & Robbins, 2007; Davids et al., 2003; Manoel & Connolly, 1995; Riley & Turvey, 2002; Wu et al., 2014), decision making (Daw et al., 2006), modulation of stress and self-regulatory strength (Geisler et al., 2010; Kryptos et al., 2011), sensory integration (Gordus et al., 2015), and differences in fitness between social groups (Modlmeiera & Foitzika, 2011).

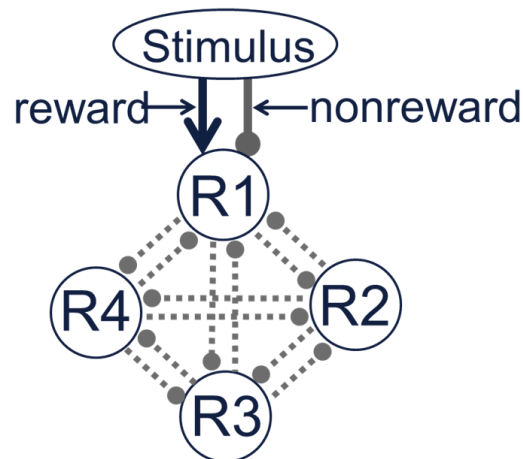
### The Modified Law of Effect

We previously noted that Epstein’s generativity theory avoids accounting for behavior in terms of psychological mechanism and instead provides solely a formal account of novel behavior (e.g., Epstein, 2014). Yet, our data suggest that outcome expectancy is an important regulator of response variation. Is there any alternative to outcome expectancy that does not rely on a cognitive (i.e., representational) account of response variation? Let’s revisit the *Law of Effect* to see what led us to propose the *Law of Expect*. According to Thorndike’s (1911) *Law of Effect*, S-R associations are strengthened when followed by a rewarding outcome, and weakened when followed by nonreward (or other dissatisfying outcomes, e.g., punishment). Once an S-R association has been weakened through nonreward, there is no way to increase it again without directly rewarding it. Thus, following acquisition, extinction of the rewarded S-R association should have no effect on the other ineffective S-R associations that had been weakened when previously followed by nonreward. Because of this, those other responses whose associations to S had been weakened should remain of low likelihood, and their resurgence should not be observed. In contrast to this, Sweeney and Shahan (2016) found responding of a non-rewarded response to increase during extinction of a previously-rewarded response.

Let’s consider modifying the *Law of Effect* to see if we can remove this constraint. First, assume that in any particular situation (a conditioning chamber, the presence of a discriminative stimulus, etc.) all



possible responses have a prior probability<sup>4</sup>. Second, we can leave the first part of the law the same. That is, Postulate 1 is that an S-R association is strengthened when followed by reward. Now consider adding a second postulate that, there exist inhibitory connections between all possible responses, such that as the S-R strength for one response increases, it exerts an increased amount of mutual inhibition to all other responses, thereby decreasing their likelihood of being expressed. This modification to the *Law of Effect* is captured by the dashed unidirectional links between all potential responses shown in Figure 13 (in which only 4 of likely many possible responses are shown to simplify exposition of the theory). By proposing a process of mutual inhibition, if R1 is followed by reward, the S-R1 association is strengthened, but inhibitory links from R1 to R2, R3, R4, & etc. increase in strength as well.



**Figure 13.** Associative structure of the Modified Law of Effect. Excitatory S-R1 association (black arrow) is strengthened when R1 is followed by reward. Inhibitory S-R1 association (gray connection) is strengthened when R1 is followed by nonreward. Two unidirectional mutual inhibitory links (dotted connections) exist between each response. Increases in excitatory S-R1 connection increases R1's mutual inhibition to all other responses. Increases in inhibitory S-R1 connection decreases R1's mutual inhibition to all other responses.

We must also postulate a change in the second part of the original law. Postulate 3 is that any response followed by nonreward does not weaken the prior S-R strength, but instead results in strengthening of an inhibitory S-R association. This is shown by the dark gray inhibitory connection between the Stimulus and R1 in Figure 13. Thus, if R1 is not followed by reward, its prior excitatory S-R association remains intact, but an inhibitory S-R association is strengthened.

The three postulates can be stated as followed:

1. Postulate 1: Response R followed by rewarding outcome O in the presence of stimulus S will strengthen an excitatory S-R association for R.
2. Postulate 2: Response R followed by rewarding outcome O in the presence of stimulus S will strengthen the inhibitory R-R connections to all other responses  $R_n$ .
3. Postulate 3: Response R followed by nonreward in the presence of stimulus S will strengthen an inhibitory S-R association for R.

<sup>4</sup> We have intentionally avoided the issue of how a response is defined. Responses can be discrete individual actions, of which no two emitted or elicited responses are likely ever identical, or classes of similar responses, such as a lever press or a key peck, which are traditional operants. In our analysis, we treat responses in an idealized fashion as every possible discrete response that can occur in a situation. We have empirically demonstrated that attributes of responses, such as a location in time or space with respect to the situation inside the conditioning chamber or in an open field, vary in a continuous fashion. As a result, variation in spatial, temporal, and choice dimensions are predicted to result from changes in S-R associations for specific Rs, and mutual inhibition between Rs.



We also need a rule that maps learning onto behavior. The probability of a response is determined by the net excitatory S-R strength of the R, inhibitory S-R strength of the R, and total mutual inhibition received from all other Rs. The operation of these three postulates results in a *Modified Law of Effect* that can potentially account for a broader range of phenomena than could the original, including the inverse relationship between probability of reward and response variation. To test whether the *Modified Law of Effect* can account for the response variation phenomena reviewed above, we compared simulations of the *Modified Law* to the original *Law of Effect*. We also include simulations of the *Modified Law* with only one Postulate 2 (mutual inhibition) or 3 (inhibitory S-R associations), but not both, to assess the necessity and sufficiency of each postulate.

### Model Simulation Methods

We compared four different models of instrumental learning involving variations of the *Law of Effect*. In all conditions, an agents' behavioral repertoire was restricted to ten possible actions. On each trial, an agent performed an action with the maximal overall excitation out of all available alternatives. If all actions were strongly inhibited (i.e. overall excitation < 0), no action was performed and the trial was recorded as an omission. In the *Law of Effect* (LoE) model, input excitation to each action determined the overall excitation. We then introduced either additional input inhibition (LoE+InInh) as a result of nonreward of an action, or mutual inhibition (LoE+mInh) between actions as additional model parameters. The final comparison model included both the input and mutual inhibition (LoE+InInh+mInh) to fully model the *Modified Law of Effect* (MLoE) as described above.

Initial input excitation and inhibition values were drawn from independent normal distributions with means of 0.5 and standard deviations (SD) of 0.01, and therefore were on average balanced at baseline. Each action inhibited the other alternatives by mutual inhibition that was set to 50% of the resulting excitation (sum of input excitation and inhibition). Additionally, we introduced input noise (mean=0, SD=0.1) to account for a variety of factors, such as attention and interference (see Chan et al., 2010; Ryan et al., 2012), that can affect animal behavior in experimental settings. A single action was set as correct for all agents. Each trial terminated with a reward if an agent selected the correct alternative and no reward otherwise. The input excitation and inhibition values were updated based on the observed outcome according to the linear operator rule (Bush & Mosteller, 1951; Kacelnik & Krebs, 1985):  $In_{t+1} = In_t(1-\alpha) + \alpha r_t$ , where  $\alpha$  (0.01) represents a learning rate,  $In_t$  represents the input value at the current time and  $r_t$  represents experienced reward. On each trial only the inputs to selected action were updated. In LoE and LoE+mInh conditions the input excitation was increased if the performed action was followed by reward and decreased if no reward was presented. In contrast, in LoE+InInh and LoE+InInh+mInh conditions, receiving a reward resulted in an increase in input excitation, whereas not receiving a reward led to increases in input inhibition.

We compared the predictions of the models about response variability under conditions of different reward magnitudes ( $r=1.0, 0.9, 0.8, 0.7$  or  $0.6$ ) and probabilities of reinforcement (1.0, 0.8, 0.5, 0.2). In a separate set of simulations we introduced an extinction phase, during which no reward was presented regardless of the action chosen. We also analyzed response variability during training and extinction under the condition when response variability (i.e. performing different actions on consecutive trials) was reinforced. For these two sets of simulations the reward magnitude during the training phase was set to 1.0 and correct responses were reinforced 100% of the time during the acquisition phase. Finally, we addressed response variability in a successive anticipatory contrast procedure, where agents are initially trained under one reward magnitude and shifted to a different magnitude during the contrast phase. We present the results with reward magnitudes of 1 (large) and 0.7 (small). We simulated the

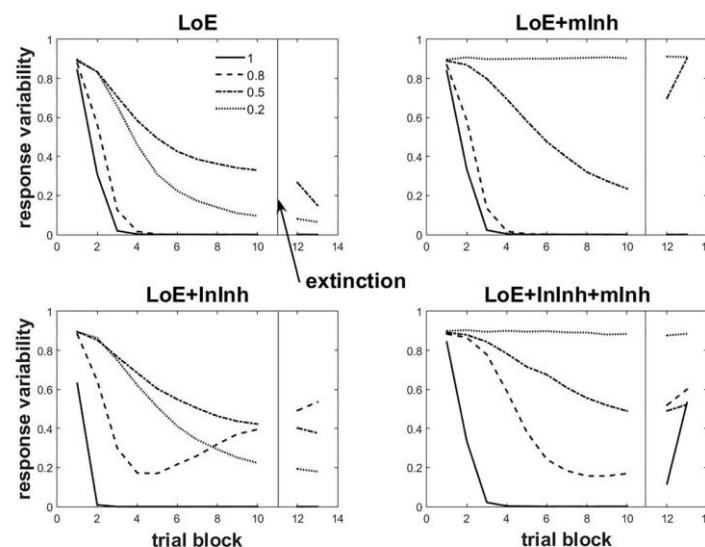




behavior of 100 agents in each condition for 10, 12 or 20 blocks of 100 trials. MATLAB R2015a (MathWorks, Natick, Massachusetts) was used in all analyses. All codes are available upon request.

### Model Simulation Results

Figure 14 shows the simulations of the four models of response variation as a function of probability of reward signaled by four different discriminative stimuli. We tested values of reward probabilities of 1.0, 0.8, 0.5, and 0.2. As is apparent in the figure, two of the simulations successfully captured the monotonic rank ordering of the inverse relationship between probability and response variation as empirically shown by our pigeons. The two models that were successful both included the process of mutual inhibition. In contrast, the other two models were only partially successful in predicting the monotonic inverse relationship, correctly ordering the low response variation observed to the two highest reward probabilities (1.0 and 0.8). Surprisingly, both the original *Law of Effect* and the modified model that only included the process of input inhibition produced more response variation to the 0.5 probability condition than to the 0.2 probability condition. Sensitivity analyses revealed that this increased variation to the 0.5 compared to the 0.2 probability condition did not depend on the value of free parameter  $\alpha$ , or noise levels (unless they were reduced to 0) and was therefore a true emergent property of the models. Although we have not systematically assessed the underlying cause of increased variation, in both LoE and LoE+InInh, the 0.5 condition generates a pattern of input updating that balances increases and decreases in overall excitation due to reward and non-reward respectively, preserving the initial input strength. Such a condition can also render agents more sensitive to input noise, promoting behavioral shifting. Conversely, the 0.2 condition could have resulted in decreases in overall excitation across responses, leading to higher impact of increases in excitation due to reward when it occurred. Additionally, since agents were allowed to omit responding when all options were strongly inhibited, decreases in response variability in the 0.2 condition may be partially explained by response withholding. This ordering stands in stark contrast to the results demonstrated in our pigeons. Thus, it appears that including mutual inhibition is necessary to correctly model the empirical data with a *Law of Effect* type reinforcement learning mechanism.



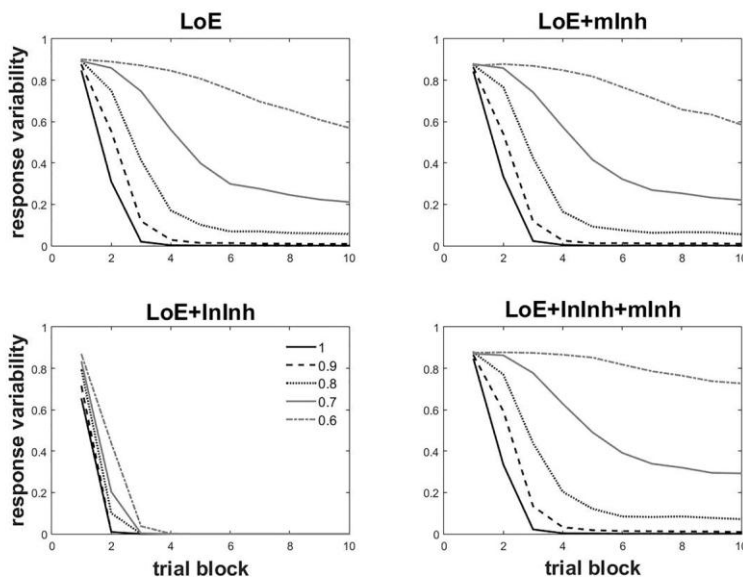
**Figure 14.** Simulations of the original Law of Effect (LoE, upper left), as well as the Modified Law of Effect (MLoE) with only inhibitory S-R links (LoE+InInh, lower left), only mutual inhibition (LoE+mInh, upper right), or both (LoE+InInh+mInh, lower right). Simulations show changes in response variability during acquisition and extinction as a function of probability of reward.





How does the *Modified Law of Effect* account for the finding that response variability is inversely correlated with probability of reward? If we consider that a stimulus associated with a high probability of reward can produce large increases in the excitatory strength of a relatively few rewarded S-R associations, then these Rs will exert strong mutual inhibition to all other responses. A stimulus associated with a relatively low probability of reward, on the other hand, would produce increases in both excitatory and inhibitory S-R associations for some Rs, thus causing these Rs to exert less mutual inhibition to all other responses. As a result, a larger subset of the total possible responses will be emitted in situations of low probability relative to high probability of reward. Since behavior occurs as an unending stream (e.g., Epstein, 2014), the overall rate of responding should be affected relatively little by the different probabilities of reward (cf., upper panels of Figures 4 & 5), while the variance in responses is predicted to differ quite a lot. That is, the subject always must be doing something, even if it's sleeping. A hungry pigeon, however, is unlikely to spend much time resting. Instead, it will spend time exploring the screen, which is a context of prior reinforcement (Racey et al., 2011). Thus, the *Modified Law of Effect* appears to provide a good account of the effects of reward probability on response variability.

Somewhat similar results were also found in how the models simulate the effects of reward magnitude (Figure 15). We tested reward magnitudes of 1.0, 0.9, 0.8, 0.7, and 0.6. Both models that incorporated mutual inhibition correctly (i.e., in relation to the empirical data) found that lower magnitudes of reward produce higher amounts of response variability in a monotonic rank-order. Unlike the case with probability of reward, the original *Law of Effect* does correctly predict an inverse relationship between reward magnitude and response variation. The *Modified Law* that incorporates only input inhibition but not mutual inhibition is completely wrong in predicting no difference in response variation at asymptote. While it does model the correct rank order of reward magnitude to response variation early in learning, responses in all cases quickly converge on a complete absence of variation.



**Figure 15.** Simulations of the original Law of Effect (LoE, upper left), as well as the Modified Law of Effect (MLoE) with only inhibitory S-R links (LoE+InInh, lower left), only mutual inhibition (LoE+mInh, upper right), or both (LoE+InInh+mInh, lower right). Simulations show changes in response variability during acquisition as a function of Magnitude of reward.

What about the case of variation of Pavlovian responses (Stahlman et al., 2010b; Figure 6-8)? Here the application of the *Modified Law of Effect* seems less clear. Pavlovian responses, by definition, are



established without any response-reward contingency. Only if the Pavlovian responses can be reinterpreted as instrumental ones would we be able to apply the *Modified Law of Effect*. Autoshaping is a Pavlovian conditioning phenomenon that bears strong resemblance to instrumental conditioning more than to Pavlovian conditioning in many ways (Gormezano & Kehoe, 1975; Herrnstein & Loveland, 1972; Hursh, Navarick, & Fantino, 1974) and can be simulated by a model using bottom-up processes that ignore the Pavlovian-instrumental distinction (Burgos, 2007). So perhaps it is not surprising to find Pavlovian response variability to be inversely related to probability of US delivery. A true test of the application of the *Modified Law of Effect* requires a more conventional Pavlovian conditioning preparation that involves goal tracking rather than sign tracking behavior (Burns & Domjan, 1996; Lesaint et al., 2014).

Can the *Modified Law of Effect* account for the observation that response variation decreases with proximity to the time of reward (Figure 8, bottom panel)? If distal responses have weaker excitatory S-R associations, while proximal responses have stronger excitatory S-R associations, thereby exerting stronger mutual inhibition against other responses, then this observation appears to emerge naturally from the *Modified Law of Effect* as well. That is, responses made closer to the time of reward will gain greater excitatory S-R strength, and thereby greater mutual inhibition, compared to more distal responses.

The original discovery of the relationship between reward versus nonreward and response variation was observed in the study of extinction of operant responding (Antonitis, 1951; Eckerman & Lanson, 1969; Frick & Miller, 1951; Herrick & Bromberger, 1965; Millenson & Hurwitz, 1961; Notterman, 1959; & Stebbins & Lanson, 1962). Thus, a critical test involves how the additional postulates of the *Modified Law of Effect* may allow for an account of this well-documented phenomenon. Figure 14 provides the simulated data for each of the four models tested. What is intriguing is that only the *Modified Law of Effect* that incorporates both input inhibition and mutual inhibition correctly models an increase in response variation during extinction. The original law and the one that includes input inhibition fail to predict any increase in response variation from zero during extinction. The modified model incorporating only mutual inhibition does predict an increase in response variation during extinction, but only for responses that have not reached a zero asymptote during acquisition. Thus, we can conclude that both postulates are necessary to model this robust empirical phenomenon.

How does the *Modified Law of Effect* account for the increase in variability during extinction of a previously reinforced response? Nonreward increases the inhibitory S-R association (input inhibition) for the previously reinforced R. When R is inhibited by the stimulus via this inhibitory S-R connection, all other responses are released from R's mutual inhibition. By releasing mutual inhibition to all other responses, their prior excitatory S-R associations can resurge (cf. Winterbauer & Bouton, 2010). With complete removal of the mutual inhibition (once extinction is complete), resurgence is complete and the previously inhibited responses return to their original prior probability. Thus, the *Modified Law of Effect* explains the increase in response variation during extinction as *resurgence* of prior responses originally observed prior to reinforcement in acquisition.

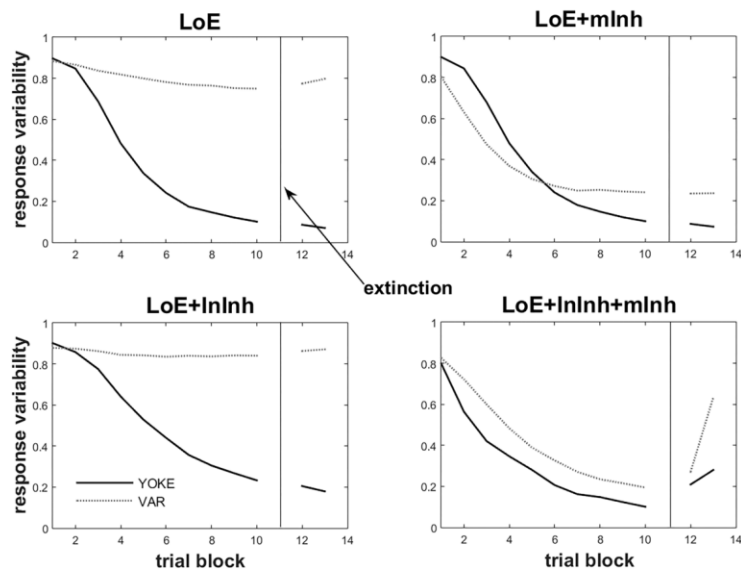
It cannot be overstated that the success of the *Modified Law of Effect* incorporating both types of inhibition to correctly model an increase in response variation during extinction is a major breakthrough. Previously, both the increase in variation in extinction and to stimuli with a history of lower probability of reinforcement has been taken as strong evidence for a cognitive expectancy on the part of the subject. That is, such relationships had been thought to be beyond the *Law of Effect* to explain. Our modifications to the law incorporating a well-known basic process of inhibition overcome this hurdle.



### Variability as an operant

Given the success of the *Modified Law of Effect*, we were interested in what other aspects of response variation could be successfully modeled within an S-R framework. There has been a great deal of research on behavioral variability as an operant (see review by Neuringer, 2002). The question Neuringer and colleagues have studied is whether response variability can be the target of selection by reinforcement, that is, whether it can be an operant. A standard procedure to show this is to compare response variability in different groups of subjects, where variability is explicitly reinforced for one group but not for others. One way to reinforce response variation is to reinforce actions that the subject has not made recently, such as in the past 50 trials. Response variability in this VAR group is typically compared to a group that is reinforced for any response (Group ANY). But the rate of reinforcement will be much higher for Group ANY than for Group VAR. To control for this, each subject in a third group (YOKE) is reinforced for any response, but only when a partner in the VAR group to which it is yoked is reinforced. Neuringer and colleagues have found using this procedure that response variability is significantly higher in the VAR group compared to the ANY and YOKE groups, thereby suggesting that variability is a reinforceable dimension of behavior.

Neuringer et al. (2001) report an additional finding that, following operant reinforcement, response variability increases even more during extinction than the amount observed during reinforcement (see Figure 4 of Neuringer et al., 2001). Thus, we were interested in how the various versions of the *Law of Effect* would handle both operant variability and the increase in variability in extinction, even in Group VAR. Figure 16 shows the simulation results for the four models. All four models, even the original *Law of Effect*, can account for operant variability. This is not surprising because even the original model predicts that repeated responses that are nonreinforced will weaken their S-R strength. Given that repeated responses are not explicitly nonreinforced for Group YOKE, but are for Group VAR, the greater response variation in Group VAR is perfectly expected.



**Figure 16.** Simulations of the original Law of Effect (LoE, upper left), as well as the Modified Law of Effect (MLoE) with only inhibitory S-R links (LoE+InInh, lower left), only mutual inhibition (LoE+mInh, upper right), or both (LoE+InInh+mInh, lower right). Simulations show changes in response variability during acquisition on a partial schedule of reinforcement on a VAR or YOKE schedule (right panel), and extinction (left panel).



The original law and the modified law that only includes input inhibition fail to show a significant drop in response variation during acquisition. Empirically, however, variability does decrease somewhat as the time of reinforcement approaches (Stahlman et al., 2010b) even for operant variability (Cherot, Jones, & Neuringer, 1996). Thus, the two models in the left panel of Figure 16 fail to accurately capture the empirical phenomenon of operant variability. The modified laws that incorporate mutual inhibition, however, do capture the drop in response variability during reinforcement learning, while also accurately capturing the maintenance of greater variation when it is reinforced (though the *Modified Law* including only mutual inhibition predicts greater variability in the YOKE group early in training which is not observed empirically). What is interesting is that none of the models except for the fully *Modified Law of Effect*, predict an increase in response variation during extinction. Thus, only the modified law accurately models the increase in response variation observed in extinction following both conventional partial reinforcement of a target response and reinforcement of operant variability.

How does the *Modified Law of Effect* account for the drop in variability during acquisition followed by the increase in variability during extinction in Group VAR? Operant variability by definition involves the delivery of reinforcement contingent upon the emission of relatively unusual responses (Blough, 1966; Neuringer, 2004; Pryor, Haag, & O'Reilly, 1969). The reinforcer strengthens the excitation to those S-R associations it follows, which increases their likelihood relative to other responses. Furthermore, each increase in an S-R association following reinforcement increases the mutual inhibition directed to all other responses, thereby reducing their likelihood. During extinction, these same originally strengthened S-R associations receive increased S-R (input) inhibition, reducing the mutual inhibition these responses direct towards all other responses. This results in resurgence in the other previously inhibited responses. Said another way, by reinforcing responses, even using the criterion that responses must vary, variation will necessarily go down relative to prior amounts that preceded the introduction of reinforcement. Extinction brings the level of response variability back to the prior state that existed before the reinforcement protocol began.

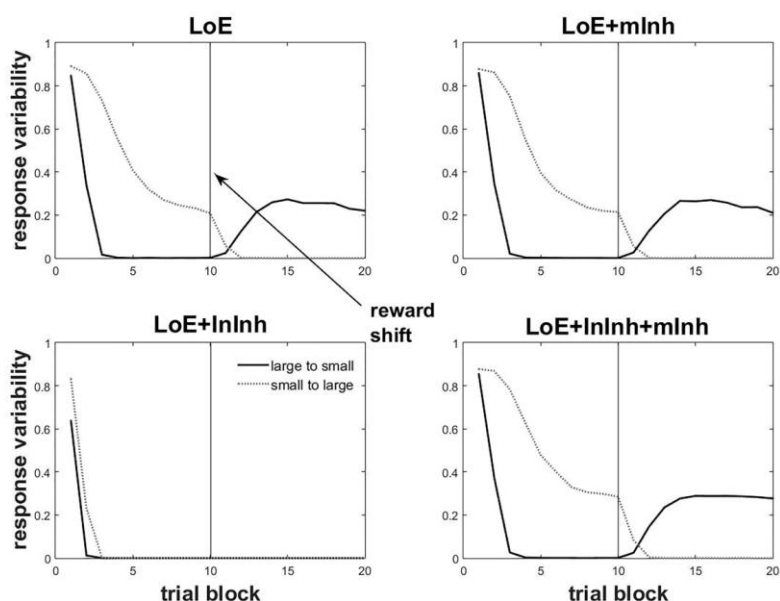
### Contrast effects

A classic demonstration of expectation representation in learning comes from demonstrations of anticipatory contrast effects (e.g., Crespi, 1942; Elliott, 1929). While larger or higher-quality rewards tend to support stronger or faster instrumental responding compared to smaller or lower-quality rewards, a change from a small to a large reward can cause an increase in response strength beyond that shown in a control group where large reward is consistently employed. This is called positive anticipatory contrast. Likewise, a change from a large to a small reward (or from a more preferred to a less preferred reward) can produce a decrease in response strength below that shown in a control group where the smaller or lower-quality reward is consistently employed. This is called negative anticipatory contrast. Positive and negative anticipatory contrast effects have been argued to show evidence that the subject encodes the value (quantitative or qualitative) of the reward, which leads the subject to have an expectation of the reward with a learned value (Flaherty & Grigson, 1988). Shifting the value of the reward larger or smaller than the expected amount changes performance of the instrumental response so that it overshoots (above in the case of upshifts; below in the case of downshifts) the level of instrumental performance displayed by subjects for which the reward was not changed. Interestingly, anticipatory contrast effects have only been found in various species of mammals, but not in non-mammalian vertebrate species, such as pigeons, turtles, toads, or goldfish (Papini, 2008; but see Mackintosh et al., 1972; & Reynolds, 1961). These comparative data suggest that outcome expectation is only found in mammals, at least so far as it can be measured by its influence on responding when outcome value changes. Recently it was shown in a button pressing task that an unanticipated decrease in reward magnitude (and extinction) induced



behavioral variation in human subjects, while upshifts had no effect on variation (possibly due to a performance floor effect; Jensen et al., 2014).

We simulated positive and negative anticipatory contrast in our four models of the *Law of Effect* (Figure 17). The original *Law of Effect* nicely captures the difference in response variation observed for small versus large rewards, with more variation shown for a smaller reward. Interestingly, adding input inhibition to the original model abolishes this success, which suggests that this addition alone is untenable (as also shown in Figure 15). Nevertheless, despite accurately capturing the difference in variation induced by differences in reward size, the original law does not capture either negative or positive anticipatory contrast effects when rewards are shifted, as have been found in mammals. It turns out that adding mutual inhibition alone or in combination with input inhibition also fails to capture the anticipatory contrast effects observed in mammals. Thus, the data from mammals, at least, appears to necessitate the inclusion of outcome expectancy as part of what is learned during instrumental conditioning. The *Modified Law of Effect*, however, can account for the lack of contrast observed in non-mammalian vertebrates. The original *Law of Effect* can, too, however.



**Figure 17.** Simulations of the original Law of Effect (LoE, upper left), as well as the Modified Law of Effect (MLoE) with only inhibitory S-R links (LoE+InInh, lower left), only mutual inhibition (LoE+mInh, upper right), or both (LoE+InInh+mInh, lower right). Simulations show changes in response variability during acquisition on a CRF schedule of reinforcement with a large or small reward (right panel), and during the contrast phase when reward magnitudes were switched between conditions (left panel).

### Discussion of the Modified Law of Effect

We found that a *Modified Law of Effect* that includes a process of mutual inhibition between all potential responses that can occur in a situation, and that replaces the weakening of S-R connections from nonreward with increasing inhibitory S-R connection (Figure 13), successfully modeled many empirical phenomena involving behavioral variation that the original *Law of Effect* cannot accommodate. This includes the inverse relationship between probability of reward and response variation, effect of reinforcement of variation on maintenance of some degree of response variation, and an increase in response variation induced by extinction, even compared to the amount of variation conditioned through reinforcement.



These modifications to the law make sense in light of both an historical and contemporary understanding of behavior during extinction. While Thorndike (1913) treated nonreward as weakening of the S-R association, Pavlov (1927) originally thought of extinction learning as an inhibitory process rather than the erasure of associative strength. Contemporary research also supports the inhibition interpretation of extinction learning (e.g., Bouton, 2002; Rescorla, 1993). Rescorla in particular provided empirical support for the acquisition of an inhibitory S-R association during extinction learning. When a previously instrumental response is followed by nonreward, original learning appears not to be erased. Rather, the original learning is suppressed through the action of an inhibitory S-R association acquired during extinction. Evidence that may support this interpretation comes from phenomena by which the original response returns, such as with the passage of time (spontaneous recovery), with the presentation of a feature of the original learning situation (reinstatement), and when testing occurs in a context different from extinction (renewal). Likewise, according to the *Modified Law of Effect*, responses that are followed by nonreward result in strengthening of an inhibitory S-R association rather than a weakening (erasure) of an excitatory S-R association. Thus, this postulate of the *Modified Law of Effect* has strong empirical support.

The inclusion of mutual inhibition was also found to be critical to some of the successes of the *Modified Law of Effect*. In fact, without mutual inhibition, adding just S-R (input) inhibition degraded the explanatory power of the original *Law of Effect*. Mutual inhibition has received less attention as a mechanism of behavior in associative learning. Nevertheless, it has strong empirical support as a fundamental mechanism of neural circuits (Tunstall et al., 2002) and in mechanisms that ‘tune’ specificity in sensory and perceptual processes. Thus, mutual inhibition likely is an important, though overlooked, mechanism of behavior that operates in many learning situations. In our model, nonreward increased inhibitory S-R strength, thereby removing other responses from mutual inhibition. Many of the modified model’s successes derive from this process.

What are the limitations of the *Modified Law of Effect*? Despite its many successes, the *Modified Law of Effect* was unable to account for anticipatory contrast effects as found in mammals. Such contrast effects are best accounted for in terms of outcome expectancy learning. Thus, the *Law of Expect* might still have more explanatory power than even the *Modified Law of Effect*, at least for anticipatory contrast effects.

The *Law of Expect* may be a preferred account for other reasons as well (cf. Bolles, 1972). Because it involves outcome expectancy, it shares many features with goal-directed action—for which there is compelling evidence, while the *Modified Law of Effect* provides an account in terms of habitual control of behavior (as does the original *Law of Effect*). A test of goal-directed action involves the use of contingency degradation and outcome devaluation procedures (Balleine & Dickinson, 1998; Colwill & Rescorla, 1985, 1986). If either degrading the contingency between the response and the outcome, or devaluing the outcome (temporarily through satiation, or more permanently through conditioned taste aversion) result in a decrease in the instrumental response, we can say that the response is goal directed and that the outcome is part of the association motivating behavior. Likewise, these same two procedures could be usefully applied to dissociate the *Modified Law of Effect* from the *Law of Expect*. If contingency degradation and outcome devaluation reduce the response rate and increase response variation in the absence of further training, then we would have direct evidence for outcome expectancy—a result that would favor the *Law of Expect* over the *Modified Law of Effect*.

Likewise, the *Law of Expect* and *Modified Law of Effect* make different predictions regarding response variability in a Pavlovian to Instrumental Transfer (PIT) design. The subject would first be trained on two stimuli, one as a High probability and the other as a Low probability Pavlovian cue, but without letting the subject sign-track. This would be followed by a nonreinforced probe test where the subject is given access to the cues so that sign tracking behavior could emerge. According to the *Law of*



*Expect*, sign-tracking responses should be more variable to the low probability than to the high probability cue. This prediction is derived from the postulate that the expectation of the outcome regulates response variability. The *Modified Law of Effect*, however, fails to make this prediction because it relies on S-R learning (both excitatory and inhibitory) to occur before differences in response variability emerge. For now, we must await such experiments to be conducted.

If there is strong empirical support for the necessity of a *Law of Expect*, what are its limitations? That is, is there still a need for a *Law of Effect* (Dennett, 1975)? Despite its many successes, the *Law of Expect* fails to account for learned behavior that is insensitive to the effects of outcome devaluation or contingency degradation manipulations. Such behavior is argued to be under habitual control rather than goal-directed (Balleine & Dickinson, 1998; Colwill & Rescorla, 1985, 1986). Habitual control of behavior is best accounted for in terms of S-R learning without (effective) representations of the outcome either through S-O or R-O associations. Thus, it appears a *Law of Effect* (S-R associations) is still needed to provide a full account of learned behavior.

### **The *Law of Expect* and creativity**

Reward expectation may also modulate creative behavior (Stahlman et al., 2013). Empirical evidence for the negative effect of reinforcement on behavioral variability with respect to creativity comes from studies showing that variability tends to decrease as an animal draws nearer to reward (Gharib et al., 2001; Neuringer, 1991; Schwartz, 1982; Stahlman et al., 2010b). The observation that variability tends to decrease with approach to reinforcers certainly suggests that reinforcement interferes with production of novel behavior (Cherot et al., 1996). Reinforcement of variability tends to increase total levels of variability (e.g., Page & Neuringer, 1985), but as outcomes become more proximal on a given trial, variability decreases. Each of these opposing effects of reinforced variability (i.e., the *Law of Effect* and the *Law of Expect*) may have implications for optimal methodology for training creativity (see also Epstein, 1990).

### **Reward Circuitry**

Since 2000, there has been a large amount of research devoted to investigating the role of the basal ganglia in instrumental behavior. Normal functioning of the components of the basal ganglia appears to be critical for instrumental learning procedures. The nucleus accumbens (e.g., Hernandez, Sadeghian, & Kelley, 2002; Koch, Schmid, & Schnitzler, 2000; Salamone, Correa, Farrar, & Mingote, 2007; Wyvell & Berridge, 2000) and the striatum (e.g., Wiltgen, Law, Ostlund, Mayford, & Balleine, 2007; Yin & Knowlton, 2006; Yin, Knowlton, & Balleine, 2006) seem to be important for the acquisition and performance of instrumental actions.

The basal ganglia are thought to be of primary importance in both expectation and in regulating behavioral variability particularly in associative preparations. In rhesus macaques, information regarding the size of an expected reward is encoded by neurons in the anterior striatum (Cromwell & Schultz, 2003); other studies have demonstrated that motor behavior in monkeys is shaped by incentive value encoded in the basal ganglia circuit (Pasquereau et al., 2007), and that neural activity in the caudate nucleus accurately predicts both rewarded and unrewarded action (Watanabe, Lauwereyns, & Hikosaka, 2003). There is evidence to support the role for prefrontal cortical structures in modulating the behavior in the striatum during reward encoding (Staudinger, Erk, & Walter, 2011). Graybiel (2005), among others, has suggested that reinforcement signals (e.g., magnitude and likelihood of reward) are instantiated in the basal ganglia. In addition, she suggests that the basal ganglia are critically important in maintaining the balance of exploration and exploitation in conditioned animal behavior, thereby optimizing response output to the expected conditions of reward. It is important to note the confluence of this suggestion with Gharib et



al.'s (2004) argument that variability in behavior must be appropriately modulated as the costs and benefits of variation change.

### Bird song learning

Neurobiological evidence from songbirds indicates that the functionality of the basal ganglia with respect to production of variation is conserved across even phylogenetically distant relatives. Brainard and Doupe (2000) discovered that lesions of the lateral magnocellular nucleus of the anterior nidopallium (LMAN, an avian cortical-basal ganglia circuit) result in unusual stereotypy in song in male zebra finches. Other studies (Brainard & Doupe, 2001; Kao & Brainard, 2006; Kao, Doupe, & Brainard, 2005) have corroborated the existence of a positive correlation between variation in song and activity of the LMAN in male zebra finches. As a zebra finch male ages, the activity of the LMAN, the variation in its song output, and its ability to modulate its song all decrease (Kao & Brainard, 2006). These findings support the hypothesis that the basal ganglia are critical for the production and modulation of song variation. For example, it allows the male finch to rapidly learn to shift the pitch of its song to avoid a disruptive exogenous stimulus (Tumer & Brainard, 2007). Dopaminergic connections within the circuitry of the basal ganglia are critically important for the modulation of song variability in adult songbirds (Leblois, Wendel, & Perkel, 2010). The adaptability of the control of variability in behavior is not confined to songbirds. Research with mice (Tanimura, Yang, & Lewis, 2008), voles (Garner & Mason, 2002), bears (Vickery & Mason, 2005), and humans (e.g., Neuringer, 2002) corroborates the relationship between variability production and the generation of adaptive action.

### Conclusions

We have presented empirical evidence for a very general and widespread role of outcome expectation in controlling response variation. This has led us to propose a new *Law of Expect*. We also developed and tested a *Modified Law of Effect* that incorporates two processes of inhibition, an inhibitory S-R connection acquired directly to the R when followed by no reward, and mutual inhibition between all possible responses in a particular S. While this *Modified Law* dramatically extended the explanatory capabilities of the original *Law*, it nevertheless cannot account for all of the empirical data. The *Law of Expect*, on the other hand, cannot account for the development of habitual control of behavior by which the learned response becomes insensitive to contingency degradation and outcome devaluation. Thus, the *Modified Law of Effect* and *Law of Expect* together play a fundamental counterbalancing role in learning. The *Law of Expect* codifies the way that S-O associations acquired by the individual while interacting with its environment drive response variation. Stronger S-O associations produce less variation than weaker ones. The resulting variation in response is thereby modulated by prevailing circumstances, placing the individual at an advantageous balancing point along the continuum from exploration to exploitation. Given the fundamental and critical role it plays in behavior during learning, we believe greater attention should be paid to the scientific investigation of the psychological and neural mechanisms that underlie the *Laws of Effect* and *Expect*.

### References

- Antonitis, J. J. (1951). Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of Experimental Psychology*, 42, 273–281.
- Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37, 407-419.





- Balsam, P. D., Deich, J. D., Ohyama, T., & Stokes, P. D. (1998). *Origins of new behavior*. Needham Heights, MA, US: Allyn & Bacon.
- Bartlett, R., Wheat, J., & Robins, M. (2007). Is movement variability important for sports biomechanists? *Sports Biomechanics*, 6, 224–243.
- Blaisdell, A. P. (2008). Cognitive dimension of operant learning. (pp 173-195). In H.L. Roediger, III (Ed.), *Cognitive Psychology of Memory. Vol. 1 of Learning and Memory: A Comprehensive Reference*, 4 vols. (J.Byrne Editor). Oxford: Elsevier.
- Blough, D. S. (1966). The reinforcement of least-frequent interresponse times. *Journal of the Experimental Analysis of Behavior*, 9(5), 581-591.
- Bolles, R. C. (1972). Reinforcement, expectancy, and learning. *Psychological Review*, 79, 394-409.
- Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, 52(10), 976-986.
- Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: Tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 248-265.
- Brainard, M. S., & Doupe, A. J. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature*, 404, 762–766.
- Brainard, M. S., & Doupe, A. J. (2001). Postlearning consolidation of birdsong: Stabilizing effects of age and anterior forebrain lesions. *Journal of Neuroscience*, 21, 2501–2517.
- Brown, P. L., & Jenkins, H. M. (1968). Auto-shaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior*, 11(1), 1-8.
- Burgos, J. E. (2007). Autoshaaping and automaintenance: A neural-network approach. *Journal of the Experimental Analysis of Behavior*, 88(1), 115-130.
- Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 58, 313-323.
- Chan, A. A. Y.-H., Stahlman, W. D., Garlick, D., Fast, C. D., Blumstein, D. T., & Blaisdell, A. P. (2010). Increased amplitude and duration of acoustic stimuli enhance distraction. *Animal Behaviour*, 80, 1075-1079.
- Cherot, C., Jones, A., & Neuringer, A. (1996). Reinforced variability decreases with approach to reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 497–508.
- Colwill, R. M., & Rescorla, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 120-132.
- Colwill, R. M., & Rescorla, R. A. (1986). Associative structures in instrumental learning. In G.H. Bower (ed.), *The psychology of learning and motivation*, Vol. 20, pp. 55-104. San Diego, CA: Academic Press.
- Crespi, L. P. (1942). Quantitative variation of incentive and performance in the white rat. *American Journal of Psychology*, 55, 467-517.



- Cromwell, H. C., & Schultz, W. (2003). Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *Journal of Neurophysiology*, *89*, 2823–2838.
- Darwin, C. D. (1859). *On the origin of species by means of natural selection*. London, UK: Murray.
- Daivids, K., Glazier, P., Araujo, D., & Bartlett, R. (2003). Movement systems as dynamical systems. *Sports Medicine*, *33*(4), 245-260.
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, S., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876-879.
- Dennett, D. C. (1975). Why the law of effect will not go away. *Journal for the Theory of Social Behaviour*, *5*, 169–188.
- 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.
- Elliott, M. H. (1929). The effect of change of reward on the maze performance of rats. *University of California Publication of Psychology*, *4*, 19-30.
- Epstein, R. (1990). Generativity theory and creativity. In M. A. Runco & R. S. Albert (Eds.), *Theories of Creativity* (pp. 116–140). Thousand Oaks, CA: Sage Publications.
- Epstein, R. (2014). On the orderliness of behavioral variability: Insights from generativity theory. *Journal of Contextual Behavioral Science*, *3*(4), 279-290.
- Flaherty, C. F., & Grigson, P. S. (1988). From contrast to reinforcement: Role of response contingency in anticipatory contrast. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(2), 165.
- Frick, F. C., & Miller, G. A. (1951). A statistical description of operant conditioning. *The American Journal of Psychology*, *64*(1), 20-36
- Garner, J. P., & Mason, G. J. (2002). Evidence for a relationship between cage stereotypies and behavioural disinhibition in laboratory rodents. *Behavioural Brain Research*, *136*, 83–92.
- Geisler, F. C. M., Vennewald, N., Kubiak, T., Weber, H. (2010). The impact of heart rate variability on subjective well-being is mediated by emotion regulation. *Personality and Individual Differences*, *49*, 723-728.
- 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.
- Gordus, A., Pokala, N., Levy, S., Flavell, S. W., & Bargmann, C. I. (2015). Feedback from Network States Generates Variability in a Probabilistic Olfactory Circuit. *Cell*, *161*, 215-227.
- Gormezano, I., & Kehoe, E. J. (1975). Classical conditioning: Some methodological-conceptual issues. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 2. Conditioning and behavior theory* (pp. 143–179). Hillsdale, NJ: Erlbaum.



- Guilhardi, P., & Church, R. M. (2006). The pattern of responding after extensive extinction. *Learning & Behavior, 34*, 269–284.
- Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Current Opinion in Neurobiology, 15*, 638–644.
- Grunow, A., & Neuringer, A. (2002). Learning to vary and varying to learn. *Psychonomic Bulletin & Review, 9*, 250–258.
- Hernandez, P. J., Sadeghian, K., & Kelley, A. E. (2002). Early consolidation of instrumental learning requires protein synthesis in the nucleus accumbens. *Nature Neuroscience, 5*, 1327–1331.
- Herrick, R. M., & Bromberger, R. A. (1965). Lever displacement under a variable ratio schedule and subsequent extinction. *Journal of Comparative and Physiological Psychology, 59*, 392–398.
- Herrnstein, R. J., & Loveland, D. H. (1972). Food avoidance in hungry pigeons and other perplexities. *Journal of the Experimental Analysis of Behavior, 18*, 369–383.
- Hursh, S. R., Navarick, D. J., & Fantino, E. (1974). “Automaintenance”: The role of reinforcement. *Journal of the Experimental Analysis of Behavior, 21*, 112–124.
- Jensen, G., Stokes, P. D., Paterniti, A., & Balsam, P. D. (2014). Unexpected downshifts in reward magnitude induce variation in human behavior. *Psychonomic Bulletin & Review, 21*(2), 436–444.
- Kacelnik, A., & Krebs, J. R. (1985). Learning to exploit patchily distributed food. In *Behavioural ecology: Ecological implications of adaptive behaviour*, R. M. Sibly, R. Smith, Eds. (Blackwell Scientific Publications, Oxford, 1985), pp. 189–206.
- Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of Neurophysiology, 96*, 1441–1455.
- Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature, 433*, 638–643.
- Koch, M., Schmid, A., & Schnitzler, H. U. (2000). Role of nucleus accumbens dopamine D1 and D2 receptors in instrumental and Pavlovian paradigms of conditioned reward. *Psychopharmacology, 152*, 67–73.
- Kryptos, A-M., Jahfari, S., vanAst, V. A., Kindt, M., & Forstmann, B. U. (2011). Individual differences in heart rate variability predict the degree of slowing during response inhibition and initiation in the presence of emotional stimuli. *Frontiers in Psychology, 2*, 278.
- Leblois, A., Wendel, B. J., & Perkel, D. J. (2010). Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *Journal of Neuroscience, 30*, 5730–5743.
- Mackintosh, N. J., Little, L., & Lord, J. (1972). Some determinants of behavioral contrast in pigeons and rats. *Learning and Motivation, 3*(2), 148–161.
- Manoel, E. D. J., & Connolly, K. J. (1995). Variability and the development of skilled actions. *International Journal of Psychophysiology, 19*, 129–147.



- Millenson, J. R., & Hurwitz, H. M. B. (1961). Some temporal and sequential properties of behavior during conditioning and extinction. *Journal of the Experimental Analysis of Behavior*, 4, 97–106.
- Modlmeiera, A. P., & Foitzika, S. (2011). Productivity increases with variation in aggression among group members in *Temnothorax* ants. *Behavioral Ecology*, 22, 1026-1032.
- Neuringer, A. (1991). Operant variability and repetition as functions of interresponse time. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 3–12.
- Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, 9, 672–705.
- Neuringer, A. (2004). Reinforced variability in animals and people: Implications for adaptive action. *American Psychologist*, 59, 891–906.
- Neuringer, A., Kornell, N., & Olufs, M. (2001). Stability and variability in extinction. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 79–94.
- Notterman, J. M. (1959). Force emission during bar pressing. *Journal of Experimental Psychology*, 58, 341–347.
- Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 429–452.
- Papini, M. R. (2008). *Comparative psychology: Evolution and development of behavior*, 2<sup>nd</sup> ed. Psychology Press, New York.
- Pasquereau, B., Nadjar, A., Arkadir, D., Bezar, E., Goillandeau, M., Bioulac, B., et al. (2007). Shaping of motor responses by incentive values through the basal ganglia. *Journal of Neuroscience*, 27, 1176–1183.
- Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford, UK: Oxford University 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.
- Pryor, K. W., Haag, R., & O'Reilly, J. (1969). The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior*, 12(4), 653-661.
- Racey, D., Young, M. E., Garlick, D., Pham, J. N-M., & Blaisdell, A. P. (2011). Pigeon and human performance in a multi-armed bandit task in response to changes in variable interval schedules. *Learning & Behavior*, 39, 245-258. PMID: 21380732.
- Rescorla, R. A. (1993). Inhibitory associations between S and R in extinction. *Animal Learning & Behavior*, 21(4), 327-336.
- Reynolds, G. S. (1961). Behavioral contrast. *Journal of the Experimental Analysis of Behavior*, 4(1), 57-71.
- Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of Motor Behavior*, 34 (2), 99-125.
- Roberts, S. (2014). How little we know: Big gaps in psychology and economics. *International Journal of Comparative Psychology*, 27(2), 48-61.



- Ryan, K. M., Blumstein, D. T., Blaisdell, A. P., & Stahlman, W. D. (2012). Stimulus concordance and risk-assessment in hermit crabs (*Coenobita chypeatus*): Implications for attention. *Behavioural Processes*, *91*, 26-29. PMID: 22643163.
- Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology*, *191*, 461-482.
- Schwartz, B. (1982). Interval and ratio reinforcement of a complex sequential operant in pigeons. *Journal of the Experimental Analysis of Behavior*, *37*, 349-357.
- Skinner, B.F. (1953). *Science and human behavior*. New York: Macmillan.
- Stahlman, W. D., & Blaisdell, A. P. (2011a). Reward probability and the variability of foraging behavior in rats. *International Journal of Comparative Psychology*, *24*, 168-176.
- Stahlman, W. D., & Blaisdell, A. P. (2011b). The modulation of response variation by the probability, magnitude, and delay of reinforcement. *Learning and Motivation*, *42*, 221-236.
- Stahlman, W. D., Leising, K. J., Garlick, D., & Blaisdell, A. P. (2013). There is room for conditioning in the creative process: Associative learning and the control of behavioral variability. In A. S. Bristol, J. C. Kaufman, and O. Vartanian (Eds.), *The Neuroscience of Creativity* (pp. 45-67). MIT Press.
- Stahlman, W. D., Roberts, S., & Blaisdell, A. P. (2010a). 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. (2010b). Response variability in pigeons in a Pavlovian task. *Learning & Behavior*, *38*, 111-118.
- Staudinger, M. R., Erk, S., & Walter, H. (2011). Dorsolateral prefrontal cortex modulates striatal reward encoding during reappraisal of reward anticipation. *Cerebral Cortex*, *21*, 2578-2588.
- Stebbins, W. C., & Lanson, R. N. (1962). Response latency as a function of reinforcement schedule. *Journal of the Experimental Analysis of Behavior*, *5*, 299-304.
- Tanimura, Y., Yang, M. C., & Lewis, M. H. (2008). Procedural learning and cognitive flexibility in a mouse model of restricted, repetitive behavior. *Behavioural Brain Research*, *189*, 250-256.
- Thorndike, E. L. (1911) Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monographs Supplement*, *vol. 2, no. 8*.
- Tolman, E. C. (1932). *Purposive behavior in animals and men*. New York: Appleton-Century-Crofts.
- Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of crystallized adult birdsong. *Nature*, *450*, 1240-1244.
- Tunstall, M. J., Oorschot, D. E., Kean, A., & Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology*, *88*, 1263-1269.
- Vickery, S. S., & Mason, G. J. (2005). Stereotypy and perseverative responding in caged bears: Further data and analyses. *Applied Animal Behaviour Science*, *91*, 247-260.



- Watanabe, K., Lauwereyns, J., & Hikosaka, O. (2003). Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. *Journal of Neuroscience*, *23*, 10052–10057.
- Williams, D. R., & Williams, H. (1969). Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior*, *12*(4), 511-520.
- Wiltgen, B. J., Law, M., Ostlund, S., Mayford, M., & Balleine, B. W. (2007). The influence of Pavlovian cues on instrumental performance is mediated by CaMKII activity in the striatum. *European Journal of Neuroscience*, *25*, 2491–2497.
- Winterbauer, N. E., & Bouton, M. E. (2010). Mechanisms of resurgence of an extinguished instrumental behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(3), 343-353.
- Wu, H. G., Miyamoto, Y. R., Castro, L. N. C., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience*, *17* (2), 312-321.
- Wyvell, C. L., & Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward “wanting” without enhanced “liking” or response reinforcement. *Journal of Neuroscience*, *20*, 8122–8130.
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews: Neuroscience*, *7*, 464–476.
- Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behavioural Brain Research*, *166*, 189–196.