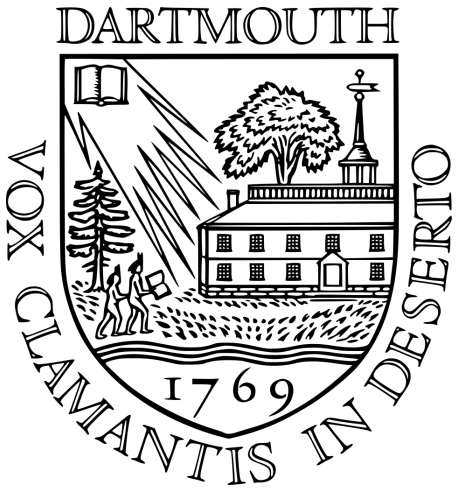
Towards an identification of state shifts in rodent CA1



Roman Huszár

Advisors: Matthijs van der Meer, PhD; David J. Bucci, PhD

Neuroscience Honors Thesis
May 30, 2017
Department of Psychological and Brain Sciences
Dartmouth College

Table of Contents

Acknowledgements

The following pages are a result of the most intense and rewarding challenge I took on during my four years at Dartmouth. The project I pursued was a collaborative effort between the labs of Dr. Bucci and Dr. van der Meer. I am immensely grateful for being able to take part in this endeavor, as it allowed me to think deeply about theories of animal behavior, as well as develop hands-on skills in electrophysiology. I find it hard to express how crucial this amalgamation of approaches has been to my development as a young researcher. I am convinced that the conceptual framework I operated under while working on my thesis will leave an indelible trace in how I think about the brain going forward.

I would like to sincerely thank both Dr. Bucci and Dr. van der Meer for their invaluable mentoring. I am especially grateful to them for allowing me to operate as independently as I did at the intersection of their labs. I would also like to express my gratitude to Dr. Todd for his illuminating insights into animal behavior and experimental design.

Furthermore, I would like to thank all members of the van der Meer lab for helping me with the research process. I am endlessly grateful to Emily Irvine for helping with behavioral experiments, to Jimmy Gmaz for all the drive building tips, to Youki Tanaka for assisting in surgery, and to Eric Carmichael for advice regarding all aspects of the process. I could not have succeeded without their presence.

Lastly, my sincerest thanks go to everyone outside the lab who provided moral support, and especially to my parents for understanding my absences in virtual communication. My deepest gratitude also goes to Yvonne Fang – your constant presence made the process considerably more joyful than it would have been otherwise.

Abstract

Behavior is constantly modulated by ever-changing constellations of cues that determine the behaviorally relevant state. For example, a vehicle approaching the crosswalk signals a shifted state in which it is no longer appropriate to cross the road. To date, very little is known about the neural substrates underlying state shifts. This is surprising given how important state identification is to exhibiting contextually appropriate behavior. First, we review the existing literature, and point to models that provide an especially useful conceptual framework for dealing with this problem. Second, we report the results of three experiments in which we set out to model state shifts in rats by employing two related tasks: serial biconditional discrimination (BD), and negative occasion setting (NOS) (both Pavlovian, and involving discrete cues). In order to successfully predict trial reinforcement value in the former task, animals had to be sensitive to the following cue combinations: V1—>A1+, V2—>A1-, V1—>A2-, and V2—>A2+, where '+' denotes reward, and '-' denotes no reward. We hypothesized a state shift in representing A1 (and A2) depending on the identity of the cue preceding it. In Experiment 1, trials were presented in pseudo-randomized order, and in Experiment 2, trial ordering was blocked. Strikingly, neither experiment promoted successful acquisition of biconditional discrimination. As such, we shifted to a NOS task (Holland et al., 1999) that included a conditional discrimination between L—>T- and T+ trials. Surprisingly, this established design was also not learned by the group. The implications of

these findings are discussed. At the outset, we intended to carry out in-vivo electrophysiological recordings from CA1 while rats performed the task in order to characterize ensemble firing of putative state shifts.

Part 1: Towards a new theory of occasion setting

The environment is replete with ambiguous events whose meaning requires moment-to-moment disambiguation. This presents a high-stakes challenge to behaving organisms, given that deciphering an event's current meaning often makes the difference between life and death. In the literature, this general problem has been illustrated through various examples. For example, a crosswalk is generally thought of as a safety cue signaling where the road can be traversed. However, we still look left and right before crossing, because we understand that the crosswalk's safety signaling property is cancelled when there is a car bolting towards it (Meyer & Bucci, 2016a). Another related example concerns the multiple meanings of individual words - for instance, when "Fire!" is cried out in the shooting gallery, it holds a meaning that is very different from when it is cried out in a bank (Bouton, 1994). Each of these anecdotal examples involves an ambiguous event that needs to be resolved by paying attention to other relevant factors. Crucially, the examples differ in terms of what these factors are; in the former case, a discrete stimulus (i.e., a moving car) mediates the resolution, while in the latter case, a complex stimulus configuration fulfills this role (i.e., a physical context such as the interior of a bank). Note, the notion of context is one that is highly relevant to the discussion of ambiguity resolution; in order to emit the appropriate response when confronted with an ambiguous stimulus, the agent must first identify the current context, or state. Framed in these terms, context simply acts as a disambiguator. Overall, there is an

extensive literature on the role of context, and a detailed treatment of the topic is beyond the scope of this review (see Bouton, 1994, 2004; O'Keefe & Nadel, 1978; Nadel & Willner, 1980; Rudy, 2009). For sake of consistency, we will refer to the disambiguating factor as the "feature", and to the ambiguous event as the "target" (Bouton, 2007). Overall, it is clear that situations involving ambiguity resolution pose a nontrivial problem - the brain needs to identify the relevant

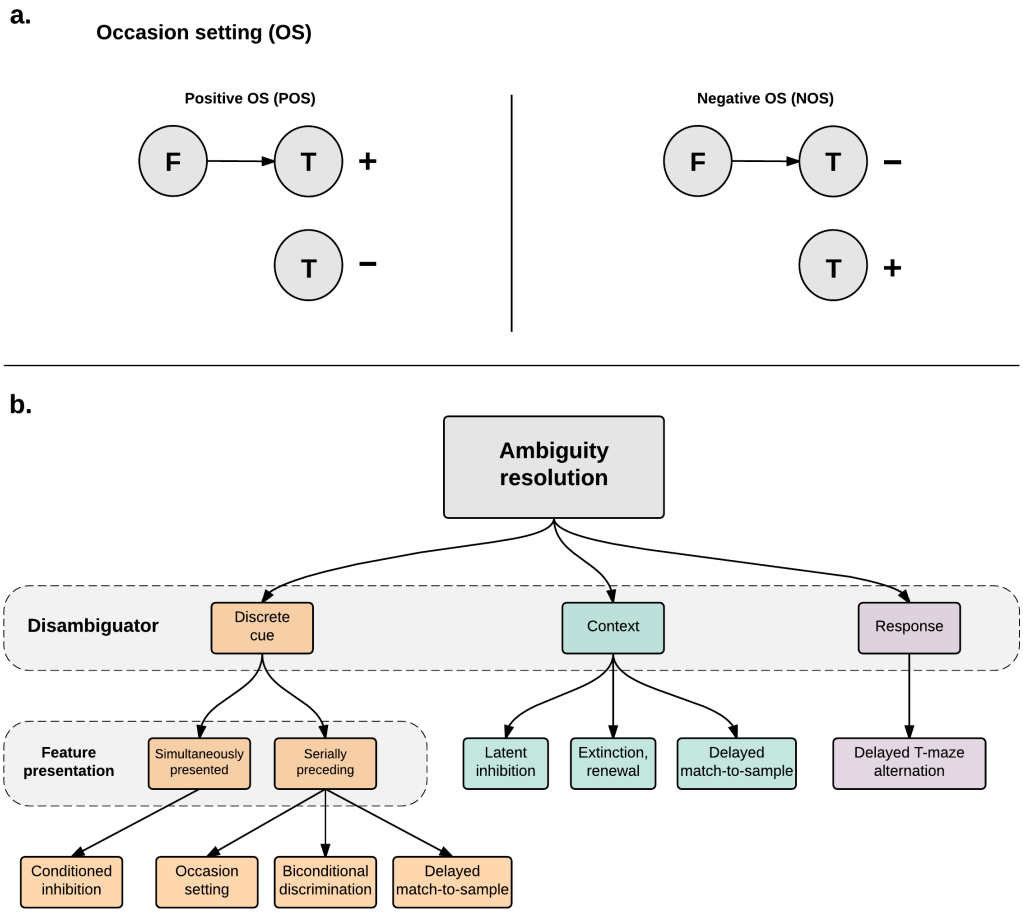


Figure 1: **(a)** Positive occasion setting (left) and negative occasion setting (right). "F" denotes the feature cue, "T" is the target, and "+" is reward ("-" is its omission). **(b)** Tree diagram describing relationships between tasks involving ambiguity resolution. Tasks are separated into three categories depending on whether the feature is discrete, contextual, or a response. Within the discrete-feature category, tasks are further categorized by whether the feature serially precedes the target or is presented with it simultaneously.

feature, and use it to represent a neural difference across otherwise perceptually identical instances of the target.

In this thesis, we review tasks that involve ambiguity resolution with special emphasis on occasion setting (see Figure 1a). First, we describe the computational problem underlying ambiguity resolution as it is embodied in OS. Next, we review existing models of OS, and discuss models that have yet to be fully explored in this research domain. More specifically, we propose a number of process models, each of which provides a unique algorithmic account of the phases a system steps through in order to solve OS. Lastly, we review the neurobiological findings pertaining to ambiguity resolution, and discuss how these might (or might not) fit within the frameworks of the different models.

Theory of ambiguity resolution

1. Occasion setting and ambiguity resolution

In OS, subjects learn to identify the meaning of a target cue (T) depending on whether a feature cue (F) precedes it or not. Tasks that involve OS therefore consist of two trial types: one in which the target is presented alone (T trials), and one in which the target is serially preceded by the feature (F—>T trials). It is clear that OS involves ambiguity resolution, given that the meaning of T requires trial-by-trial disambiguation - to this end, animals must learn to use the presence or absence of F as a signal to inform the present meaning of T.

In positive occasion setting (POS), F—>T trials result in reward, while T trials result in no reward - in this case, the feature is referred to as a positive occasion setter, since it "sets the occasion" for the target's reward prediction value. Conversely, in negative occasion setting (NOS), F—>T trials result in no reward, while T trials result in reward - here, the feature is referred to as a negative occasion setter, as it cancels the target's reward prediction value. The arrow symbol ("—>") between F and T represents the fact that OS paradigms typically include a temporal gap (i.e., an inter-stimulus interval, or ISI) between the feature and the target - indeed, the presence of the ISI has proven to be a crucial factor for promoting the acquisition of OS as opposed to other forms of learning - e.g., conditioned inhibition (Rescorla, 1969) - that are learned when F offset and T onset are temporally contiguous, or when F and T are presented simultaneously.

Overall, OS has received a great deal of attention from learning theorists, as occasion setters hold properties that make them distinct from regular conditioned stimuli (CSs). First, it has been demonstrated that CSs with positive reward predictive value (i.e., exciters) often come to elicit a CS-specific conditioned response. In rats, for instance, auditory exciters elicit head jerking, while visual exciters elicit rearing (Holland, 1992). In OS, the feature is only used in order to inform the current value of the target; thus, any conditioned responding is specific to the modality of the target, not the feature. Furthermore, occasion setters are immune to counterconditioning. For instance, when a

negative occasion setter is repeatedly paired with reward, its capacity to act as a negative occasion setter goes largely unaffected (Holland, 1991, 1992). Additionally, an occasion setter's properties are specific to its target, and do not transfer well to cues that have not been trained as targets of other occasion setters (i.e., non-target cues). Interestingly, this transfer effect is weakened when the non-target cue is associated with the same reinforcer as the target (reinforcer-specific transfer). Lastly, there is a remarkable difference between NOS and POS that stems from the amount of training that is required until acquiring a robust discrimination between the two trial types - it turns out that NOS is acquired significantly slower than POS (see Figure 2). This has often been referred to as the feature-positive effect (Jenkins & Sainsbury, 1970), and

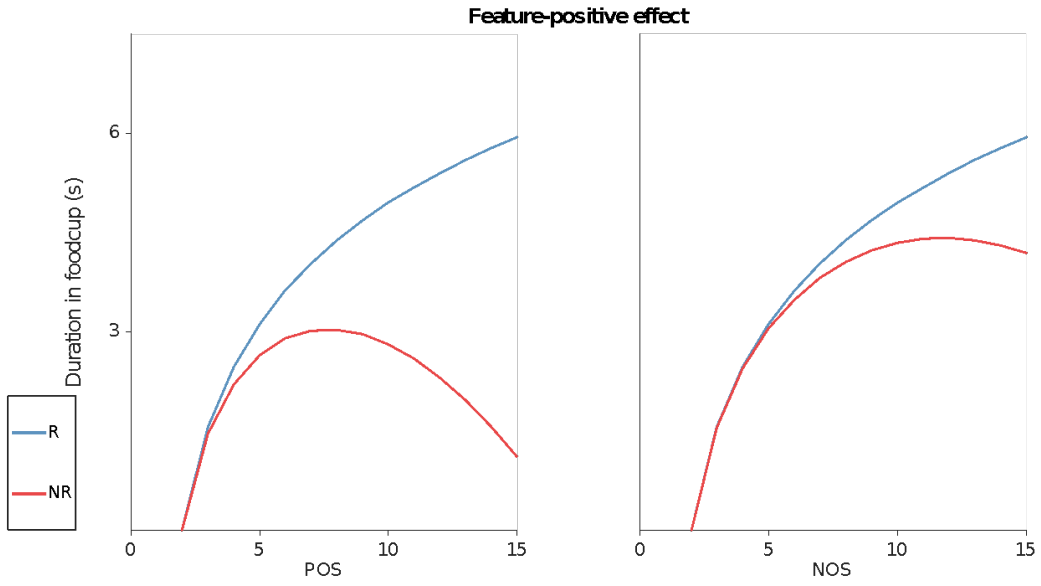


Figure 2: Feature-positive effect. R denotes the rewarded trial type, and NR the unrewarded trial type. The simulated acquisition curves demonstrate the feature-positive effect, as in the case of POS, discrimination between R and NR trials occurs earlier than in NOS. The diagram is based on data in Holland et al., (1999), and Bouton & Hendrix (2011).

has typically been observed in behavioral paradigms involving a simultaneous presentation of the feature and the target (i.e., a FT compound). More specifically, it has been demonstrated in animal models (González, Quinn, & Fanselow, 2003) and in humans (Wheeler, Amundson, & Miller, 2006) that when the FT compound is first paired with reward, and then T is presented individually in extinction, the decrement in responding to T is significantly greater than when the reverse scenario takes place - i.e., when T is first reinforced, and then tested as compound FT. The feature-positive effect has often been described as an asymmetry in stimulus generalization between discrete cues and compounds. It is plausible that in the case of OS, the feature-positive effect is explicable in similar terms. In any case, any realistic account of OS must be able to make sense out of this phenomenon, as it likely reflects an important difference in the neural processes that govern the acquisition of POS as opposed to NOS. Overall, we propose the following four criteria a model of OS must satisfy with respect to the distinct properties of occasion setters:

1. Response form is target-specific.
2. Occasion setters are immune to counterconditioning.
3. The properties of an occasion setter do not transfer well to a non-target cue, except when the non-target predicts the same reinforcer as the target (reinforcer-specific transfer).
4. NOS is acquired significantly slower than POS (feature-positive effect).

Indeed, a large number of behavioral tasks apart from OS have modeled ambiguity resolution. Figure 1b displays the tasks as categorized by identity of the feature (e.g., discrete cue, context, response), and by whether the feature serially precedes or co-occurs with the target (see Figure 1b). It is interesting to note that some of the tasks were not originally designed to tackle ambiguity resolution. For example, the delayed match-to-sample design was employed to study working memory (Miller, Erickson, & Desimone, 1996). This said, the task clearly involves trial-by-trial ambiguity - the design involves instances of a target cue whose immediate meaning depends on whether a previously presented feature is identical to it or not. Before delving into the specifics of each task, however, it is useful to have an algorithmic description of the steps that must take place for the brain to resolve ambiguities successfully. To this end, the OS literature is especially informative, as it involves models that deal with this general problem. Overall, OS holds a unique position in the task hierarchy depicted in Figure 1b. First, it is arguably the simplest instantiation of ambiguity resolution, as both the target and the feature are discrete cues, as opposed to more complex stimulus configurations. Furthermore, the discreteness of the feature implies that it can be presented at will; this grants the experimenter unique control over the current meaning of the target. Indeed, this is particularly useful for studying the neural substrates underlying the different meanings of the target across rewarded and unrewarded OS trial types.

2. Three accounts of OS

Overall, the unique properties of occasion setters have made them a challenge to standard models of associative learning, such as the Rescorla-Wagner model (Rescorla & Wagner, 1972). The model developed by Rescorla and Wagner details the process by which environmental stimuli come to predict the occurrence (or non-occurrence) of behaviorally salient events (e.g., reward, footshock). These events are thought to support a threshold of conditioning (λ) that reflects their surprisingness. When confronted with such an event, the model assumes that all neutral cues (e.g., lights, tones, etc.) present at the time of the event come to accrue 'associative strength' (parameter V in the model) that decrements the surprisingness of the event - in other words, the neutral cues become predictive of the upcoming event by virtue of their non-zero V. For each cue, the change in associative strength occurs on a trial-by-trial basis, and is proportional to the difference between the surprisingness of the event, and the sum of the associative strengths of all cues present ($\lambda - \sum V$). Given that Rescorla-Wagner does not model temporal dynamics, it has no way of using the discontiguous occasion setter to acquire a discrimination between F—>T and T trials. As such, the model predicts that F remains a neutral cue, while each rewarded T accrues positive V, and each unrewarded T accrues negative V. Clearly, the discrimination is never achieved.

This challenge has precipitated the development of novel accounts aiming to explain the properties of occasion setters, as well as reconcile them with associative accounts of classical conditioning (see Figure 3). For instance,

hierarchical theory claims that the occasion setter comes to modulate the specific association between the target and reward (Figure 3a). A specific instantiation of of this idea is given by Bouton & Nelson (1998); in NOS, for example, the target is thought to develop a simple excitatory association with reward on T+ trials. On F—>T- trials, however, the target forms an inhibitory association with reward, such that this association is only active in the presence of the feature. In this manner, the feature comes to hierarchically gate the active association between the target and reward (Bouton & Nelson, 1998). Indeed, the specificity of this modulation is consistent with immunity to counterconditioning and transfer effects; In the case of counterconditioning, it has been argued that altering the feature's associative relationship with the reinforcer does not change its relationship with the target-US associative link ('US' denotes unconditioned

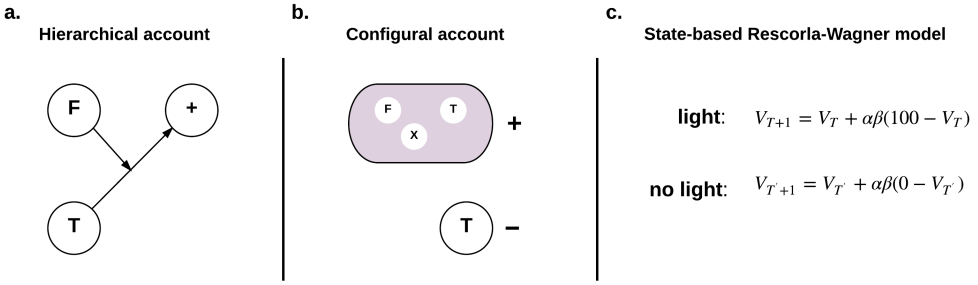


Figure 3: Three accounts of OS. **(a)** The hierarchical account of OS posits that the feature (F) gates the target's (T) association with reward (+). **(b)** The configural account assumes that all cues present on a given trial are encoded as a configural stimulus comprising of cue-specific and configure-specific components; for example, on an F—>T+ trial, cues are represented as a configure consisting of F-specific, and T-specific components, as well as component X that is unique to the configure. **(c)** A state-based Rescorla-Wagner model in which the presence or absence of the feature determines which target-outcome association (T or T') is currently active.

stimulus, which is reward in this case). On the other hand, reinforcement-specific transfer can be explained as a generalization effect that stems from the similarity of associations that involve the same reinforcer; more specifically, if F acts on a T-sucrose associative link (where T was trained as a target of F), then the occasion setting properties of F will likely generalize to a C-sucrose link, even though C was never trained as a target of F - this is due to a similarity between the T-sucrose and C-sucrose associations by virtue of their shared reinforcer.

In contrast to the hierarchical account stands configural theory, which assumes that on F—>T trials, the feature and the target are encoded together as a configural cue (Figure 3b). This configure is composed of feature-specific elements, target-specific elements, and elements X that are unique to the configure (Rescorla, 1972). Indeed, configural theory is attractive as it makes assumptions about cue encoding that make OS tractable for associative accounts of conditioning - the configural cue is merely a composition of elements that enter into associative relations with behaviorally significant events. In this sense, a configural cue is just like any other neutral cue. In the specific case of OS, the configural account suggests that any responding to the configure on F—>T trials by virtue of its similarity to T trials is counteracted by building associative strength to X. Notice that under the configural assumption, OS can easily be solved by the Rescorla-Wagner model. In the case of the NOS paradigm, the negative associative strength V accrued on F—>T- trials is split equally among all elements of the configure, while the positive V accrued on T-US trials is gained

uniquely by the target. As such, on F—>T- trials the net-positive V accrued to the target is counteracted by the negative V accrued to non-target specific elements (e.g., X) of the configure. Therefore, the model correctly predicts the differential responding in NOS. Furthermore, the model can account for the feature-positive effect, which would be explained as an asymmetrical generalization effect. In the case of POS, positive V would be split equally amongst elements of the F—>T configure; as such, the small positive V accrued to T would be quickly counteracted on T- trials. On the other hand, in NOS, the large positive V accrued to T would take more time to counteract on F—>T trials, thereby resulting in a delay in acquiring the discrimination. Note, however, that under the assumption of configural cues, the Rescorla-Wagner model mispredicts that the feature develops a direct associative relationship with reward. This is a major imperfection of this rendition of the model. Overall, configural accounts struggle to explain some of the more nuanced phenomena that distinguish OS from other forms of learning. For instance, configural theory fails to explain reinforcement-specific transfer. This is the case because the account only assumes generalization by virtue of cue similarity (e.g., responding to an AX configure might generalize to a BX configure by virtue of the shared 'X' elements), but not by virtue of shared reinforcer identity. For this reason, hierarchical accounts have often been deemed to better reflect the process by which OS is learned (Bonardi et al., 2017).

It is notable that the OS-specific deficits in the Rescorla-Wagner model disappear as soon as configural cues are assumed. In fact, this is not the only assumption that allows the model to solve OS. For instance, let us assume that on T-alone trials, the target holds a completely independent set of associative relations from those it holds on F—>T trials (Figure 3c). Moreover, let us assume that only one set can be effective at any given time, and that the identity of the active set is controlled by the presence or absence of the feature. Under this assumption, F acts as a disambiguating factor by defining the appropriate set of target-related associations on a trial-by-trial basis. Each set of associations can be thought of as the 'active state' currently occupied by the target. Clearly, solving OS would be trivial for a state-based version of the Rescorla-Wagner model: the positive associative strength V accrued to T in the rewarded trial type would be distinct from the negative V gained by T in the unrewarded trial type. Note that the state-based Rescorla-Wagner model can be framed as a simplified version of the temporal difference reinforcement-learning (TDRL) algorithm - it is only composed of two states, such that each possesses a distinct value reflecting the signed magnitude of the associative strength accrued to the target. The disadvantages of this model are identical to those of the TDRL algorithm (discussed in the following section).

As discussed above, each OS-specific account involves a tradeoff between aspects that make it consistent with OS along certain dimension, but not along others. For example, the hierarchical account can explain

counterconditioning and transfer effects (Bonardi et al., 2017), but it cannot explain the feature-positive effect. On the other hand, certain variations of the configural account are well poised to explain the feature-positive effect, but not certain transfer effects (Bonardi et al., 2017). Interestingly, a disadvantage common to each of these accounts is their OS-specificity, which restricts explanatory power to OS, a particular instantiation of the general problem of ambiguity resolution. The following section discusses computational and statistical models that are more broadly applicable to problems involving ambiguity resolution.

3. Computational models of OS

This section introduces a number of computational models that have yet to be fully explored as potential explanatory frameworks for OS, and ambiguity resolution more broadly. Each model is explored with the intention to give the reader an intuitive understanding of how it might go about solving OS; however, such a level of description necessarily leaves out a great deal of the mathematical and statistical underpinnings that underlie each model's inner workings – for these details, we refer the reader to the publications cited throughout the text.

Overall, the unique problems associated with the elemental[1] and configural theories of OS invite a unified model exploiting the advantages of

[1] Both the hierarchical account of OS and the Rescorla-Wagner model are elemental theories, since they assume cues are encoded as discrete units, or elements.

each, whilst avoiding their respective pitfalls. This account amalgamation was achieved in the SLH connectionist model (Schmajuk, Lamoureux, & Holland, 1998). Overall, the model architecture is that of a connectionist network comprising layers of units that represent relevant events (e.g., features, targets, configures, etc.), and edges between vertices that hold associative strengths (Figure 4). Conditioned stimuli (contexts and discrete cues) are represented at the input-unit layer, while configural stimuli are represented at the hidden-unit layer. Each input-unit forms a direct association with a corresponding output-unit (e.g., VS_1 for input unit A), as well as with every unit in the hidden layer (e.g.,

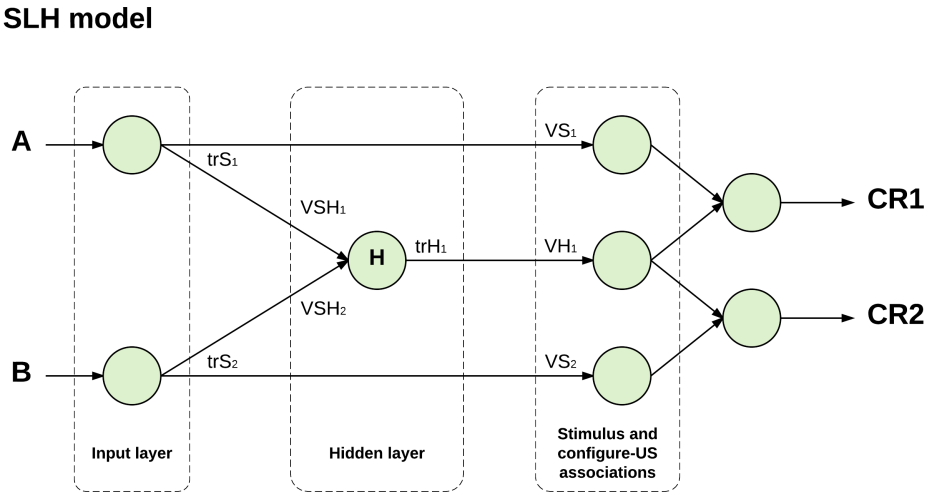


Figure 4: The SLH connectionist model comprises an input layer, a hidden layer, and an output layer. A, B = conditioned stimuli; trS = trace memory of a stimulus; H = hidden-unit representing a configuration of A and B; VSH = association between conditioned stimulus and hidden unit; trH = trace memory of a hidden unit; VS = direct association between conditioned stimulus and outcome; VH = direct association between hidden unit and outcome; CR = conditioned response.

VSH_1 for input unit A). Hidden units also form associations with the output layer (e.g., VH_1 for hidden unit H), which means that each CS forms a direct association with the output layer, as well as an indirect one via the hidden layer. Note that each unit holds an additional trace memory parameter - trS for input units, and trH for hidden units - which reflects the activation level of the corresponding unit. In other words, when the unit is activated, its trace parameter value goes up, and decays over time. As suggested above, associations are updated according to the Rescorla-Wagner rule, with the exception that change in associative strength is additionally weighted by the trace parameter: $\lambda - B$, where $B = \sum VS_i trS_i + \sum VH_j trH_j$, the overall US expectancy. Crucially important is the fact that the subset of output units associated with a given input (e.g., VS_1 and VH_1 for input-unit A) feeds into a final unit that defines the conditioned response. This allows the quantification of CS-specific response form, a crucial asset for determining the nature of underlying learning upon simulation of OS.[2] In the original paper (Schmajuk, Holland, & Lamoureux, 1998), the authors demonstrate that SLH can simulate POS and NOS in manner that is consistent with appropriate response form, counterconditioning phenomena, and transfer effects - qualities that are unique to occasion setters.

Despite these strong points, the model does not seem to account for the feature-positive effect; while this phenomenon is never explicitly mentioned in the text, the simulated acquisition curves for POS and NOS do not suggest any

[2] This aspect is an extension to a previous version of the same model; see Schmajuk and DiCarlo, 1992.

differences in acquiring the former over the latter (see Figures 15 and 18 in Schmajuk, Holland, & Lamoureux, 1998). Overall, the connectionist model substitutes the problem of acquiring task structure for that of acquiring appropriate parameter weights; while this approach is effective for learning about complex task spaces, it has been argued that it leads to data overfitting, detectable through cross-validation on a separate data set (Fuhs & Touretzky, 2007). In addition, the model lacks prior beliefs about the environment, which serve to constrain what can be learned (Gershman, Blei, & Niv, 2010; Kemp & Tenenbaum, 2008).

As mentioned above, the SLH model can reproduce a large number of phenomena that distinguish OS from other forms of learning. This is achieved by having CSs affect outcome expectancy via direct CS-US associations, as well as via indirect associations that pass through a hidden-unit layer. The hidden layer supports stimulus configuration, which is how occasion setters come to affect responding. Despite the model's strengths, however, it does not explain the feature-positive effect. Furthermore, the model's "learning of weights" approach makes it liable to overfitting, and detracts from its potential to act as a model for ambiguity resolution more generally. To this end, it is important to note that OS-specific models are not the only models capable of explaining OS.

Interestingly, OS is explicable in terms of temporal difference reinforcement learning (TDRL) (Sutton & Barto, 1998; Maia, 2009), a learning algorithm with impressive scope and generality. TDRL assumes that the

environment can be parsed into distinct states, such that each state is associated with a set of possible actions and rewards. Performing a state-related action may lead to a state transition, and associated reinforcement. These reinforcement values are then propagated back in time, and become associated with states in the past in order to signal the amount of reinforcement that is reachable from them. Overall, TDRL is an algorithm aimed at identifying sequences of state-related actions that reap the greatest expected reinforcement. Assuming that F—>T and T trials can be encoded as separate states, TDRL could easily solve OS by learning that T signals reward in one state, but not in the other. If the algorithm's parameters were set up so that F would only serve to toggle the correct state, its properties would likely be consistent with those of an occasion setter. Indeed, by reframing OS as a problem involving the acquisition of state-dependent event relations, TDRL may be better positioned to explain ambiguity resolution more broadly - for example, the same process can be applied in order to give an algorithmic description of the delayed match-to-sample task.

Despite TDRL's generality, the default model involves the significant setback of assuming that the state space is given. The major challenge of OS is to identify the correct states; as already sketched out in the state-based Rescorla-Wagner model, learning the appropriate action values is a rather trivial problem once the states are known. Thus, unless it specifies how the correct states are inferred, TDRL cannot be a viable explanation of OS. As is becoming increasingly clear, the problem of state learning is highly relevant to the

discussion of OS. In order to begin to address this problem, we review statistical models that have yet to be explored in the domains of OS and ambiguity resolution more broadly.

Courville, Daw, & Touretzky (2006) present a generative Bayesian model that reframes conditioning paradigms as inferential statistics problems. According to the model, animals assume that events in the environment arise from unobservable (or latent) causes that can be inferred based on specific patterns in the animal's experience. To illustrate this idea, consider the example of serial reversal learning; the task involves two responses (R1 and R2), such that in the initial phase, emitting R1 is rewarded, while emitting R2 is punished. At some point into the task, however, the experimenter reverses the values of R1 and R2. Clearly, there is a causal structure to the task, in the sense that each experimental phase is causally linked to a specific pattern of observations. The model discussed by Courville, Daw, & Touretzky (2006) assumes that animals aim to reconstruct this causal structure. The inferred causes are latent (i.e., unobservable, or hidden), however, since animals never possess explicit knowledge regarding the experimental phase they are in. Broadly speaking, latent causes play a role similar to that of state representations in temporal difference reinforcement learning models.

Before interacting with the environment, the agent maintains a naive priori model that consists of latent causes (C1 and C2) linked to specific environmental observations (T, +, and F; abbreviations are explained further in the text). Given

an active cause (e.g., C1) and environmental observation (e.g., T), the weights **w** specify the probability of making the observation, P(T | **w**,C1) (see Figure 5). Parameters **w** delineate the animal's prior model of the environment, and effectively constrain what it can learn; the animal could never learn about observation X, for example, since X is not included in its prior model. Overall, the goal is to estimate parameters **w** so as to maximize the likelihood of making correct predictions regarding observations in the environment - this corresponds to the computation of posterior P(**w** | observations), which is achieved through Bayes' theorem.[3]

It can be shown that this model can acquire the stimulus contingencies specific to trial types in OS. For example, if a reward (+) is consistently delivered

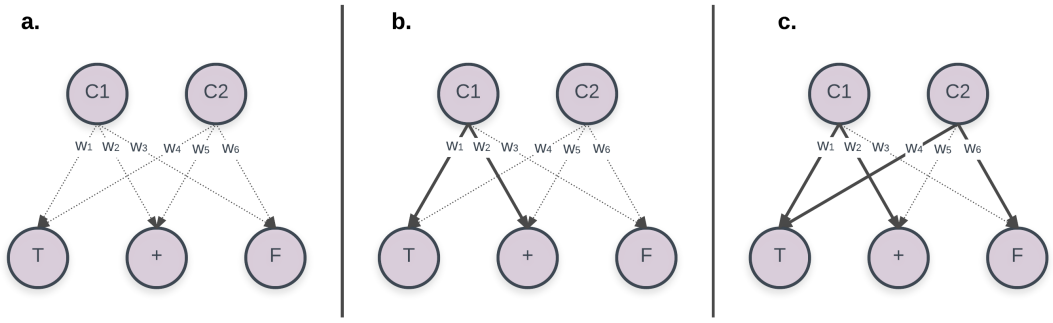


Figure 5: Latent causes C1 and C2 are assumed to generate certain environmental observations - namely, observations of the feature (F), the target (T), and reward (+); given an active cause, weight w gives the likelihood of making the associated observation. In general, OS is solved by associating the individual trial types with distinct latent causes. **(a)** Prior assumptions of the model are given by pre-existing connections between causes and events. **(b)** Upon experiencing a T+ trial, cause C1 is inferred by strengthening the appropriate weights. **(c)** L—>T- observations become linked with latent cause C2.

[3] The following is a statement of Bayes' theorem that has been adapted for the purpose estimating parameters **w** in the context of the model discussed in the text:

$$P(w|observations) = \frac{P(observations|w)P(w)}{P(observations)}$$

whenever a target tone (T) sets off (i.e., a T+ trial), weights w_1 and w_2 will be adjusted, thereby reflecting the knowledge that tone predicts reward (Figure 5a and 5b). This effectively corresponds to the inference of latent cause C1, since the correlation of observations T and + is causally attributed to C1 by virtue of the adjusted parameters **w**.

It is relevant to consider what would happen if something in the environment changed. For example, if all of a sudden the animal received a feature light cue (F) followed by the familiar T, but this time resulting in no reward (i.e., a F—>T- trial, exactly as in NOS), the model should ideally detect this novelty, and update beliefs about cue relations appropriately. To account for this situation, Courville, Daw, & Touretzky (2006) designed the model to make trial-by-trial inferences as to whether parameters **w** have changed. As such, upon experiencing a F—>T- trial, the model would infer a likely change in parameters **w** due to the low value of the prior P(observation I **w**). This would result in a Bayesian update of the posterior, thereby revising beliefs about the structure of the world. More specifically, the model would learn event relations on F—>T- trials by strengthening weights w_4 and w_6 on the links between a previously irrelevant cause C2, and the appropriate events. This learning update is reflected in Figure 5 in the difference between parts b and c. Clearly, the model presented by C Courville, Daw, & Touretzky (2006) can handle occasion setting - inferring C1 or C2 corresponds to an expectancy of a T+ or a L—>T- trial, respectively.

Despite this success, the model involves a number of disadvantages that are worth pointing out. Firstly, the model is only effective in learning to predict observations insofar as these are incorporated in its priors; clearly, this assumption becomes unfeasible in hyperdimensional, real-world situations composed of innumerable events (see Gershman, Cohen, & Niv, 2010). A further complication is the assumption of a finite parameter space **w** - the number of latent causes included is essentially stationary, and so any novelty in the environment must be modeled as a change in weights on existing causal links. This is a significant limitation, given that the model constrains the animal's experiences to a finite number of observations.

In summary, the model proposed by Courville, Daw, & Touretzky (2006) provides a useful framework for delineating how agents might maintain a

Box 1: The Dirichlet process

We briefly discuss the Chinese Restaurant metaphor (which illustrates the Dirichlet process), as it gives insight into the parameters that affect the inference of novel causes. Imagine a restaurant with an infinite number of tables, such that each of the tables has an infinite number of seats. The overall challenge is to assign customers (observations) to specific tables (latent causes). Note that this situation is analogous to an infinite parameter space in that it is possible to assign customers to a hypothetically infinite number of different tables. Let n-1 be the number of customers that have entered the restaurant, and let k be the number of distinct tables they have been assigned to. The likelihood that the n-th customer will be seated at a given table is governed by the following random process: the probability of being assigned to an occupied table t is proportional to the number of customers seated at that table, and the probability of being assigned to a so-far unoccupied table t' is proportional to a parameter α. The value of α defines how biased the model is to state splitting. When $\alpha = 0$, the probability of assigning the n-th customer to an unoccupied table t' is 0; in other words, the model never infers a novel latent cause. On the other hand, when $\alpha = \infty$, the likelihood of seating a customer at t' is 1; in this case, the model infers a novel cause for every new observation. Clearly, neither of these α values are optimal, and a value such that $0 < \alpha < \infty$ is desired - for the purpose of simulations, Gershman, Blei, & Niv (2010) set $\alpha = 0.1$. In this case, the model prefers a small number of latent causes, but expands by inferring novel causes whenever an unfamiliar observation is made.

decomposition of its environment into relevant states; while the model discusses latent causes, rather than states, these serve effectively the same purpose in that both serve to cluster meaningful events in order to maximize the predictability of environmental occurrences. The acquisition of states is effectively achieved by updating the weights on a priori links between latent causes and specific events. As discussed above, the model cannot infer a greater number of latent causes than are already incorporated in its prior assumptions – this is a considerable limitation that constrains the agent's experience, thereby limiting the model's generality.

This particular problem is addressed in a variant of the generative model proposed by Gershman, Blei, & Niv (2010). The authors present a normative Bayesian statistical framework that involves generative and inferential models designed to tackle conditioning phenomena such as renewal and latent inhibition. The model was inspired by a previously described TDRL algorithm that had additionally been designed to carry out state splitting (Redish et al., 2007). A crucial aspect of the Gershman, Blei, & Niv (2010) model is that it involves an infinite parameter space - given an observation, the model classifies it into a cluster of observations that corresponds to a common latent cause. In case the observation is not easily categorized into existing clusters, it is classified into its own separate cluster, thus corresponding to a novel latent cause. This procedure is formally known as the Dirichlet process, and has often been illustrated by a metaphor dubbed as the Chinese Restaurant process (see Box 1). Overall, this

allows the model to account for the large number of distinct observations that animals encounter.

While the Bayesian generative model can easily acquire occasion setting and related context-dependent phenomena (renewal, latent inhibition, etc.), it does not provide a complete account of OS. Especially relevant to our discussion is the acquisition of NOS - the Gershman, Blei, & Niv (2010) model solves OS by inferring a distinct latent cause for each trial type. However, given that it assumes no differences in the acquisition of latent causes based on the contents of trial observations (i.e., the salient difference between NOS and POS), it is not equipped to explain the feature-positive effect. Interestingly, the model does predict differences in the rate of learning depending on the number of latent causes that are known to it at the start of acquisition;[4] when starting out with a single cause, the situation would be as described in the Chinese Restaurant metaphor (see Box 1); acquiring POS would occur at the same rate as NOS, however, since the inference of a novel cause is proportional to α (see Box 1) irrespective of the OS paradigm. On the other hand, provided the model starts out with more than one cause, acquisition ought to be swifter. This is due to an aspect of the generative model, which states that for each observation, a known cause is sampled according to a mixing distribution P;[5] thus, when starting with

[4] This is assuming that in both of the to-be-described situations, α remains constant and between 0 and ∞.

[5] Note, this occurs regardless of the number of known causes at the start of acquisition. Upon making an observation, the model samples a known cause c; whether the model attributes the observation at hand to cause c or a completely novel cause is then subject to the Dirichlet process (see text).

more than one cause, the distinct OS trial types would be probabilistically attributed to more than one cause since the start of acquisition. While this would make acquisition faster, clearly it would hold for NOS as much as it would for POS. Surely, the feature-positive effect could be obtained if the model assumed a single cause at the start of NOS, and more than one cause at the start of POS; however, such an assumption imputes a certain degree of clairvoyance to the agent, and is therefore unlikely to shed any light on the neural mechanisms that behaving animals realistically employ in order to solve occasion setting.

An additional downside is that the model involves significant computational costs. Clearly, encoding the appropriate stimulus-reward relations across OS trial types necessitates the formation of separate latent causes (one per trial type) for situations that are perceptually very much alike. Indeed, the only differences across trial types in occasion setting are the presence (or absence) of a discrete cue (i.e., the feature), and the presence (or absence) of reinforcement. It is without a doubt that these differences are key, and must be included in any model that solves the task. Nonetheless, representing the differences by generating a novel latent cause comes at the cost of increased model complexity. If each latent cause is associated with a set number of parameters, then the inference of a latent cause is equivalent to a quadratic increase in complexity (Schwartz, 1978). While this problem might not seem as pressing in the case of occasion setting, an expanding parameter space is a crucial consideration when dealing with real-world, infinite task spaces.

In summary, the model proposed by Gershman, Blei, & Niv (2010) solves the limitation of a finite number of latent causes by assuming an infinite parameter space. This equips the model with impressive generality in terms of what can be learned. However, this framework still falls short of adequately modeling OS, as it fails to account for the feature-positive effect; furthermore, the model is associated with a number of computational costs.

Overall, this last problem can be avoided with a careful formulation of the situations that engender novel state representations. Surely, when put into a situation consisting of largely unfamiliar events and stimulus configurations (e.g., an animal's first experience in an operant chamber), a state representation of the novel context ought to be stored. This said, the encoding of observations into state representations cannot be reduced to a "novel-familiar" binary process (i.e., pattern separation versus pattern completion), since most observations are combinations of novelties and familiarities, and therefore do not fall neatly into either category. This complexity could be modeled by having states consist of both unique parameters (the distinguishing components), as well as parameters shared by a number of separate, yet related states (the familiar components).

Fuhs and Touretzky (2007) developed a hidden Markov model (HMM) that exploited this very idea. The model's building block is a state *s* comprising parameters θ_s, which specify the transition probabilities to related states, as well as define the likelihood of making specific observations P(*data* | θ_s). Additionally, the model defines dependent states that share parameters with the independent

states they belong under; for instance, if state *s'* is dependent on *s*, it will include parameter sets θ_s and $\theta_{s'}$, such that the relative contribution of each is weighed by a specific mixture parameter. Indeed, parameter sharing has the advantage of computational efficiency, as it restricts parameterization to only a subset of the parameters - those that are unique to state *s'*. More broadly, the HMM aims to select the most appropriate model given past observations. As part of this process, a tradeoff needs to be made between models of varying complexity, such that the more complex ones incur a greater computational cost. Models are compared by computing a Bayes' factor, which consists in calculating P(model | data) for each of the models in question. Computing this posterior involves weighing the prior likelihood of model parameters, such that greater parameter spaces come at the cost of low priors. In this way, the HMM remains biased against complex models while being rather lenient with respect to model expansions that consist in the addition of dependent states (since these are somewhat less costly).

Conceptually speaking, given a state S1, the HMM has an intriguing solution to encoding observations that are highly similar to S1, yet different along certain dimensions. In this case, the HMM encodes the novel observations as S2, but only insofar as they differ from S1. More specifically, S2 comes to share common parameters with S1 (the familiar components), as well as hold its own S2-specific parameters (the distinguishing components). Furthermore, the stability of S2-specific parameters depends on repeated experience of S2-

specific contents, which is the case since repeated S2 observations increase the posterior likelihood of the expanded model. Overall, these characteristics ensure that the HMM is fitting genuine environmental differences (rather than noise), as well as making the state splitting process less computationally demanding. For these reasons, the HMM is a more biologically plausible candidate for state classification than the generative models considered above. Note, the HMM proposed by Fuhs & Touretzky (2007) was intended to give insight into context learning, and thus remains mute regarding the acquisition of action values. This

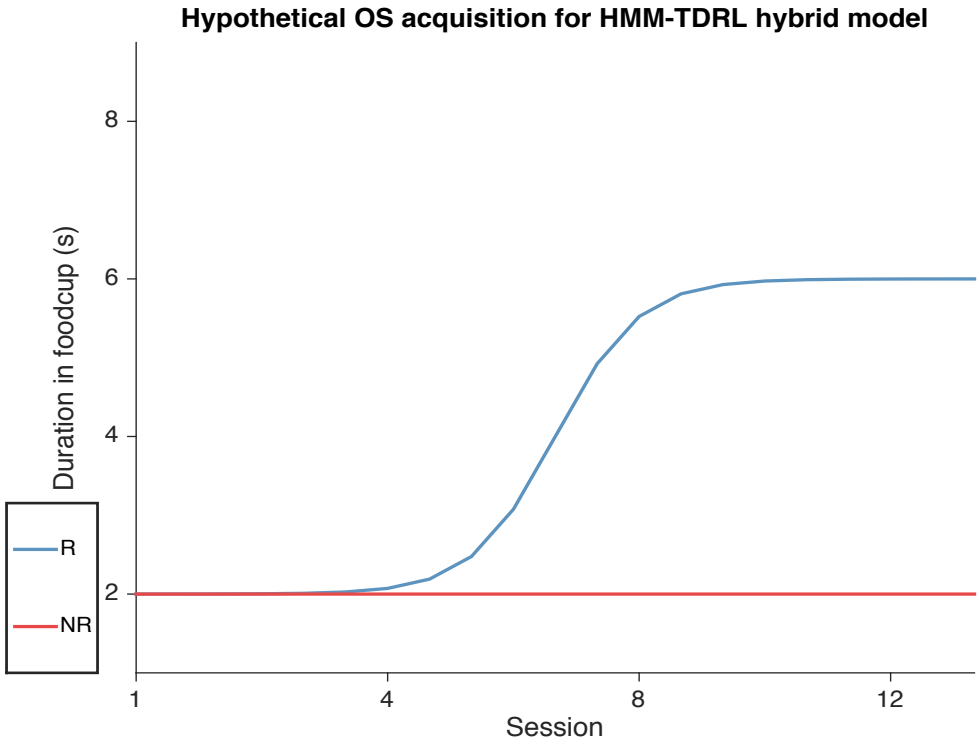


Figure 6: A hypothetical acquisition curve of OS for the combined HMM-TDRL model. The two curves are flat over the first 5 sessions – this reflects the initial stage of processing, in which the model is learning to assign trial types into distinct states. Once these states have been acquired, TDRL is employed in order to assign state values – this transition occurs at the point the two curves start diverging.

said, the model can easily be extended to incorporate state-specific reward learning. One possibility would be to first identify relevant states using the HMM approach (as described above), and then run TDRL on these in an unsupervised manner. This implies a two-step process, with separate algorithms acting at each step. A potential complication would then be to decide at what point TDRL should kick in to assign state values; presumably, this would happen upon reaching some sort of criterion reflecting confidence in how well relevant states have been identified.[6] As described above, the HMM is always making tradeoffs between models of various complexity; this is achieved by comparing the models' posterior likelihoods, such that high posteriors reflect a more reliable partitioning of the environment into relevant states (as described above). For this reason, the posterior might be an appropriate criterion. Setting a threshold it would need to surpass would effectively establish the critical level at which TDRL would be run in order to assign values. Needless to say, the acquisition curves for POS and NOS would look very different from those in Figure 2 - responding would be flat over the first few sessions (reflecting initial state learning), and eventually start increasing in the case of the rewarded trial type (reflecting TDRL onset once states have been acquired, see Figure 6). Furthermore, since the HMM assumes no differences in the acquisition rates of different states, it certainly cannot explain the feature-positive effect. Presumably, the model would need to weigh

[6] This would prevent assigning state values to models that poorly capture the structure of the environment.

environmental features in a specific way in order to acquire POS- and NOS-related states at different rates.

In conclusion, the above-described statistical models provide novel insights into the acquisition of OS. Each of the models solves OS by associating each trial type with a distinct state. This achieves a useful partition of the agent's experience, in that an otherwise ambiguous target can easily hold two distinct meanings (one in each state). As described above, however, this notion can be extended, and the target can effectively hold as many meanings as the number of states it appears in. This idea is especially powerful, as it provides a general framework for resolving ambiguity.

In the following section, we review neurobiological findings that pertain to ambiguity resolution, and couch these in terms of the conceptual framework explored above.

Neurobiology of ambiguity resolution

Interestingly, the neurobiological underpinnings of OS remain largely unknown. For this reason, we extend the discussion of OS neurobiology with ideas and evidence from related tasks that involve ambiguity resolution (see Figure 1b). Lastly, we relate the relevant findings to the models described above.

Broadly speaking, the appropriate representation of otherwise ambiguous events is related to the idea of cognitive control. While cognitive control has been defined in a number of ways (Miller, 2000; Botvinick et al., 2001; Braver, 2012), a

common aspect to all definitions is the self-regulation of behavior so as to favor the achievement of the agent's long-term goals. This necessitates knowledge of the environmental contingencies; in other words, to act in a way that furthers their goals, agents must know how their actions relate to specific outcomes, and how these relations are updated as the environment changes. This implies knowledge of task structure - a fundamental component of all ambiguity resolution tasks. In appetitive OS, for example, subjects are typically food restricted, and presumably, one of their goals is to maximize food reward. Clearly, when the stimulus contingencies have been acquired, subjects must still regulate their behavior appropriately in order to obtain maximum reward. This is particularly true when reward delivery is contingent on emitting a particular response (i.e., instrumental OS, see Holland 1992). However, insofar as conditioned responding involves energy expenditure in preparation for the occurrence of food reward (Domjan, 2005), optimal behavioral performance is just as desirable in Pavlovian OS, even though it is not necessary for reward retrieval. As such, optimal performance in OS requires the acquisition of stimulus contingencies, as well as optimal regulation of behavioral performance. In short, ambiguity resolution involves cognitive control. For this reason, the neural substrates of cognitive control may be especially relevant to our understanding of the neurobiology underlying ambiguity resolution.

A large body of research has implicated the prefrontal cortex (PFC) in cognitive control. For instance, human neuroimaging studies showed that

increased connectivity with the prefrontal region correlates with better performance in a go-nogo task[7] (Liston et al., 2005), and that developmental changes in PFC are predictive of improvements in cognitive control as adolescents become adults (Casey, Galvan, & Hare, 2005). Interestingly, a similar set of observations has been made in rodents (Haddon & Killcross, 2006, 2007). The role of PFC has often been framed as that of an executive controller using task-dependent event-outcome relations in order to mediate appropriate processing in related brain structures (Miller, 2000).

In order to perform such a function, the PFC would need to be sensitive to changes in environmental contingencies; indeed, a host of behavioral and electrophysiological studies suggest that this is the case. For instance, rats with lesions of the orbitofrontal cortex (OFC, a subregion of the PFC) show severe deficits in acquiring serial reversal learning. In the initial phase, sham and lesion animals show no differences in learning that a response R is reinforced in the presence of stimulus S1, and punished in the presence of S2. However, the lesioned rats show a significant acquisition deficit once these contingencies are reversed (i.e., R is reinforced during S2, and punished during S1) (Schoenbaum, Setlow, & Nugent, 2003; Delameter, 2007). Interestingly, electrophysiological studies revealed that across reversal phases, certain OFC neurons fire in response to stimuli predictive of the same outcome - an example would be a

[7] A typical go-nogo task involves two distinct cues - one that requires making a response ('go' trials), and one that requires withholding a response ('nogo' trials). The goal is to perform the correct response as swiftly as possible following the cue presentation. The amount of 'cognitive control' that is required to perform optimally can be controlled by varying the number of 'go' trials that precede a given 'nogo' trial.

neuron firing to S1 in the initial phase, and S2 in the reversal phase (the outcome signaled by these stimuli is identical). Furthermore, studies of ensemble firing have extended this function to other regions of the PFC as well. Durstewitz et al. (2010) simultaneously recorded mPFC neurons while rats performed a rule-switching task, and found that sudden changes in task rules were accompanied by abrupt shifts in network activity, perhaps reflecting the acquisition of novel task rules. In a similar vein, Karlsson, Tervo, & Karpova (2012) showed that in the face of abruptly changing task rules, mPFC ensembles undergo rapid transitions between a number of distinct network states - the authors suggest that these correspond to periods when the animals forge novel beliefs about the structure of the environment. Together with the lesion studies, the electrophysiological findings suggest that the OFC is involved in processing and updating environmental contingencies.[8] Indeed, the fact that PFC is sensitive to environmental changes is relevant to both cognitive control and ambiguity resolution, as both require the maintenance and updating of specific sets of event relations.

Moreover, there is an interesting connection between the role of PFC described above, and the idea of state learning discussed in the previous section. In short, generative Bayesian models demonstrate that ambiguity resolution can be achieved by acquiring distinct states (or latent causes), such that each comprises a unique association between the ambiguous target and the

[8] For additional evidence from stimulus devaluation studies, see Gallagher, McMahan, & Schoenbaum (1999), and Pickens et al. (2003, 2005).

outcome. Wilson et al. (2014) proposed an insightful OFC model that relies on this exact principle; the study suggested that sudden changes in task contingencies (e.g., as in serial reversal learning) are handled by storing distinct OFC states, which define separate sets of event-outcome relations. In addition, OFC lesions were modeled as an inability to maintain more than one state at a time. Amazingly, these assumptions were sufficient to model a large body of research detailing the effects of OFC lesions on behavior in a number of related tasks (e.g., reversal learning, extinction, devaluation, delayed alternation, etc.). Note, the Wilson et al. (2014) model suggests that OFC only maintains an effective state decomposition (see Blanchard et al., 2015 for a related experimental finding), and that relevant states are passed down to separate structures, which acquire state-dependent action values via a reinforcement learning algorithm. The nucleus accumbens (NAC) is a suitable candidate for this downstream processing, as it has been shown to invigorate reward-related responding based on its interactions with frontal and temporal regions that define the most relevant task contingencies (Floresco, 2015). According to this general model, OFC operates as the disambiguator in orchestrating contextually driven representations of otherwise ambiguous events; interestingly, this is one of the central claims propounded by Miller (2000) regarding the function of PFC.[9]

Overall, given the parallels between cognitive control and ambiguity resolution, it is likely that tasks such as OS also rely on PFC-dependent processing. Meyer & Bucci (2016a) provided some of the first causal evidence

[9] See Box 2 in the original publication.

suggesting that this is the case. The authors found that simultaneously increasing activity in NAC and decreasing activity in OFC resulted in significantly protracted acquisition of NOS. These results can be interpreted within the OFC framework described above. If OFC is indeed involved in encoding state-dependent event relations, its inactivation would result in an inability to encode trial type cue-relations as belonging to distinct OFC states (Wilson et al., 2014). Together with an overactive NAC driving approach to reward related cues (i.e., the target), this would result in a deficit in using the feature to cancel the target's reward predictive value on unrewarded trial types. A related study showed that pretraining lesions of the prelimbic (PL) cortex also retard NOS acquisition - this is consistent with PL's role in the acquisition of behavioral strategies (MacLeod & Bucci, 2010; also see Meyer & Bucci, 2014a). Lastly, it has been shown in human subjects that negative occasion setting (dubbed 'context monitoring' in the original publication) is associated with increased activity in ventrolateral PFC (Chatham et al., 2012). Overall, these results strongly suggest that there is an important overlap between the neural substrates of OS (and ambiguity resolution more broadly), and those of cognitive control.

In fact, NOS has previously been used as a model of proactive inhibition, a process by which cues in the environment are used to withhold inappropriate responding (Braver 2012). In this manner, NOS has been used to study the neurobiological underpinnings of cognitive control. For instance, Meyer & Bucci (2014b) showed that adolescent rats are significantly slower at acquiring NOS

than are adults, and offer the interpretation that this is due to an imperfectly developed adolescent PFC (for a review, see Meyer & Bucci, 2016b). This idea is consistent with human imaging studies suggesting that adolescents have limited cognitive control capacities due to a developing PFC (Liston et al., 2005; Casey et al., 2005). In order to test whether the NOS impairment reflected a learning or a performance deficit, the authors conducted the following experiment: adolescents were trained for 6 consecutive sessions, such that training was resumed once they became adults. In this manner, the subjects acquired NOS in the same total number of sessions as adults (Meyer & Bucci, 2014b). Given these results, the authors concluded that protracted NOS acquisition in adolescents was due to a performance deficit. While this interpretation is at odds with the notion that delayed acquisition may be due to an inability to encode task contingencies into distinct states (Wilson et al., 2014) - a proposed learning deficit - the results of the study itself are not inconsistent with this idea. Indeed, learning is often thought of as a gradual process; however, careful trial-by-trial analyses have shown that uncertainty can bring about rapid PFC network transitions that correlate with the rapid acquisition of novel task contingencies (Gallistel, Fairhurst, & Balsam 2004; Durstewitz et al., 2010; Karlsson, Tervo, & Karpova, 2012). As such, it is plausible that during the 6 days of training as adolescents (immature PFC), subjects only learned to encode the discrete cues, and that this encoding promoted rapid acquisition of conditional cue contingencies once the subjects experienced additional training as adults

(mature PFC). This alternative interpretation would suggest that protracted NOS acquisition is a learning deficit.

As argued by Wilson et al. (2014), the OFC (and by extension the PFC) is likely not the only brain region that supports state learning. For instance, there is copious evidence to suggest that the hippocampus may play an important role in state splitting. In spatial tasks, hippocampal pyramidal cells preferentially fire in specific locations in the environment (O'Keefe & Dostrovsky, 1971). It has been shown that these place cells come to non-topographically tile the entire extent of the animal's current spatial environment, and 'remap' once the animal is placed in an entirely novel environment (Muller & Kubie, 1987; Muller, 1996); in other words, place cells that remapped show preferential firing in completely independent spatial locations across the two physical contexts. Furthermore, when the animal is placed in a novel context that is largely similar to some [10]previously experienced environment, remapping occurs gradually, and is contingent on repeated experience with the novel environment (Jeffery, 2000). Interestingly, this finding was one of the motivations for developing the HMM for context learning (Fuhs & Touretzky, 2007), which primarily served as a model of hippocampal place cell remapping. Overall, the hippocampus is involved in representing spatial contexts by sparse place cell subpopulations, which may be viewed as distinct neural states. Accordingly, remappings can be regarded as state transitions. Given that it is well equipped to deal with problems that can be

[10] For further evidence of hippocampal involvement in context processing and state learning (e.g., pattern separation and pattern completion), see Kim & Fanselow (1992), Wiltgen et al. (2006), and Knierim & Neunuebel (2016).

solved by maintaining state decompositions of the environment, the hippocampus may be crucially involved in ambiguity resolution.

A number of studies have explored the effects of hippocampal lesions on OS. Early studies were marked by inconsistencies in the outcomes of various lesion techniques. For instance, while aspiration lesions prevented acquisition of POS (Ross et al., 1984), ibotenic acid lesions had no effect (Jarrad & Davidson, 1990). Both of these lesion results were replicated by Jarrad & Davidson (1991), who further observed that aspiration lesions produced significantly greater extra-hippocampal damage than ibotenic acid lesions. As such, the authors concluded that the deficit in POS resulted from damage to structures outside the hippocampus, and not from damage to the hippocampus proper. A number of later studies corroborated this conclusion (Moreira & Bueno, 2003; Holland et al., 1999). Interestingly, Holland et al., (1999) further found that neurotoxic hippocampal lesions caused a marked deficit in acquiring NOS. Indeed, it is curious that an intact hippocampus is necessary for NOS, but not for POS. It has been speculated that this asymmetry might in part be due to the inhibitory demand present in NOS, where the feature must be used to cancel (or inhibit) the target's association with reward. This is a plausible explanation, given that early theories of hippocampal function imputed to the structure an important role in inhibitory learning (Benoit et al., 1998; Schmajuk & DiCarlo, 1991). An alternative explanation would be that the asymmetry reflects an underlying difference in state acquisition across POS and NOS. If state splitting occurred at

a significantly slower rate in NOS, then presumably, this form of OS would be particularly sensitive to hippocampal damage. In order to model this properly, however, the rate of state splitting would need to somehow be modulated by the individual features that make up the states (e.g., aversive reward omission, appetitive reward delivery, number of features in a state, etc.). The generative Bayesian models (Courville, Daw, & Touretzky, 2006; Gershman, Blei, & Niv, 2010), and in particular the HMM for context learning (Fuhs & Touretzky, 2007), could be adapted to include feature identity as an additional state splitting parameter. Interestingly, this extension would also account for the feature-positive effect. This would suggest that in OS, the feature-positive effect occurs for the same reason that hippocampal lesions selectively impair NOS: due to differences in the rates of state acquisition between NOS and POS. Clearly, these predictions would need to be tested experimentally.

Furthermore, it has been demonstrated that the rate at which NOS is acquired can be manipulated experimentally. Subjects treated with nicotine - a general acetylcholine receptor (ACHR) agonist - discriminated in significantly fewer days than controls, as well as learned at a faster rate (MacLeod et al., 2006; MacLeod, Vucovich, & Bucci, 2010; Meyer, Chodakewitz, & Bucci, 2016). Strikingly, subjects that received a broad-spectrum ACHR antagonist required a greater number of days to discriminate. While this effect was not a mirror image of that produced by nicotine administration (e.g., there were no differences in learning rates), as a whole, the results suggest that ACHRs play an important

role in the acquisition of NOS. A relevant interpretation the authors suggest is that cholinergic input into the hippocampus may restrict the ambiguity produced by interfering stimuli, thereby increasing the rate of acquisition (MacLeod et al., 2006; Baxter, Holland, & Gallagher, 1997). Overall, this is consistent with the notion that hippocampus plays a role in OS (and ambiguity resolution more broadly) by supporting the acquisition of trial-type specific stimulus contingencies as distinct states. Thus, cholinergic input may simply act by speeding up the rate by which states are acquired. This could be modeled computationally as an alteration in the parameters that promote state splitting; for example, the generative Bayesian model proposed by Gershman, Blei, & Niv (2010) could account for this by dynamically altering the value of α to reflect the amount of cholinergic input.

Relevant to this discussion, the hippocampus has been shown to play a crucial role in a number of other tasks that involve ambiguity resolution more broadly. For instance, hippocampal lesions have a pronounced effect on contextual OS (Yoon et al., 2011) where the physical context (e.g., operant box, etc.) takes on the role of the disambiguating feature (Bouton & Swarzentruber, 1986; Urcelay & Miller, 2014). Indeed, this is not surprising given hippocampus' obvious role in contextual processing. Relatedly, inactivating the hippocampus can affect the renewal effect. Renewal comprises three distinct phases; first, a CS-US contingency is acquired in a conditioning context. Next, the CS undergoes an extinction phase in which conditioned responding is reduced by

repeatedly presenting the CS in the absence of the US. In a crucial test, the animal is moved outside the extinction context, and expresses resurgence of responding (i.e., renewal) when it next encounters the CS (Bouton, 2004). It is thought that the extinction phase produces a novel CS-US contingency whose retrieval depends on contextual disambiguation. In this sense, renewal carries similarities to contextual as well as discrete OS. Xu et al. (2015) showed that fear renewal is disrupted upon optogenetic inactivation of ventral hippocampal neurons that project to the central nucleus of the amygdala - the authors argued that this pathway is specifically involved in the contextual modulation of cued fear (Bouton & Swarzentruber, 1986). Furthermore, Corcoran & Maren (2004) showed that when there is ambiguity regarding the CS's meaning in the final test phase,[11] renewal crucially depends on an intact hippocampus; presumably, the hippocampus is needed to resolve the ambiguity in order to appropriately express renewed responding during the test phase. The role of the hippocampus in ambiguity resolution is further demonstrated in a spatial pattern separation task – or spatial delayed match-to-sample task (see Figure 1a) - developed by Gilbert, Kesner, & DeCoteau (1998). In an initial phase, the experimenters placed an identifying object on top of a baited food well (one of 15 identical wells), and allowed a rat to navigate towards it in order to retrieve the reward. In a

[11] Test phase ambiguity arises in AAB, ABC, and AAA renewal (Corcoran & Maren, 2004); the letters denote the three phases of the experiment. For example, AAB renewal means that the conditioning and extinction phases occurred in context A, while the test phase occurred in a novel context B. In both AAB and ABC renewal, the test phase involves retrieval ambiguity, since the CS had never been experienced in the test context. In the case of AAA renewal, however, the CS had undergone both conditioning and extinction in A, resulting in a degree of ambiguity regarding CS meaning in the test phase.

subsequent test phase, the same well was baited, but two distinct wells were covered with the identifying object - the baited well, and a separate foil well. Thus, the additional object (on top of the foil well) acted as a source of interference in recognizing the baited well. It introduced ambiguity regarding the reward location, and required resolution by holding in mind information from the initial phase - the location of the baited well with respect to distal cues. Interestingly, hippocampal lesions (and specifically dentate gyrus lesions - see Gilbert, Kesner, & Lee, 2001) produced a marked deficit in identifying the baited well. These results are consistent with the idea that hippocampus represents separations in spatial contexts - in other words, it is involved in state learning. This provides further evidence that ambiguity resolution is tightly coupled with state learning, and that hippocampus may be a crucial component of the circuitry implementing these processes.

Lastly, there is electrophysiological evidence to suggest that hippocampus performs state learning in ambiguity resolution tasks involving non-spatial cues; this is crucial information for dispelling the suspicion that hippocampal state learning may be restricted to tasks with strong spatial components. Firstly, a delayed match-to-sample odor discrimination task had been employed to characterize temporal modulation of ensemble firing in the hippocampus (MacDonald et al., 2011, 2013). Performing a response to a target odor was reinforced if and only if the feature odor matched. In this sense, the value assigned to the response upon target odor presentation was ambiguous, and

needed to be resolved by determining whether the target odor was a match. This disambiguation was done on a trial-by-trial basis, exactly as in OS. The authors found that temporally modulated CA1 firing patterns during the delay were largely distinct depending on feature odor identity, and were predictive of trial performance (MacDonald et al., 2013). This suggests that the hippocampus is involved in retaining feature information in order to inform the value of a subsequently presented target. A similar framework of interpretation applies to the findings made by Pastalkova et al. (2008). The authors employed a delayed-alternation task that required rats to shuttle between two arms of a maze in order to obtain reward; the task was made more difficult by including a delay in between alternations during which the animal had to run on a wheel. This delay added an element of interference that likely led to some degree of ambiguity regarding the upcoming reward location; indeed, animals had to hold online which arm they came from in order to correctly select (or disambiguate) the rewarded arm on the next trial. In this sense, the most recent response can be viewed as a "feature" that determines the reward value of the subsequent "target" response. During the delay period, the authors observed temporally modulated firing sequences in the hippocampus that were predictive of subsequent performance. Overall, the observed sequences give credence to the view that hippocampus binds discontiguous events in time in order to subserve episodic memory (MacDonald et al., 2011, 2013; Pastalkova et al., 2008). In the case of hippocampal theta sequences that extend before or behind the animal as it

navigates through a maze, it has been suggested that sequences serve to segment the environment into chunks (e.g., between landmarks) that may be important for correct navigation (Gupta et al., 2011). The sequences observed during delay periods in match-to-sample and delayed alternation tasks might reflect a similar chunking process, in that discontiguous, yet decidedly related events become bound in time. The chunking of related events is similar to the idea of state learning, as both involve the clustering of events that are decidedly interrelated, and necessary to know about for the purpose of correct performance.

In summary, there is experimental evidence to suggest that certain brain regions (e.g., OFC, hippocampus) might encode stimulus relations into distinct neural states. This is a crucial piece of evidence, as it suggests that solving task structure by maintaining information in separate states is a realistically plausible mechanism. This lends further credence to the statistical models of state learning discussed above, underscoring that they are indeed apt conceptual frameworks for reasoning about the neural processes that resolve ambiguity. Undoubtedly, the notion of state learning in ambiguity resolution opens the door to a number of exciting experiments. The following section introduces an experiment that was designed to test whether hippocampal states are associated with ambiguity resolution in OS.

Part 2: Experiments

As suggested above, OS may be solved by maintaining relevant stimulus relations as distinct states that are toggled on a trial-by-trial basis. One possibility is that OS-specific state learning is carried out by the hippocampus, and that states are encoded in the form of trial-type specific ensemble firing patterns. If this were the case, one might reasonably expect to observe sequences of neural activity such as those discovered by Pastalkova et al. (2008), and MacDonald et al. (2011, 2013); presumably, such sequences would occur during the inter-stimulus interval that separates feature offset, and target onset (F—>T trials). Therefore, a meaningful analysis might be to compare ensemble activity during the inter-stimulus interval on F—>T trials with activity during an interval of the same length in the period that immediately precedes T trials. Furthermore, if F—>T and T trial types are indeed maintained as distinct hippocampal states, instances of the target belonging to different trial types ought to be distinguishable in their neural profile. These predictions can be tested experimentally by training rats in OS, implanting them with 16-tetrode hyperdrives aimed at the CA1 pyramidal layer, and acquiring neural data as they perform the task.

Unfortunately, training rats in an OS task involving F—>T and T trial types does not guarantee the acquisition of OS. For example, in NOS, the modulatory control a feature exerts over the target's association with reward is distinct from its potential role as a conditioned inhibitor. Briefly, in conditioned inhibition

subjects learn that a stimulus signals a decrease in the likelihood of experiencing some behaviorally significant event (e.g., reward, foot shock, etc.). Thus, acquiring the feature as a conditioned inhibitor would imply a direct inhibitory relationship between the feature and reward.[12] Learning to discriminate between F—>T- and T+ by using F as a conditioned inhibitor eliminates any ambiguity in the meaning of T, as the task essentially reduces to a discrimination between F- and T+. This form of learning is likely distinct from the mechanisms of state learning that are hypothesized above for the case of OS.

In order to avoid the problem of identifying whether rats acquired OS or a simple feature-reward association, we opted for a serial biconditional discrimination task that eliminated the latter possibility in the underlying learning. In its simplest form, biconditional discrimination involves two distinct features and two distinct targets, such that during each trial, the animal must use the unique feature-target cue relations in order to predict the trial's reward value (Honey & Watt, 1998). In our experiments, two visual stimuli were used as the features (V1 and V2), and two auditory stimuli were used as the targets (A1 and A2). Specific trial types were as follows: V1—>A1+, V1—>A2-, V2—>A2+, and V2—>A1-. The primary variable of interest was how vigorously animals would respond to the auditory target given the visual feature that preceded it. Note, since each stimulus alone is rewarded at a rate of 50%, optimal performance can be attained by no means other than learning the unique stimulus combinations. In

[12] Note, in the case of POS, the situation would be similar if the feature were acquired as a conditioned exciter. In this case, the feature and reward would develop an excitatory relationship.

Experiment 1, 8 rats were trained in biconditional discrimination where each session consisted of a unique, pseudo-randomized trial ordering (see Experiment 1, Methods). Given that after extensive training (51 sessions), the group did not show a discrimination in responding between rewarded and unrewarded auditory cues, training was discontinued. In the hopes of easing acquisition, within-session trial ordering was modified so that trials were presented in blocks of 4 of the same type; we reasoned that having animals experience a number of instances of the same trial type in sequence might aid them in acquiring the discrimination. Thus, in Experiment 2, a second set of 8 subjects was trained in a biconditional discrimination task with blocked trial ordering. After 34 training sessions, however, discrimination was no more robust than it was in Experiment 1.

Given the difficulties in the acquisition of biconditional discrimination, we decided to carry out a final Experiment 3, in which a group of 8 rats were trained in a negative occasion setting task (NOS) inspired by the behavioral design in Holland et al. (1999). In addition to a conditional discrimination between L—>T- and T+ trials,[13] the design of Experiment 3 included a non-conditional discrimination between a reinforced click (C+) and a non-reinforced noise (N-). The additional cues were especially desirable, as they provided an opportunity to test for OS using a transfer test consisting of L—>C trials; if rats suppressed their responding to C during these trials, L would be considered a conditioned inhibitor; if they responded no differently than they would on C trials, however,

[13] Note, L denotes light, and T denotes tone.

OS would be inferred. This reasoning is rooted in evidence showing that a true occasion setter's properties do not transfer well to cues that have not been trained as its target (Holland, 1992; Holland et al., 1999). Overall, rats received 35 training sessions, upon which they received 2 transfer test sessions (see the Methods section under Experiment 3). Following the transfer test and 5 days of additional training, a single rat showing the best conditional discrimination was selected for further training in a custom designed operant chamber that was adapted for electrophysiological recordings.

At the time of writing, the project is still ongoing. The goal is to reach stable asymptotic responding in the selected subject, and eventually implant a 16-tetrode hyperdrive directed at the CA1 pyramidal layer to quantify putative state shifts across rewarded and unrewarded instances of T.

Experiment 1

Methods

1. Subjects

Subjects were 8 naïve, adult male Long Evans rats obtained from Harlan Laboratories (Indianapolis, IN). All subjects were ~60 days old when they arrived, and weighed ~250 g. Animals were individually housed in a climate-controlled colony room on a 12hr light/dark cycle, and were allowed unlimited access to food (Teklad Global 14% Protein Rodent Maintenance Diet, Harlan Laboratories). After a one-week acclimation period, all rats were handled on a daily basis, and gradually food-restricted to 85% of their starting weight. All throughout behavioral acquisition, subjects' weights were maintained by feeding them supplemental chow following each training session. All procedures were performed in accord with the National Institute of Health's Guide for the Care and Use of Laboratory Animals, and all protocols were approved by the Dartmouth College Animal Care and Use Committee.

2. Apparatus

Behavioral acquisition took place in eight identical standard conditioning chambers (model # ENV-007, Med Associates, Georgia, VT; dimensions 24 cm W x 30.5 cm L x 29 cm H). Each chamber was enclosed in its own sound attenuating chamber (Med Associates, ENV-017M; dimensions 66 cm W_ x 56 cm H x 56 cm L) that was equipped with an exhaust fan to ensure constant

airflow and background noise of ~68 dB. The sidewalls and ceiling of each conditioning chamber were made of clear plastic acrylic, while the front and back walls were made of brushed aluminum. The floor was composed of 19 stainless steel rods that were 5 mm in diameter and mounted 1.5 cm apart. The center-most part of the front wall was equipped with a recessed food magazine. A photobeam across the magazine entrance was used to detect magazine entries. The delivery of two 45-mg food pellets (Bioserv) served as the reinforcer. To the left and right of the food magazine were retractable levers (Med Associates, ENV-112CM) that remained drawn in throughout all behavioral sessions. A speaker (ENV-224AM) mounted ~20 cm above and to the right of the food cup was used to deliver distinct auditory stimuli. The chamber was equipped with four panel lights (Med Associates, ENV-221M) - one above each lever (left and right panel lights), another immediately above the food cup (center-bottom panel light), and the last ~16 cm above the grid floor and over the food cup (center-top panel light). Furthermore, a house-light (Med Associates, ENV-215M) was located ~24 cm above the grid floor at the back of the chamber. A red light source (a 2.8-W bulb with a red cover) was attached to the top of the sound-attenuating chamber, and provided illumination that was invisible to the subjects, but visible to experimenters. The sound attenuating chambers were equipped with surveillance cameras that recorded behavioral sessions. The apparatus was controlled via MED-PC software located on a computer in an adjacent room.

All stimuli (visual and auditory) were 10 seconds in duration. The center-top panel light and the house-light served as the two visual stimuli. During stimulus presentation, the center-top panel light was constantly illuminated (steady light), and the house-light flashed at a rate of 2 Hz (flashing light). The auditory stimuli were a pure tone (1500 Hz, 100 dB),[14] and a white noise (85 dB)[15] generated by a Med Associates stimulus generator (ANL-926).

3. Behavioral procedures

Subjects first received a 64-minute magazine training session that consisted of 16 deliveries of the reinforcer (i.e., two food pellets), with an average inter-trial interval (ITI) of 4 minutes. Biconditional discrimination training began the day after magazine training. In each session, subjects received trials that consisted of a visual stimulus (V), followed by a 5-second inter-stimulus interval (ISI) that led to an auditory stimulus (A), which co-terminated with reinforcer delivery depending on the trial type. Overall, subjects received the following trial types: V1—>A1+, V1—>A2-, V2—>A2+, and V2—>A1- (Note: "+" denotes reinforcer delivery, while "-" represents lack of reinforcer delivery. The "—>" symbol represents the 5-second gap between the visual (V) and the auditory (A) stimulus). For all subjects, A1 was a pure tone, and A2 was a white noise. For half of the subjects, V1 was the steady light, and V2 was the flashing light - for the other half, identities of V1 and V2 were reversed. This ensured that the

[14] Note, 100 dB was the input parameter into the Med Associates sound generator; the actual sound intensity was ~90 dB (measured by a sound meter).
[15] The actual sound intensity was ~80 dB.

contingencies between visual and auditory stimulus identities were properly counterbalanced. Subjects received eight presentations per trial type, with 4-minute ITIs on average - thus, each session consisted of 32 trials in total. Since each trial was 25 seconds in duration, the entire session lasted ~142 minutes. To ensure a relatively even spacing of different trial types, each session was divided into 4 blocks of 8 trials, where each block consisted of 2 each of the same trial type. Trials within blocks were pseudo-randomized, such that within each session, at most 3 consecutive trials of the same reward value were permitted. We refer to this session type as *Pseudo-randomized*. Trial ordering was unique for each session, and subjects received two sessions per day (i.e., bi-daily training) - one at the start of their light cycle, and one at the end. Overall, biconditional training with the *Pseudo-randomized* session type continued for 34 consecutive sessions. The following session type is referred to as *Blocked*: for the next 11 sessions (from session 35 up until session 45), trials within training sessions were rearranged into blocks of 4 trials of the same trial type. Each session consisted of 2 blocks of the same trial type (8 blocks per session in total), such that overall, the number of trials within a session remained at 32. Within-session trial blocks were arranged so that no two consecutive blocks were of the same reward value. Block ordering in each session was unique, and subjects received two sessions per day. Following session 45, training was paused for 17 days (*Blocked 1*),[16] but then resumed for 6 consecutive days (*Blocked 2*). Note, these training consisted of daily sessions (not bi-daily as in

[16] All members of the lab were attending the SfN Annual Meeting conference in San Diego.

previous training) - trials were still arranged into blocks. Following a 2-day break, daily training was resumed for two subjects, and lasted for 4 days; as the second day consisted of two training sessions, the selected subjects received a total of 5 additional training sessions.

4. Data analysis

Following established procedures (Holland et al., 1999; MacLeod et al., 2006, etc.), we set out to analyze responding during auditory cue presentations with the prediction that responding to rewarded cues would be more vigorous than responding to unrewarded cues. All of the subsequent tests were performed to examine this a priori hypothesis.

The amount of time the photobeam in the food magazine was broken served as a measure of conditioned food-cup behavior. Of primary interest were the measure's readouts during presentations of the auditory cues, such that data across trial types of the same reward value were combined. More concretely, times spent in the food magazine during A1+, and A2+ trials were combined (resulting in A+), and the same was done for A1-, and A2- (resulting in A-). For each of these combined trial types, the group average within each session was computed. The same procedure was repeated in the case of the visual cues (i.e., mean responding during V- and V+ was calculated across all subjects, and for each session), which permitted a visualization of average group responding across individual training sessions (see Figure 7, and Results).

First, group-level differences in responding to visual as opposed to auditory cues were quantified. To this end, within-session responding during all auditory cues was averaged irrespective of their reward value (i.e., mean within-session responding to A), and the same was done for visual cues (yielding mean responding to V). Discrimination between A and V was assessed using a two-sample t test. [17] The crucial discrimination between A+ and A-, a metric for the acquisition of biconditional discrimination, was tested on a session-by-session basis by employing a two-sample t-test. Group differences across individual auditory trial types (i.e., A1+, A1-, A2+, A2-) were also analyzed by comparing beta coefficients of logarithmic curve estimates. This test served to compare the acquisition rates of individual trial types, and supplemented the session-by-session analysis stated above. Beta coefficients were computed for each subject in the following manner: for a given trial type, the base-ten logarithm of all consecutive trial presentations was computed (MacLeod et al., 2006), a linear

---

[17] Note, employing a t-test implies certain assumptions about the population distributions of the tested samples. The assumptions in question are those of normality, and equal variance between test samples. When in violation of these assumptions, p values resulting from the t test could turn out to be unreliable.

A more cautious statistical approach would be to first check for normality via the Lilliefors test; in the case of non-normality, an alternative non-parametric test could be employed - e.g., the Wilcoxon rank sum test (checks the null hypothesis that samples came from a continuous distribution with the same median). On the other hand, if the Lilliefors test judged the samples as belonging to the family of Gaussians, a suitable test for equal variance ought to be employed. Among the various alternatives (e.g., two-samples F-test, Bartlett's test), Levene's test might be most appropriate, as it is not too sensitive to deviations from normality in the distributions of tested samples. Despite passing the Lilliefors test, the samples' distributions could at best be judged *approximately* normal. Thus, an equal variance test that is robust to non-normality might be most suited. If Levene's test suggested equal variances, a t-test would indeed be appropriate. Else, Welch's test could be employed.

Overall, this proposal begs the question of whether employing the appropriate test would make a difference in the conclusions about the tested samples. This is something that ought to be tested, and is beyond the scope of this thesis.

model was fit to the data, and the beta coefficient was extracted. In order to quantify group differences, a two-way ANOVA was conducted[18] to compare the main effects and interactions of trial reward value and auditory cue identity on the value of the beta coefficient. All analyses were performed with an alpha level of 0.05.

Results

Average group responding during presentations of the combined trial types A+, A-, L+, and L- (see Methods) is shown in Figure 7. Due to technical difficulties, data from training session 1 could not be loaded, and was excluded from analysis. Overall, subjects began to discriminate between the visual and auditory stimuli on day 1 ($t(62) = 17.25$, $p \leq 0.01$),[19] and the discrimination remained significant for all subsequent sessions. Within the *Pseudo-randomized* section of behavioral acquisition (see Figure 7, and Methods), only two of the 33 analyzed sessions involved a significant discrimination between the A+ and A- combined trial types ($t(30) = 2.21$, $p = 0.02$; $t(30) = 2.19$, $p = 0.02$). Within the *Blocked 1* section, 8 out of the 11 sessions involved a significant discrimination between A+ and A-. Significance levels were such that $p \leq 0.1$ for each of the 8 significant sessions, and $p \leq 0.01$ for 7 of the 8 significant sessions. In section *Blocked 2*, only 2 out of the 6 sessions had a significant discrimination between the relevant

[18] The concern described in the previous footnote applies.

[19] Note, since training was bi-daily, and the data from session 1 could not be loaded, the reported result is based on analysis of session 2.

trial types (t(30) = 2.14, p = 0.02; t(30) = 1.89, p = 0.03). Given the instability of a significant discrimination across individual sessions,[20] we suspected that sporadic discrimination at the group level could be attributed to a small number of subjects that had acquired the task contingencies. Figure 8 shows two subjects (RH005 and RH008) whose discrimination levels were thought to be responsible

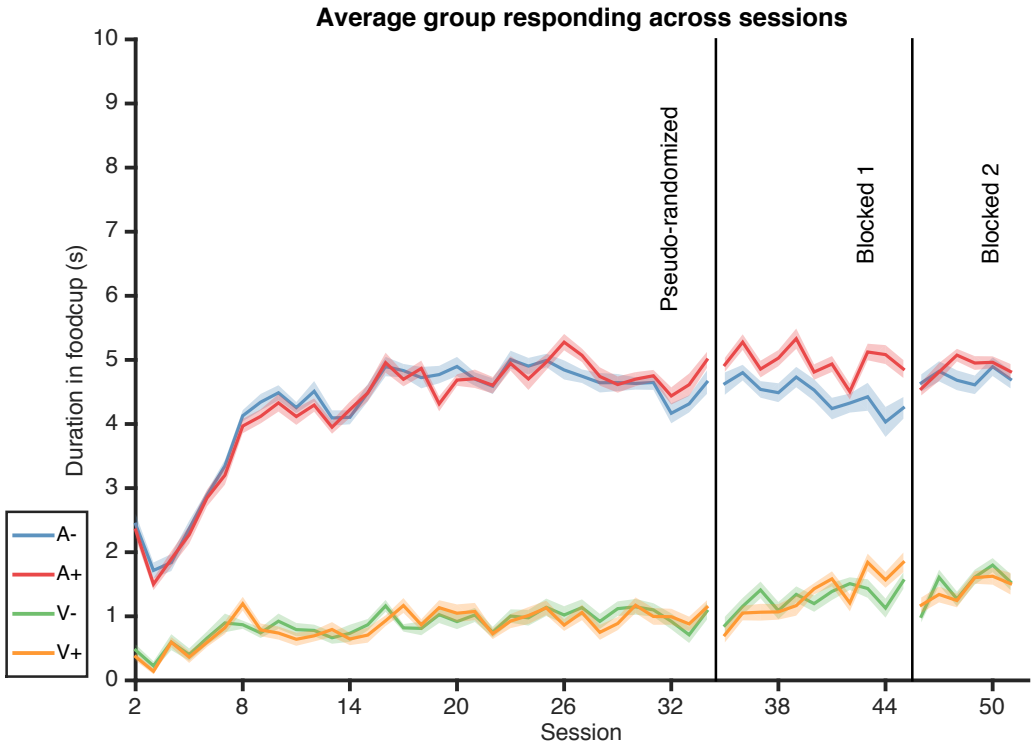


Figure 7: Group responding across daily training sessions. A- is a combined trial type produced by averaging responding across all instances of unrewarded auditory cues (A1- and A1+); combined trial types A+, V-, and V+ were produced in a similar manner. By the end of the *Pseudo-randomized* section, there was no detectable difference in responding between A- and A+; interestingly, blocking the individual trial types – section *Blocked 1* – produced a notable difference in responding. Unfortunately, this difference was lost when training was resumed following a 17-day break in training (see *Blocked 2*).

[20] Note, there is a specific worry associated with running multiple t tests; every time the test is performed, the likelihood of making a Type I error increases. In this specific instance, however, this problem is not a cause for concern, as the majority of the t tests did not reject the null hypothesis.

for this effect. Following the *Blocked 2* section, training was terminated for all subjects with the exception of RH005 and RH008 for whom training continued for another 5 consecutive sessions. While statistical analysis was not conducted for individual subjects,[21] visual inspection of responding across sessions 52 through 56 suggested a lack of discrimination between A+ and A-, and so training was terminated for these subjects as well.

A group-level ANOVA of the beta coefficients of best-fit curves for each trial type supported the above findings. Neither the main effect of reward value ($F(1,31) = 0.78$; $p = 0.38$), nor the main effect of auditory cue identity ($F(1,31) =$

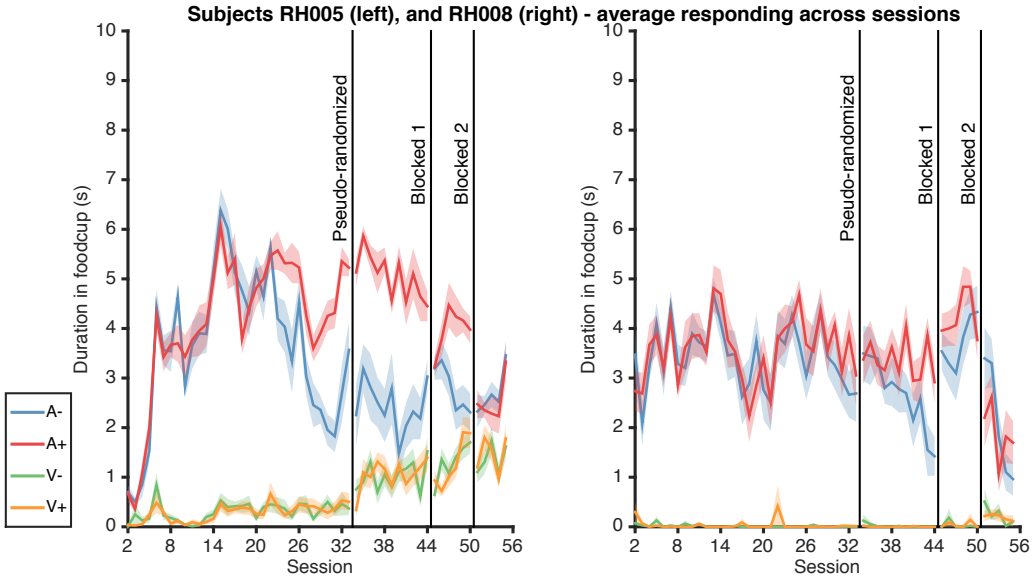


Figure 8: Responding across daily training sessions for two subjects. In the latter half of training with the *Pseudo-randomized* session type, subject RH005 showed a notable discrimination between A+ and A-. Importantly, this discrimination persisted across sections *Blocked 1* and *Blocked 2*. In a similar vein, subject RH008 first exhibited the desired discrimination towards the end of section *Blocked 2*. In the case of both subjects, however, discriminatory responding disappeared during the last section of training (sessions 51 through 56).

[21] This decision was made due to small number of trial presentations, and therefore low power at the level of individual subjects.

0.03; p = 0.87), nor any interactions (F(1,31) = 0.21; p = 0.65) were significant. As such, both analyses strongly suggested that at the level of the group, biconditional discrimination was not acquired.

Experiment 2

The results from Experiment 1 strongly suggested that at the level of the group, biconditional discrimination was not acquired. Nonetheless, a group level discrimination was apparent for a small number of sessions once the *Pseudo-randomized* session type was exchanged for the *Blocked* session type (see Figure). This suggested to us that perhaps the rate of acquisition could be considerably sped up if subjects experience the *Blocked* session type since the start of training. This hypothesis was the chief rationale motivating Experiment 2.

Methods

1. Subjects

Subjects were 8 naive, male Long Evans rats – exactly as described in Experiment 1.

2. Apparatus

In the present experiment, 7 subjects received training in Med Associates chambers (described in the Methods section of Experiment 1), and 1 subject was trained in a custom-designed conditioning chamber.

A schematic of the custom-designed box is shown in Figure 9. The floor dimensions (23.8 cm W x 29.14 cm L) were designed to be identical to those in the original Med Associates chambers (24 cm W x 30.5 cm L); thus, the floor consisted of 19 steel rods characterized by the same diameter and spacing as those in the original box. The walls were made of aluminum (alloy 5052), and were structured to angle out at 20° starting at a height of 3.84 cm (the total height of the box was 21.72 cm). The angling of the walls was designed to limit the degree to which implanted rats would hit their drives and attached headstages against the walls. The side of the box, which is not shown in Figure 9, was made up of a polycarbonate side door. The ceiling was made of the same material as the side door, and included an opening that was large enough to pass a cable

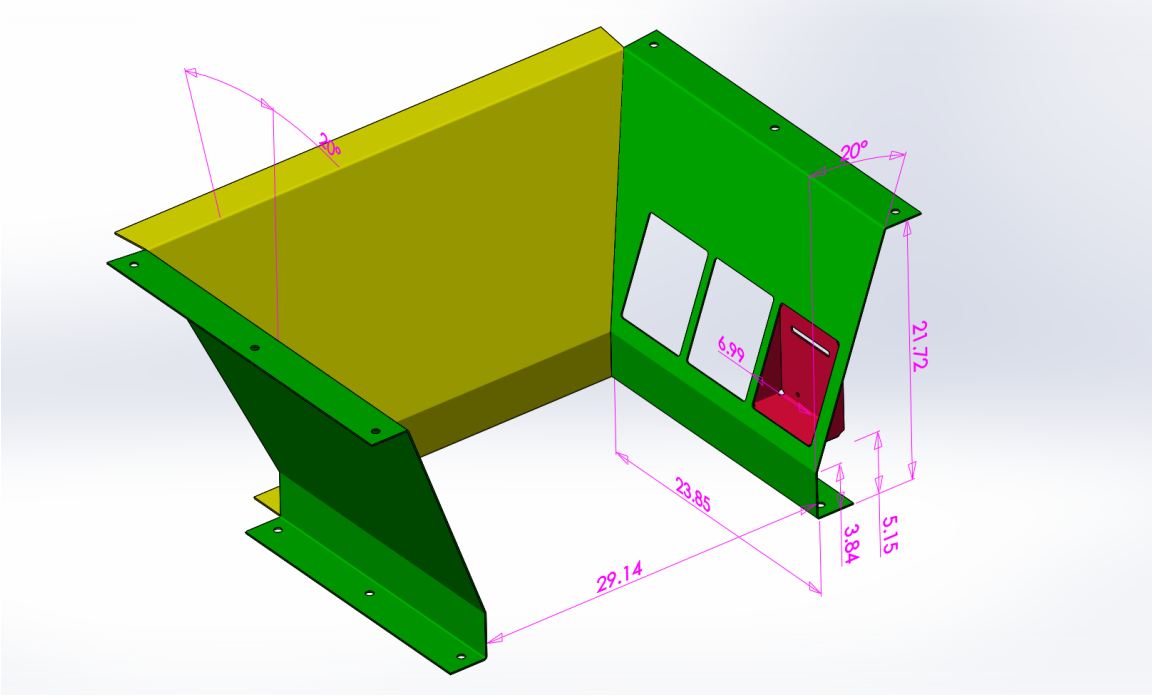


Figure 9: A schematic illustrating the general shape and relevant dimensions of the custom designed operant chamber. See text for details.

with a headstage. At a height of 5.15 cm, the center-most part at the front of the box was outfitted with a recessed adapter piece (made of acrylonitrile butadiene styrene, or ABS) that contained a high density polyethylene foodcup for collecting 45-mg food pellets (Bioserv). The pellets acted as reinforcement, and were dispensed two at a time. Furthermore, the sides of the ABS adapter were equipped with a photobeam whose breaks were used in order to detect magazine entries. Aluminum cut-outs were used to seal the rectangular spaces to the left and right of the center-most space containing the ABS adapter. The chamber contained a total of four panel lights – one above each rectangular space at the front of the box, and one in the central aspect of the back wall. A speaker (CUI Inc., model CDS-27208) mounted at a height of ~10 cm above and to the right of the right panel light was used to deliver the distinct auditory cues. The operant chamber was located in a separate room with walls made of grounded conductive material that acted as a Faraday cage. The room included a white noise generator that provided background of ~68 dB. A red ceiling light provided constant illumination that allowed the experimenters to observe subjects using a surveillance camera – the red light was invisible to the rats. The operant chamber was controlled by a Digital Lynx SX data acquisition system (Neuralynx) using a MATLAB script.

The training parameters described in this section were identical across all training chambers. As in the previous experiment, all stimuli (visual and auditory) were 10 seconds in duration. Besides the major difference that subjects received

Blocked session types since the start of training, there were a number of other minor changes in the experimental parameters; the following gives a brief description of these changes. First, in Experiment 2 the left and right panel lights served as the two visual stimuli. During stimulus presentation, the left panel light was constantly illuminated (steady light), and the right panel light flashed at a rate of 2 Hz (flashing light); furthermore, the red light was turned off during visual stimulus presentations.[22] The auditory stimuli were a click (10 Hz, 100 dB),[23] and a white noise (85 dB) generated by a Med Associates stimulus generator (ANL-926) in the case of Med Associates boxes, and a CUI Inc. speaker in the case of the custom designed box (CDS-27208).

3. Behavioral procedures

Subjects received a magazine training session whose parameters were as described in the previous experiment. Biconditional discrimination training began on the day following magazine training, and trial types were as described above: V1—>A1+, V1—>A2-, V2—>A2+, and V2—>A1-. Note, however, that the ISI duration (represented by "—>") was reduced to 1 second. Furthermore, A1 was a click sound (as opposed to a pure tone), and A2 was a white noise. With the exception of the physical sources of the steady and flashing lights in the conditioning chambers (see above), the identities and counterbalancing of V1

[22] Note, this was only possible in the Med Associates chambers. In the custom designed operant chamber, the red light was illuminated throughout the entire duration of each experiment.
[23] 100 dB was the value if the parameter given to the Med Associates sound generator. Using a sound meter, it was established that the actual sound intensity was ~70 dB.

and V2 were as described previously. Each session consisted of 8 presentations per trial type, and the total number of trials per session was 32. With an average 4-minute ITI, each session lasted approximately 139 minutes. Trials within each session were arranged into blocks of 4 trials of the same trial type, such that each session consisted of 2 blocks per trial type (8 blocks in total). Within-session blocks were ordered so that rewarded and unrewarded blocks alternated. Moreover, to ensure that each session comprised a sufficiently distinct block ordering, all session pairs had to have a Levenshtein distance of at least 4 blocks.[24] With these constraints, 24 distinct sessions were generated. To prevent too many consecutive sessions from starting with a block of the same reward value, sessions were ordered such that for every 3 consecutive days starting with a rewarded block, the next 2 days started with an unrewarded block. Training consisted of daily sessions (not bi-daily, as in previous experiment), and lasted for 34 days.

4. Data analysis

The a priori hypothesis was the same as in the previous experiment – it was expected that subjects would respond more vigorously during rewarded than unrewarded auditory cues. As such, the analyses of session-to-session

[24] Levenshtein distance (or Edit distance) is a measure of similarity between two strings. Its value reflects the the number of operations (insertions, deletions, and substitutions) that are required to make one string into another. In the case of strings "AAB" and "BAA", for example, the Levenshtein distance is 2, as making the former string into the latter requires 2 substitutions (the first letter "A" needs to be substituted for a "B", and the last letter "B" needs to be substituted for an "A").

responding were identical to those described in Experiment 1. Both an ANOVA of trial type beta coefficients, and a two-sample t test to assess A+/A- discrimination were conducted. In addition, an exploratory analysis led to a characterization of trial-to-trial responding within blocks of the same reward value was performed. This was done in order to assess the effects of block structure on conditioned responding. As described above, sessions consisted of 2 blocks per trial type (e.g., Block 1 and Block 2 for trial type V1—>A1+), such that each block comprised 4 trial presentations. The goal was to identify any notable patterns of responding across individual trials within a block. For each subject, responding was averaged across blocks of the same reward value - i.e., A1+ and A2+ blocks were averaged. This was done for each session, and in a way that preserved within-block trial ordering. In the case of unrewarded trial types A1- and A2-, for example, trials 3 of their respective Blocks 1 were averaged, trials 4 of their respective Blocks 2 were averaged, etc. Moreover, averaging was performed across days with a sliding window of 4 sessions, such that the window moved 2 sessions at a time. Resulting windows were averaged across subjects, which enabled visualization of within-block responding at the level of the group (Figure 9, see Results).

The statistical analysis rested on comparing responding on consecutive trials within a block of given reward value (A+ or A-), such that each within-block trial was an average across a 4 session sliding window (as described above). Clearly, a rather small number of trials went into computing the window trial

average for each subject. As such, subject-specific trial averages were further averaged across the entire group in order to afford more statistical power. A paired t test was employed in order to make the comparisons. As explained in more detail in the Results section, an exploratory analysis based on results of the group-level comparison resulted in performing the same comparisons for individual subjects as well.

Results

Average group responding during presentations of the combined trial types A+, A-, L+, and L- is shown in Figure 9. As in the previous experiment,

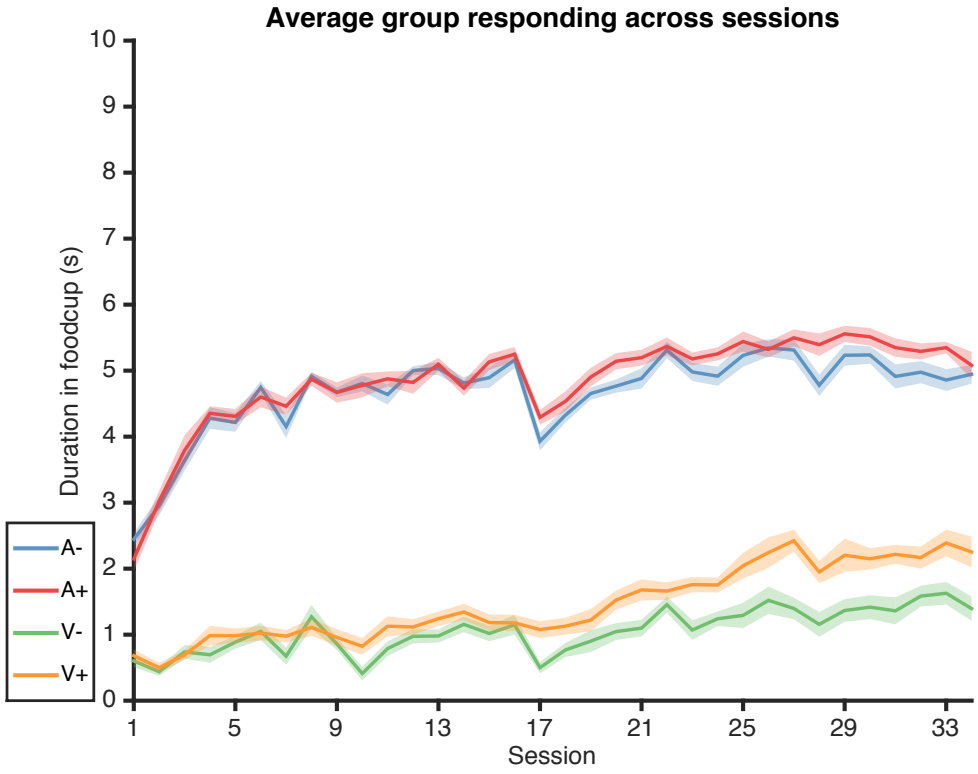


Figure 9: Group responding across daily training sessions. The combined trial types were exactly as in Experiment 1. Most notably, there was no apparent discrimination between A+ and A-. The sudden dip in responding at session 17 was caused by a single rat that exhibited highly unusual behavior – our suspicion was that the subject might have had a seizure, and did not respond for that reason. In subsequent sessions, this subject's behavior was no different from the rest of the group.

subjects showed a significant discrimination between the visual and auditory stimuli as soon as day 1 ($t(62) = 13.17$, $p \leq 0.01$). This discrimination remained significant for all subsequent sessions. On session 17, subjects first showed a significant discrimination between A+ and A- combined trial types. Among the following 18 sessions, 6 non-consecutive sessions involved a significant discrimination between A+ and A- ($p \leq 0.05$ for each of these). As in the previous experiment, these results were corroborated by an ANOVA of beta coefficients of the best-line fit for each trial type. Neither the main effect of reward value ($F(1,31) = 0.07$; $p = 0.8$), nor the main effect of auditory cue identity ($F(1,31) = 0.16$; $p = 0.67$), nor any interactions ($F(1,31) = 0.01$; $p = 0.94$) were significant. These analyses suggested biconditional discrimination was not acquired at the level of the group; as in the previous experiment, significant discrimination between A+ and A- was not stable across training sessions. As such, it was suspected that any group level differences were driven by a handful of subjects who had acquired the task rules. Average responding of the identified subjects (R115 and R119) is shown in Figure 10. Even though within-subject analyses were not conducted (due to small number of trial presentations per subject), Figure 10 suggests that even for the selected subjects, the discrimination was not particularly robust - in other words, the curves for A+ and A- do not show a consistently large separation across sessions.

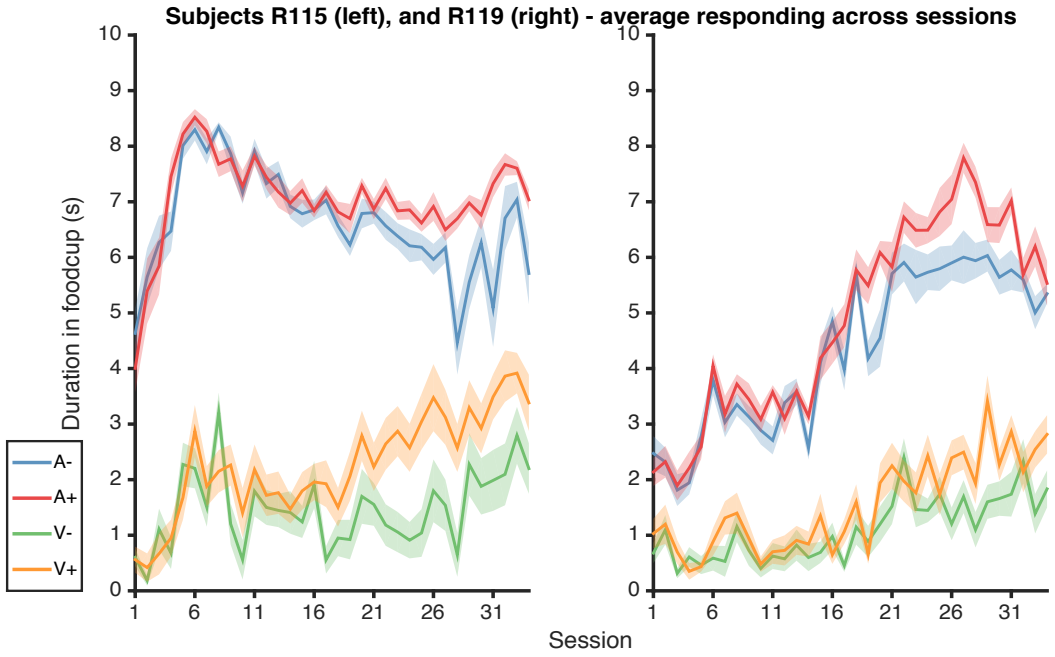


Figure 10: Responding across training sessions for two subjects. Though highly variable, in the latter half of training both subjects exhibited a discrimination between A+ and A- trial types.

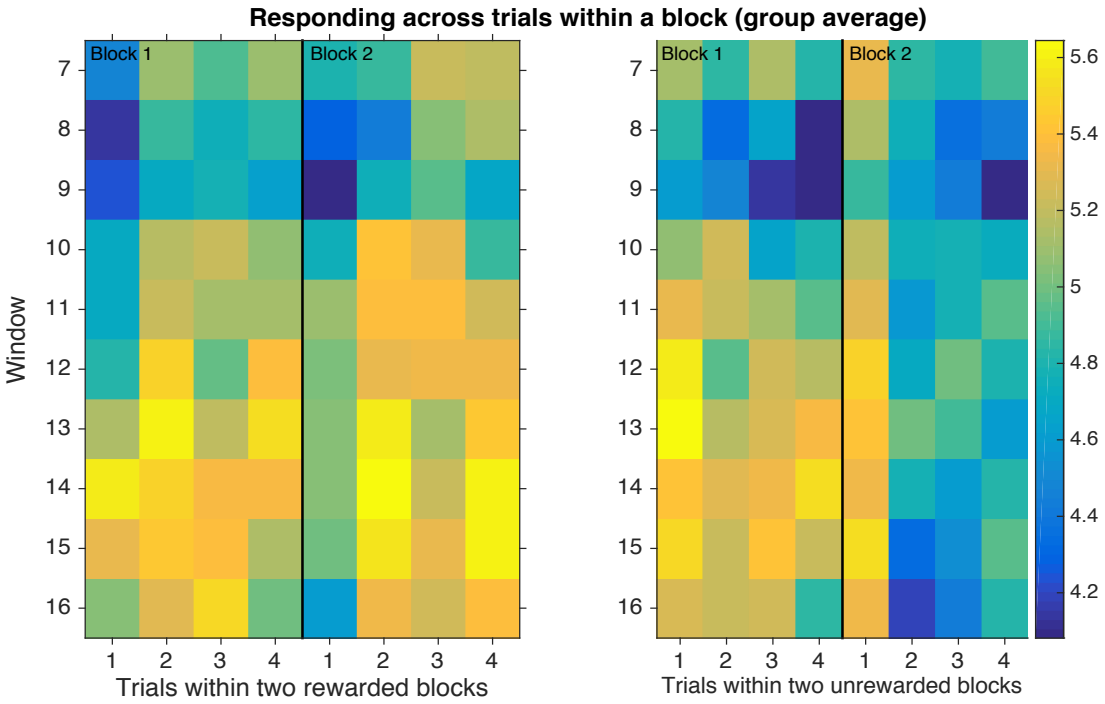


Figure 11: Average group responding across individual trials within rewarded blocks (left), and unrewarded blocks (right). In the case of rewarded blocks, subjects increased their level of responding going from trial 1 to trial 2 within the block (left panel); in the case of unrewarded blocks, subjects decreased their responding between trial 1 and 2 (right panel). For both rewarded and unrewarded blocks, this effect was especially pronounced in Block 2. Note, colder colors reflect lower rates of responding, while warmer colors reflect higher rates of responding.

In order to query further what the subjects had learned, an analysis of trial-to-trial responding within blocks of given reward value was conducted. If subjects used within-trial cue contingencies to predict reward on a trial-by-trial basis, there would be no reason to expect differences in responding across trials within a given block. However, an alternative strategy would be to adjust responding on each block by inferring its reward value. In this case, differences in within-block trial responding would be expected, as trial cue contingencies would presumably not be the primary determinant of responding. In order to probe these questions, trials within blocks of the same reward value were averaged for every subject; furthermore, within-block trials were averaged across a 4-day sliding window (see Methods for more details). The window was slid 2 sessions at a time, resulting in a total of 16 windows across all training sessions.

Figure 11 displays windows 7 through 16 (10 in total). As suggested by the session-by-session analysis reported above, no discrimination was apparent in the first 16 sessions. For this reason, the initial 6 windows are not shown. Group analysis of unrewarded blocks revealed that among the 16 windows, 3 non-consecutive windows involved a significant *decrease* in responding going from trial 1 to trial 2 within Block 1 ($p \leq 0.05$ for each of these). In the case of unrewarded Block 2, windows 12 through 14 showed a significant *decrease* in responding ($p \leq 0.01$ for the first two windows, and $p \leq 0.05$ for the third). For both Block 1 and Block 2, differences in responding between trials 2 and 3, and trials 3 and 4 were insignificant across all but 2 of the computed windows.

Similarly, a group analysis of rewarded blocks showed that within Block 1, 7 out of the 16 windows involved a significant *increase* in responding between trials 1 and 2 ($p \leq 0.05$ for each of these); the greatest number of significant consecutive windows was 3, and 5 of the 7 significant windows occurred in the latter half of the total 16. Within Block 2, 10 out of the 16 windows involved a significant increase between trials 1 and 2 ($p \leq 0.05$ for each of these), such that 6 of these were consecutive, and occurred in the latter half of the total 16 windows. As in the case of unrewarded blocks, only a small number of windows (a total of 8, $p \leq 0.05$ for each) across Block 1 and Block 2 showed a significant increase in responding between trials 2 and 3, and trials 3 and 4. Together with a visual inspection of Figure 11, these results tentatively suggest that within rewarded blocks, subjects increased their responding going from trial 1 to trial 2, and did the opposite in the case of unrewarded blocks. As the above results suggest, however, trial-to-trial shifts in responding were somewhat unstable across consecutive windows.

As such, it was suspected that any significant differences at the group level were due to a handful of subjects. For this reason, the analysis of within-block responding was additionally performed on data from R115, the subject showing the best discrimination between A+ and A- (see Figure 9). Between trials 1 and 2, a total of 7 out of 16 windows involved a significant decrease in responding for unrewarded Block 1, as well as Block 2 ($p \leq 0.05$ for each of these); in the case of Block 2, significant windows were all consecutive, and

occurred in the latter half of the 16 windows. A similar result held for rewarded Block 1, and Block 2. For the former, 8 windows involved a significant increase in responding, while for the latter, 7 consecutive windows showed the significant increase ($p \leq 0.05$ for each of these).

As a result of these analyses, it was concluded that subjects that acquired a discrimination (i.e., R115) were responding by inferring the reward value of a given block - this is apparent in the direction of shifts in responding going from trial 1 to trial 2 in rewarded as opposed to unrewarded blocks. For this reason, training sessions were discontinued.

Experiment 3

As suggested by the results above, the parameters employed in Experiment 2 proved just as inefficient at acquiring biconditional discrimination as those used in Experiment 1. For this reason, we decided to make a drastic change in experimental design, and carry out a NOS experiment that had been demonstrated to work in the past (Holland et al., 1999). The reasoning was that even if only a subset of the subjects acquired OS - and the rest acquired conditioned inhibition - this ought to be sufficient for recording purposes given the small number of subjects that are necessary in order to gather large neural data sets. Furthermore, since NOS is acquired in considerably fewer training sessions than biconditional discrimination (Honey & Watt, 1998; Holland et al., 1999; Meyer & Bucci, 2016a), it would presumably allow a much higher throughput of fruitful experiments.

Methods

1. Subjects

Subjects were 8 naive, male Long Evans rats – exactly as described in Experiment 1.

2. Apparatus

A total of 8 subjects received training in Med Associates chambers. As in the previous experiments, all stimuli (visual and auditory) were 10 seconds in

duration. The center-top panel light served as the visual stimulus. During stimulus presentation, the center-top panel light flashed at a rate of 2 Hz (flashing light). The auditory stimuli were a pure tone (3000 Hz, 90 dB), a click (10 Hz, 100 dB), and a white noise (85 dB) generated by a Med Associates stimulus generator (ANL-926).

3. Behavioral procedures

On their first day, subjects received a magazine training session with parameters as described above. Acquisition of negative occasion setting (NOS) started on the day following magazine training, and consisted of the following trial types: a flashing light, followed by a 5-second ISI, and leading to a pure tone that resulted in no reward (L—>T-); a standalone tone resulting in reward (T+); a non-reinforced white noise (N-); and lastly, a reinforced click sound (C+). Subjects received daily training sessions, such that within a session, each rewarded trial type was presented twice, while each non-rewarded trial type was presented 6 times. In total, there were 16 trials in a session, and with an average 4-minute ISI, each lasted ~70 minutes. Within-session trial ordering was arranged so that at most 3 consecutive trials of the same reward value were permitted. With these constraints, 26 sessions with unique trial ordering were generated. Training lasted for a total of 35 days, which were followed by 2 consecutive transfer test days. Each transfer test session consisted of the following trial types: L—>T, T, L—>C, C, and L; each trial type was presented in extinction, and occurred twice

per session. Apart from these differences, session parameters were as described above. Overall, transfer sessions consisted of 10 trial presentations, and lasted ~43 minutes. Following the transfer sessions, daily training resumed for 5 days, at which point a single subject was selected for further training in the custom designed operant box.

4. Data analysis

Analyses of session-to-session responding were as described in previous experiments. First, a two-sample t test was employed to quantify discrimination between N- and C+; the identical test was also used to assess discrimination between rewarded and unrewarded instance of the tone (L—>T- versus T+ trial types). In addition, a one-way ANOVA of trial type beta coefficients was conducted. Lastly, transfer test sessions were analyzed by pooling responding for each trial type across all subjects, and both transfer test days; a one-way ANOVA was performed on the resulting trial type averages.

Results

Average group responding during presentations of trial types L—>T-, T+, N-, and C+ is shown in Figure 12 (note, the LT- curve shows average responding during L—>T- trial type presentations). As in the previous experiments, subjects discriminated between the visual and auditory stimuli since day 1 ($t(6) = 5.3$, $p \leq 0.01$), and continued to do so for the remainder of the sessions. On session 3,

subjects first showed a significant discrimination between N- and C+ trial type ($t(20) = 3.08$, $p \leq 0.01$). This discrimination remained for all subsequent training sessions. Among the total 40 training sessions, only 5 non-consecutive sessions involved a significant discrimination in responding to the tone during L—>T- and T+ trial types ($p \leq 0.05$ for each of these). As in the previous experiment, these results were corroborated by an ANOVA of beta coefficients of the best-line fit for each trial type; the test revealed a significant effect of trial type on the value of

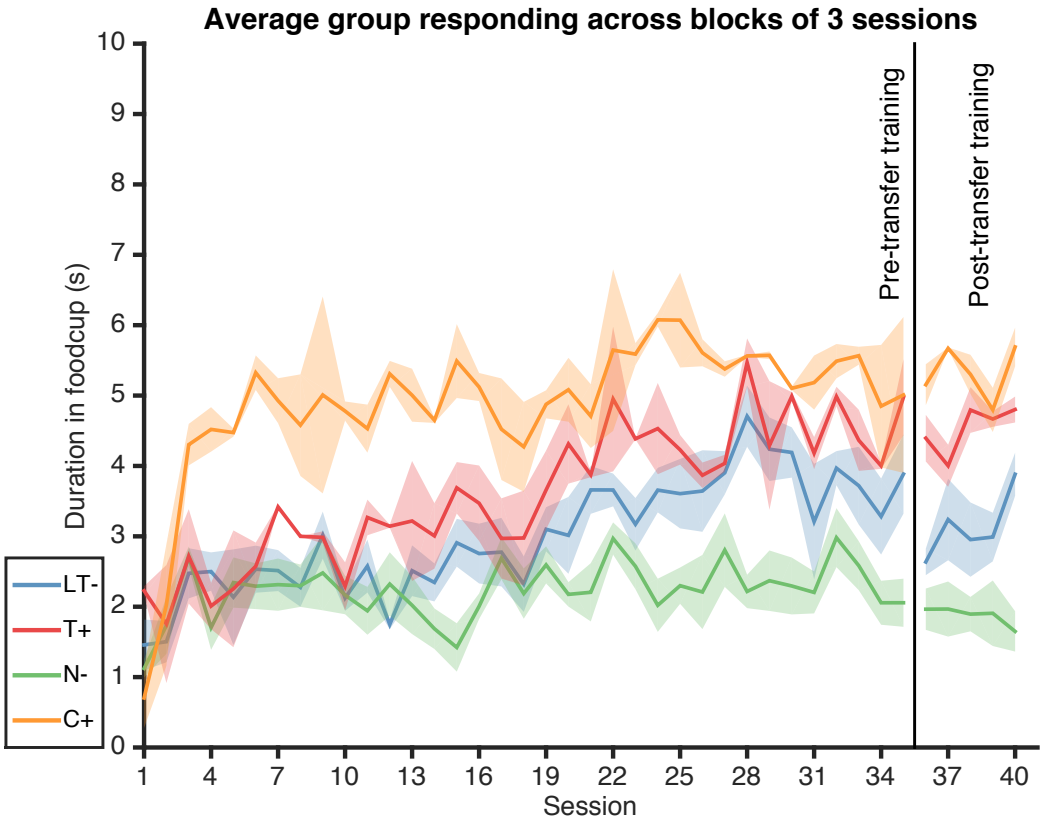


Figure 12: Average group responding across individual sessions. The subjects readily acquired the non-conditional discrimination between C+ and N-. However, a visual inspection of responding on LT- (referred to as L—>T- in the text) and T+ trials suggests that subjects did not acquire the conditional discrimination. Note, the session-by-session responding is rather noisy, and might benefit from averaging across multiple sessions.

the beta coefficient (F(3,28) = 4.18; p = 0.01). Post-hoc comparisons using the multiple comparisons test indicated that the mean beta coefficient of C+ (M = .0082, SD = .002) was significantly greater than that of N- (M = -.000124, SD = .002); furthermore, the mean beta coefficient of T+ (M = .0078, SD = .002) was significantly greater than that of N-. These results suggested that at the level of the group, subjects learned to discriminate between the non-conditional trials (N- and C+), but failed to acquire a discrimination between tones on negative occasion setting trials (L—>T- and T+). Despite these results, a visual inspection of Figure 12 reveals a separation between the curves that correspond to trial

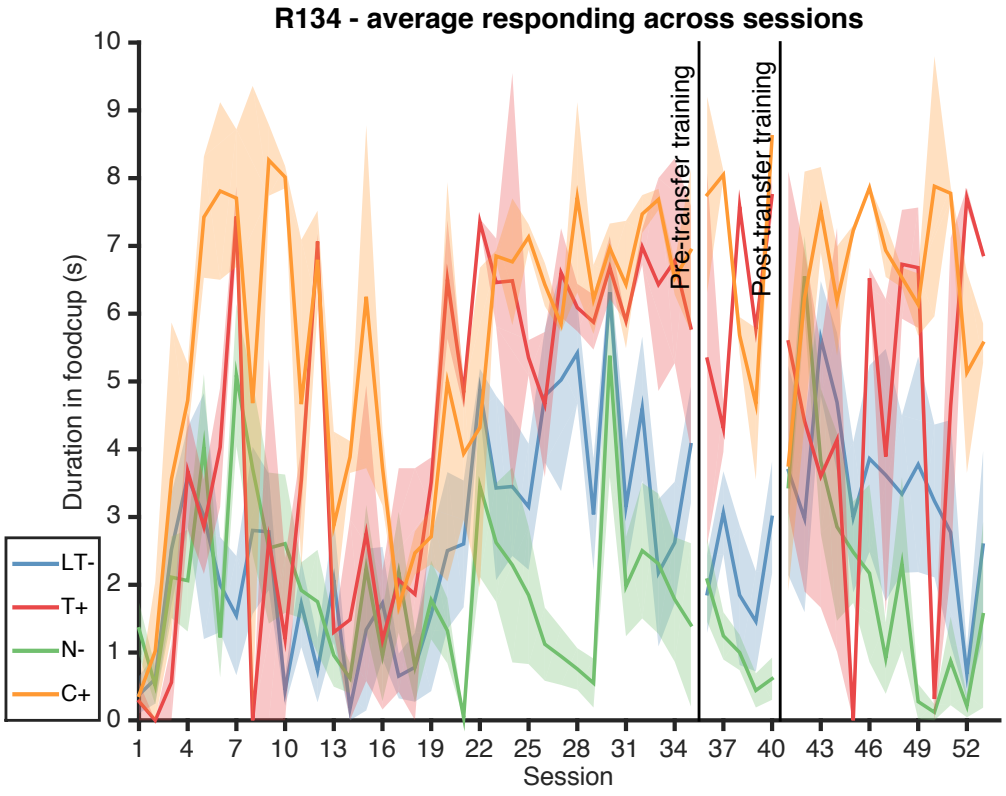


Figure 13: Responding across training sessions for an individual rat. Sessions 1 through 40 were run in Med Associates chamber with the rest of the group; sessions 41 onward were run for R134 individually in the custom designed operant chamber. Note, from session 48 onward, the subject seems to be exhibiting a relatively robust discrimination.

types L—>T- and T+. It is possible that the small number of trial presentations - especially in the case of rewarded trial types - resulted in a lack of power to make statistical statements about the discrimination. As in both of the previous experiments, it was suspected that group differences were driven by a small number of subjects who had acquired the discrimination (see Figure 13).

As such, training continued, and ultimately a transfer test was conducted in order to probe the nature o f any underlying learning; a one-way ANOVA was employed to quantify group-level differences in responding to the individual transfer trial types - these were L—>T, T, L—>C, C, and L (see Figure 14). Test results revealed a significant effect of trial type on conditioned responding ($F(4,15) = 7.08$; $p \leq 0.01$). A post-hoc multiple comparisons test showed that the rate of responding during the L trial type (M = 0.35 s, SD = 0.65 s) was significantly lower than responding during the C trial type (M = 4.22 s, SD = 0.65 s), the T trial type (M = 4.68 s, SD = 0.65 s), and the L—>T trial type (M = 4.04 s, SD = 0.65 s). There were no other differences. While these results did not allow the inference of occasion setting (responding on L—>T was no different from responding on L—>C), they were just as inconclusive with respect to the acquisition of conditioned inhibition; indeed, if L were a conditioned inhibitor, it would suppress responding to T as much as it would to C. However, results of the transfer test suggested that L did not suppress responding to either of these auditory stimuli. Given these results, the subject showing best discrimination (R134; see Figure 13) was selected for further training in the custom designed

operant chamber. Acquisition sessions were discontinued for the remainder of the group.

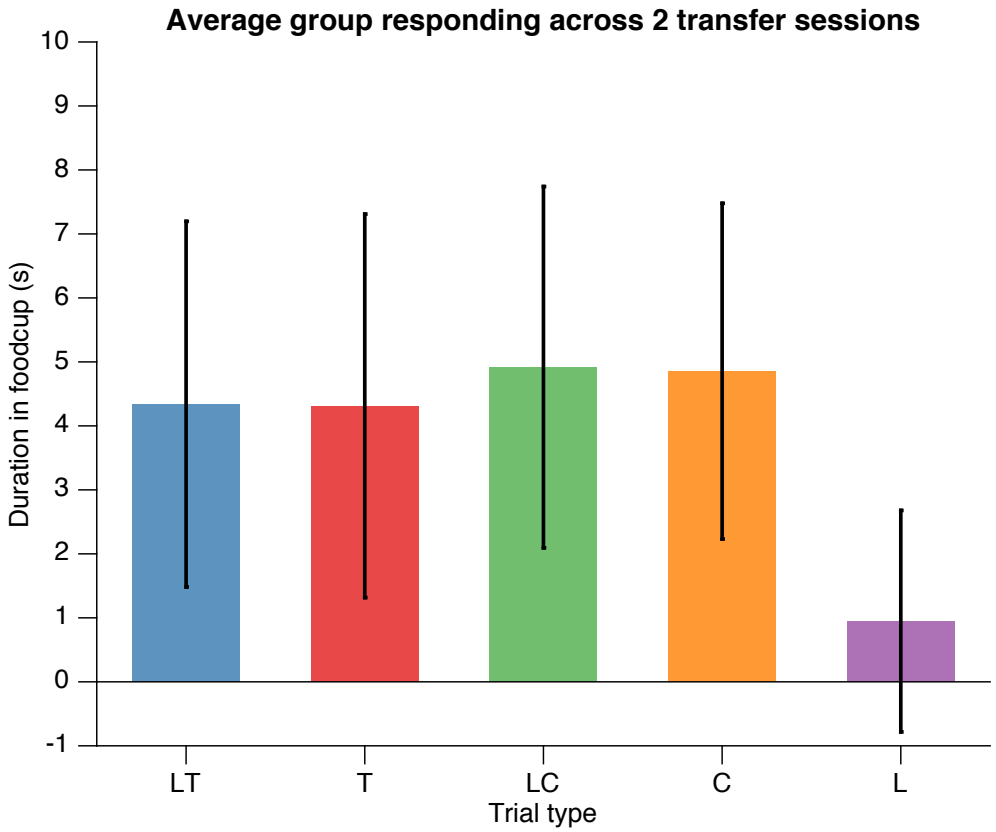


Figure 14: Average group responding across two consecutive transfer test sessions. There were no differences in responding across the trial types of interest: LT (L—>T in the text), T, LC (L—>C in the text), and C. These results were inconclusive, and did not allow inference of neither OS, nor conditioned inhibition.

Discussion

The series of experiments we carried out was borne out of a difficulty in establishing an experimental design that would reliably promote the acquisition of discriminatory responding between rewarded and unrewarded target cues. First, we sum up and discuss the results from Experiment 1 and Experiment 2, since these involved the same behavioral design - biconditional discrimination - albeit with different parameters. Next, we explain the rationale for Experiment 3 – a negative occasion setting (NOS) design - and go on to discuss the results. We conclude by sketching out a possible future research direction.

In Experiment 1, a biconditional design was selected in order to eliminate the possibility of solving the task by establishing a simple association between any single cue and the trial outcome. In biconditional discrimination, each cue is predictive of reward only 50% of the times; as such, reliable predictions regarding trial value can only be made by being sensitive to the specific sequence of cues that are present within each trial. Given that the features must necessarily act as occasion setters in order for the task to be solved, serial biconditional discrimination is ideal for tackling the question of whether the meaning of ambiguous events is resolved by encoding relevant cues as part of distinct neural states.

Unfortunately, acquisition of biconditional discrimination proved rather complicated, as suggested by the results from Experiment 1. During the first 33 sessions in which trial orderings were pseudo-randomized, the group never

exhibited a stable discrimination between rewarded and unrewarded presentations of the target. Given this null result, we decided to change the within-session parameters by grouping trials of a given type into blocks of four and alternating by reward value. Notably, in 7 out of the 11 sessions with this blocked trial ordering, the group showed a significant discrimination. Due to inconvenient circumstances, however, the experiment had to be paused – at the time, all members of the lab were scheduled to attend the SfN Annual Conference in San Diego. When training with blocked sessions was resumed, however, the group no longer exhibited a stable discrimination. In Experiment 1, our primary goal was to pilot the biconditional task, and potentially acquire neural data from a single subject; as such, two subjects showing the best discrimination were selected for further training. Nonetheless, as in the case of group level responding, stable discrimination in the two subjects disappeared following a 2-day break from behavioral training.

Clearly, the failure to acquire biconditional discrimination was not due to general deficits in performance, or learning. Indeed, if either of these were the case, animals would not exhibit any discrimination between the auditory and visual stimuli, which they did since the first day of training. The failure to acquire biconditional discrimination was especially puzzling in the face of prior evidence suggesting that rats can acquire a considerably more complex version of the task. Honey & Watt (1998) employed a biconditional design involving four distinct auditory stimuli (A, B, C, and D), such that half of each cue's presentations were

immediately followed by visual stimulus X, and the other half by visual stimulus Y (e.g., A—>X and A—>Y trials). Furthermore, half of the trial types were reinforced (A—>X+, B—>X+, C—>Y+, D—>Y+), and the other half were not (A—>Y-, B—>Y-, C—>X-, D—>X-). Sessions consisted of 5 presentations per trial type, so that each session comprised 40 trials overall. Stimuli were 10 seconds in duration, and as implied above, cues within each trial were presented serially with a 0 second delay. The average ITI duration was 2 minutes. Despite the apparent challenge posed by a biconditional task that included 8 unique trial types, subjects began to discriminate between rewarded and unrewarded target instances between sessions 16 and 24.

Relatedly, a number of studies carried out biconditional discriminations in which stimuli were presented in compound. While compound discriminations are likely acquired via a different strategy from their serial counterparts (e.g., configural strategies), a brief discussion of the past compound designs remains appropriate, as it allows a closer inspection of the session parameters that proved successful for acquisition. For instance, Harris et al. (2008) employed a biconditional design with compound cues whose components were identical to those in Experiment 1: a pure tone, a white noise, a steady light, and a flashing light. Each compound consisted of two stimuli from distinct modalities, and each trial was 30 seconds in duration. Sessions consisted of 16 presentations per trial type, with 64 trials per session in total. The average length of the ITI was 2 minutes. With these parameters, subjects started discriminating between

rewarded and unrewarded compounds at approximately session 25 (Harris et al., 2008). In a similar fashion, Delameter et al. (2017) carried out a series of biconditional experiments with parameters much like those used by Harris et al. (2008) - cues were presented in compound, and the average ITI duration was 2 minutes. Strikingly, a number of the parameters were identical to those employed in Experiment 1: there were 8 presentations per trial type (32 per session), stimulus presentations were 10 seconds in duration, and within-session trial orderings were pseudo-randomized across 4 8-trial blocks. One distinctive aspect of the design employed by Delameter et al. (2017) was that rewarded compounds were differentially reinforced; unsurprisingly, biconditional discrimination with differential reinforcement - each compound was associated with a distinct reward - was acquired more readily than biconditional discrimination with non-differential reinforcement (this was previously established by Delameter et al., 2010). Lastly, Delameter et al. (2017) carried out a biconditional experiment with parameters as above, except with half of the reinforced trials resulting in sucrose, and half resulting in pellets; this is similar to a design with nondifferential reinforcement - as in Experiment 1 - since reinforcer identity cannot be exploited to facilitate learning about the compound cues. Importantly, with these specific parameters, biconditional discrimination was never acquired.

Indeed, as suggested above, a number of successful biconditional designs made use of parameters (cue duration, cue identity, number of trials per session,

etc.) that were much like those employed in Experiment 1. Interestingly, ITI duration was the one parameter in which all of the above-described designs differed - past designs used 2-minute ITIs, as opposed to the 4-minute ITIs employed in our experiment. While shorter ITIs come with the advantage of being able to include a greater number of trials within each session (e.g., Harris et al., (2008) presented 64 trials per sessions), this might not necessarily aid in the acquisition of discriminatory responding. More specifically, each rewarded trial offers an opportunity to accrue excitation to the context. With short ITIs, contextual excitation is less likely to extinguish, and may come to block any learning to the individual cues (Kamin, 1969). Nonetheless, the number of successful biconditional designs that have used this parameter in the past suggest that piloting a biconditional experiment with 2-minute ITIs might turn out to be fruitful.

Overall, given the results from Experiment 1, we decided to run an additional study - Experiment 2 - which included changes in parameters that were hypothesized to ease the acquisition of biconditional discrimination. Notably, trials within each session were blocked as described in the previous paragraph. Given that in Experiment 1, blocking the trials brought about a notable improvement in the group-level discrimination, we concluded that subjects might benefit from such a design were they to receive blocked sessions since the first day of training. Additionally, the ISI between feature offset and target onset was shortened from 5 seconds to 1 second - this was done in hopes

of decreasing the working memory demand imposed by longer ISIs; note, while it has been shown that longer feature-target ISIs are more likely to result in OS (Holland, 1992), this finding does not necessarily extend to biconditional discrimination, since the task cannot be solved by other means than using the feature to inform the present value of the target. A potential criticism of this reasoning would be that any reduction of the temporal discontiguity between the feature and target increases the likelihood of acquiring the task contingencies by means of a configural strategy (Holland, 1992). However, it has been shown that even when the feature is immediately followed by the target (i.e., 0 second ISI), the learning that underlies biconditional discrimination is not consistent with a configural strategy (Honey & Watt. 1998). Next, the 1500 Hz pure tone, which acted as auditory stimulus A1 in Experiment 1, was exchanged for a 10 Hz clicker in Experiment 2; this change was made, as a discrimination between a white noise (A2 in both Experiment 1 and 2) and a clicker had been obtained in the same operant boxes on a previous occasion (Todd et al., 2016). The reasons for the last adjustments were primarily anecdotal, and stemmed from observations that had been made over the course of running behavioral studies in the lab. First, the background red light was temporarily switched off during visual stimulus presentations, as it had been observed that this tended to speed up learning about visual stimuli. Second, the sources of the steady (V1) and flashing (V2) lights were selected as the left and right panel lights, respectively.

We hypothesized that having the two visual stimuli emitted from sources that were equidistant from the food magazine might help behavioral acquisition.

At the level of group responding, however, these changes had very little effect. Indeed, among the total of 34 training sessions, the group exhibited a significant discrimination on a mere 6 non-consecutive sessions. Clearly, the changes we had decided to implement were not effective in speeding up the acquisition of the task. One variable of interest was whether the blocked structure had any effect on trial responding in a given block. Strikingly, it turns out subjects were using the structure of trials in order to infer the reward value of a given block, and adjust their responding accordingly. This conclusion was arrived at through a visualization of group responding across rewarded and unrewarded blocks (Figure 9), and an exploratory analysis that led to an investigation of within-block responding in the rat showing the best discrimination (see Figure 11). The identified subject shifted his level of responding going from trial 1 to trial 2 within a given block, and maintained it for the remaining trials within the block. More specifically, responding increased significantly between trials 1 and 2 in the case of rewarded blocks, and decreased significantly in the case of unrewarded blocks.

Most notably, these block-by-block adjustments are strongly reminiscent of serial reversal learning; most consistent with this interpretation is the finding that the adjustment to an appropriate level of responding requires a single trial. Indeed, when subjects are trained in a serial reversal task that involves a large

number of contingency reversals, all it takes is a single trial for the subjects to adjust their responding appropriately (Pubols, 1962). Interestingly, employing the reversal strategy in a blocked biconditional design is consistent with computational models, which posit that organisms internally decompose their environments into effective states (Courville, Daw, & Touretzky, 2006; Gershman, Cohen, & Niv, 2010; Fuhs & Touretzky, 2007). In the case of biconditional discrimination with pseudo-randomized trial orderings, an agent that stores stimulus contingencies into distinct states would have to maintain a distinct state for each trial type in order to perform optimally. In the case of blocked trial types, however, it suffices to store two states - one per block reward value. This offers an elegant way out of encoding specific stimulus contingencies, as correct performance merely requires inference of block value. While in the specific case of Experiment 2, the blocked trial orderings were counterproductive to the acquisition of biconditional stimulus relations, blocking trials could still prove to be a useful method for easing acquisition. For instance, ordering trials into blocks of 4 comprising the same feature might prove especially effective - an example of a specific block would be V1—>A1+, V1—>A2-, V1—>A2-, and V1—>A1+; given that the reward value of trials within such a block would vary, it would be impossible to infer the value of any individual trial by inferring the reward value of the block. However, the block structure could still be used to infer the reward value of specific targets - in the block described above, for example, A1 is rewarded, while A2 is not. The agent could establish this by noting that they

find themselves in a “V1 block”; in this sense, exploiting the block structure to perform optimally would necessitate using the feature in order to inform the value of the targets. This possibility is being actively pursued in the lab at the time of writing.

Given the repeated complications in the acquisition of biconditional discrimination, we decided to make a considerable change in the experimental design. More specifically, we opted for a NOS design (Holland et al., 1999) whose parameters inspired a number of experiments that had previously been run in the lab (e.g., Meyer, Putney, & Bucci, 2015; Meyer & Bucci, 2016a). Since the experimental design in Holland et al., (1999) formed the basis of Experiment 3, we briefly discuss the specific parameters it involved.

As in Experiment 3, the design employed by Holland et al. (1999) involved a conditional (L—>T1- and T1+), as well as a non-conditional (T2+ and N-) discrimination. Note, T1 and T2 were pure tones of different frequencies, N was a white noise, and L (the occasion setter) was a steady light. All of the cues, as well as the gap between the feature (L) and the target (T1), were 5 seconds in duration. Each unrewarded trial occurred 3 times per session, while each rewarded trial occurred once. Thus, each session comprised 8 trials in total, with trials fairly spaced out within the session - the average ITI duration was 10 minutes. With these parameters, control subjects showed a discrimination starting at approximately session 24.

The design employed in Experiment 3 was highly similar to that used by Holland et al. (1999). The cues in Experiment 3 were identical to those in the initial experiment, except for the 10 Hz clicker C, which was originally a 300 Hz tone T2, the 3000 Hz target tone, which was originally a 1500 Hz tone T1, and the flashing light, which was a steadily illuminated light in the original experiment. We opted to make these changes in order to increase the salience of each cue; for example, we reasoned that between a 300 Hz pure tone and a 10 Hz clicker, the clicker was more saliently different from a 3000 Hz tone. Presumably, learning would be less likely to generalize across these cues. Furthermore, we decided to shorten the average ITI lengths from 10 to 4 minutes - this allowed for each session to consist of twice as many trial presentations as in the original experiment without changing the overall session duration. Lastly, cue durations were 10 seconds, and not 5 seconds as in the original experiment. The transfer test was identical to that employed by Holland et al., (1999).

Remarkably, even in the case of this well-established design, the acquisition of NOS was fraught with puzzling complications. Out of the total 40 training sessions, only 5 non-consecutive sessions involved a significant discrimination between rewarded and unrewarded conditional trials (i.e., L—>T- and T+). A possible explanation for this surprising result is that due to the small number of trial presentations, it was not possible to make statements of statistical significance as a result of low power. Indeed, in the original experiment, Holland et al. (1999) averaged responding over blocks of 4 sessions, and analyzed

responding across individual session blocks; it is possible that a similar manipulation of the data would help establish statistical significance. However, the lab has repeatedly demonstrated that with parameters similar to those in Holland et al. (1999), statistically significant NOS discriminations can be obtained by analyzing the data on a session-by-session basis (e.g., Meyer, Putney, & Bucci, 2015; Meyer & Bucci, 2016a).

Importantly, subjects readily acquired a stable discrimination between the non-conditional cues C+ and N-. In addition to showing that the subjects did not suffer from a general learning or performance deficit, this result further suggests that the rats' ability to acquire discriminatory responding was unimpaired.

Interestingly, the specific choice to make the feature cue a flashing light (as opposed to keeping the original - and less salient - steady light) may have been one of the pitfalls of Experiment 3. Past experiments have shown that one of the crucial conditions determining whether a feature becomes an occasion setter or a simple Pavlovian CS is its relative salience with respect to the target; more specifically, it was established that features ought to be considerably less salient than the targets in order for OS to be favored (Holland, 1992). As such, one might argue that subjects did not acquire OS since the feature (a flashing light) was not sufficiently less salient with respect to the target. Given this information, however, it is curious that conditioned inhibition was not learned instead. Indeed, the transfer test was mute regarding what form of learning was acquired – its outcome suggested that neither OS nor conditioned inhibition was

acquired. Importantly, it was shown that discontiguity between feature offset and target onset is a crucial factor that discourages the acquisition of conditioned inhibition, and promotes acquisition of OS. Overall, it is possible that the salience of the feature discouraged OS, while at the same time the temporal asynchrony between feature offset and target onset discouraged conditioned inhibition – such a situation would result in a lack of discriminatory responding, which is entirely consistent with results from Experiment 3.

In conclusion, given the results of Experiments 1 through 3, it is unclear what sort of experimental design would be guaranteed to result in stable discriminations characterized by a desirable form of underlying learning (e.g., OS, and not simple conditioning). One possibility would be to completely move away from designs involving discrete features (e.g., OS and biconditional discrimination) and towards designs with contexts acting as features (Yoon et al., 2011). A significant downside associated with this approach is that subjects would be confined to a single context - and set of stimulus associations - in each session; this would eliminate any putative state shifts that might occur within a session, since the appropriate set of associations would presumably be established at session onset. However, this is not an insurmountable obstacle – in a recent study, Jezek et al. (2011) studied the effect of instantaneous context switches on CA3 ensemble firing. Using complex lighting patterns, the authors were able to artificially create two distinct contexts within the same box – the assurance that these were in fact encoded as distinct stemmed from the result

that the contexts elicited orthogonal patterns of hippocampal firing. This suggested they were encoded as separate hippocampal states. A design of this sort would not only maintain the precise temporal control over the occurrence of individual state shifts, but would also ensure the hippocampus-dependence of these states.

References

Baxter MG, Holland PC, Gallagher M (1997) Disruption of decrements in conditioned stimulus processing by selective removal of hippocampal cholinergic input. Journal of Neuroscience 17(13):5230-6.

Benoit SC, Davidson TL, Chan KH, Trigilio T, Jarrard LE (1999) Pavlovian conditioning and extinction of context cues and punctate CSs in rats with ibotenate lesions of the hippocampus. Psychobiology 27(1):26-39.

Blanchard TC, Hayden BY, Bromberg-Martin ES (2015) Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. Neuron 85(3):602-14.

Bonardi C, Robinson J, Jennings D (2017) Can existing associative principles explain occasion setting? Some old ideas and some new data. Behavioural Processes 137:5-18.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychological review 108(3):624.

Bouton ME, Swartzentruber D (1986) Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. Journal of Experimental Psychology: Animal Behavior Processes 12(4):333.

Bouton ME. Context, ambiguity, and classical conditioning (1994) Current directions in psychological science 3(2):49-53.

Bouton ME, Nelson JB (1998) Mechanisms of feature-positive and feature-

negative discrimination learning in an appetitive conditioning paradigm. Occasion setting: Associative learning and cognition in animals 69-112.

Bouton ME (2004) Context and behavioral processes in extinction. Learning and memory 11(5):485-94.

Bouton ME (2007) Learning and behavior: A contemporary synthesis. Sinauer Associates.

Braver TS (2012) The variable nature of cognitive control: a dual mechanisms framework. Trends in cognitive sciences 16(2):106-13.

Casey BJ, Galvan A, Hare TA (2005) Changes in cerebral functional organization during cognitive development. Current opinion in neurobiology 15(2):239-44.

Chatham CH, Claus ED, Kim A, Curran T, Banich MT, Munakata Y (2012) Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. PloS one 7(2):e31546.

Corcoran KA, Maren S (2004) Factors regulating the effects of hippocampal inactivation on renewal of conditional fear after extinction. Learning & Memory 11(5):598-603.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends in cognitive sciences 10(7):294-300.

Delamater AR (2007) The Role of the Orbitofrontal Cortex in Sensory-Specific Encoding of Associations in Pavlovian and Instrumental Conditioning. Annals of the New York Academy of Sciences 1121(1):152-73.

Delamater AR, Kranjec A, Fein MI (2010) Differential outcome effects in Pavlovian biconditional and ambiguous occasion setting tasks. Journal of Experimental Psychology: Animal Behavior Processes 36(4):471.

Delamater AR, Garr E, Lawrence S, Whitlow JW (2017) Elemental, configural, and occasion setting mechanisms in biconditional and patterning discriminations. Behavioural Processes.

Domjan M (2005) Pavlovian conditioning: a functional perspective. Annu. Rev. Psychol. 56:179-206.

Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron 66(3):438-48.

Floresco SB (2015) The nucleus accumbens: an interface between cognition, emotion, and action. Annual review of psychology 66:25-52.

Fuhs MC, Touretzky DS (2007) Context learning in the rodent hippocampus. Neural Computation 19(12):3173-3215.

Gallagher M, McMahan RW, Schoenbaum G (1999) Orbitofrontal cortex and representation of incentive value in associative learning. Journal of Neuroscience 19(15):6610-4.

Gallistel CR, Fairhurst S, Balsam P (2004) The learning curve: implications of a quantitative analysis. Proceedings of the national academy of Sciences of the United States of America (36):13124-31.

Gershman SJ, Blei DM, Niv Y (2010) Context, learning, and extinction.

Psychological review 117(1):197.

Gershman SJ, Cohen JD, Niv Y (2010) Learning to selectively attend. In: 32nd Annual Conference of the Cognitive Science Society.

Gilbert PE, Kesner RP, DeCoteau WE (1998) Memory for spatial location: role of the hippocampus in mediating spatial pattern separation. Journal of Neuroscience 18(2):804-10.

Gilbert PE, Kesner RP, Lee I (2001) Dissociating hippocampal subregions: A double dissociation between dentate gyrus and CA1. Hippocampus 11(6):626-36.

González F, Quinn JJ, Fanselow MS (2003) Differential effects of adding and removing components of a context on the generalization of conditional freezing. Journal of Experimental Psychology: Animal Behavior Processes 29(1):78.

Gupta AS, van der Meer MA, Touretzky DS, Redish AD (2012) Segmentation of spatial experience by hippocampal theta sequences. Nature neuroscience 15(7):1032-9.

Haddon JE, Killcross S (2006) Prefrontal cortex lesions disrupt the contextual control of response conflict. Journal of Neuroscience 26(11):2933-40.

Haddon JE, Killcross S (2007) Contextual control of choice performance. Annals of the New York Academy of Sciences 1104(1):250-69.

Harris JA, Livesey EJ, Gharaei S, Westbrook RF (2008) Negative patterning is

easier than a biconditional discrimination. Journal of Experimental Psychology: Animal Behavior Processes 34(4):494.

Holland PC (1991) Acquisition and transfer of occasion setting in operant feature positive and feature negative discriminations. Learning and Motivation 22(4):366-87.

Holland PC (1992) Occasion setting in Pavlovian conditioning. In: The psychology of learning and motivation (Medin D, ed), pp 69–125. San Diego: Academic Press.

Holland PC, Lamoureux JA, Han JS, Gallagher M (1999) Hippocampal lesions interfere with Pavlovian negative occasion setting. Hippocampus 9(2):143-57.

Honey RC, Watt A (1998) Acquired relational equivalence: implications for the nature of associative structures. Journal of Experimental Psychology: Animal Behavior Processes 24(3):325.

Jarrard LE, Davidson TL (1990) Acquisition of concurrent conditional discriminations in rats with ibotenate lesions of hippocampus and of subiculum. Psychobiology 18(1):68-73.

Jarrard LE, Davidson TL (1991) On the hippocampus and learned conditional responding: Effects of aspiration versus ibotenate lesions. Hippocampus 1(1):107-17.

Jeffery KJ (2000) Plasticity of the hippocampal cellular representation of space.

Neuronal mechanisms of memory formation: concepts of long-term potentiation and beyond. In: Neuronal mechanisms of memory formation (C. Holscher, ed), pp 100–124. Cambridge: Cambridge University Press.

Jenkins HM, Sainsbury, RS (1970) Discrimination learning with the distinctive feature on positive or negative trials. In: Attention: Contemporary theory and analysis (D. Mostovsky, ed), pp 239–273. New York: Appleton-Century-Crofts.

Jezek K, Henriksen EJ, Treves A, Moser EI, Moser MB (2011) Theta-paced flickering between place-cell maps in the hippocampus. Nature 478(7368):246-9.

Karlsson MP, Tervo DG, Karpova AY (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science 338(6103):135-9.

Kamin LJ (1969) Predictability, surprise, attention, and conditioning. Punishment and aversive behavior 279-96.

Kemp C, Tenenbaum JB (2008) The discovery of structural form. Proceedings of the National Academy of Sciences 105(31):10687-92.

Kim JJ, Fanselow MS (1992) Modality-specific retrograde amnesia of fear. Science 256(5057):675-7.

Knierim JJ, Neunuebel JP (2016) Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics. Neurobiology of learning and memory 129:38-49.

Liston C, Watts R, Tottenham N, Davidson MC, Niogi S, Ulug AM, Casey BJ

(2006) Frontostriatal microstructure modulates efficient recruitment of cognitive control. Cerebral Cortex 16(4):553-60.

MacDonald CJ, Lepage KQ, Eden UT, Eichenbaum (2011) H. Hippocampal “time cells” bridge the gap in memory for discontiguous events. Neuron 71(4):737-49.

MacDonald CJ, Carrow S, Place R, Eichenbaum H (2013) Distinct hippocampal time cell sequences represent odor memories in immobilized rats. Journal of Neuroscience 33(36):14607-16.

MacLeod JE, Potter AS, Simoni MK, Bucci DJ (2006) Nicotine administration enhances conditioned inhibition in rats. European journal of pharmacology 551(1):76-9.

MacLeod JE, Bucci DJ (2010) Contributions of the subregions of the medial prefrontal cortex to negative occasion setting. Behavioral neuroscience 124(3):321.

MacLeod JE, Vucovich MM, Bucci DJ (2010) Differential effects of nicotinic acetylcholine receptor stimulation on negative occasion setting. Behavioral neuroscience 124(5):656.

Maia TV (2009) Reinforcement learning, conditioning, and the brain: Successes and challenges. Cognitive, Affective, & Behavioral Neuroscience 9(4):343-64.

Meyer HC, Bucci DJ (2014a) The contribution of medial prefrontal cortical regions to conditioned inhibition. Behavioral neuroscience 128(6):644.

Meyer HC, Bucci DJ (2014b) The ontogeny of learned inhibition. Learning & Memory 21(3):143-52.

Meyer HC, Putney RB, Bucci DJ (2015) Inhibitory learning is modulated by nicotinic acetylcholine receptors. Neuropharmacology 89:360-7.

Meyer HC, Bucci DJ (2016a) Imbalanced activity in the orbitofrontal cortex and nucleus accumbens impairs behavioral inhibition. Current Biology 26: 2834–2839.

Meyer HC, Bucci DJ (2016b) Setting the occasion for adolescent inhibitory control. Neurobiology of Learning and Memory.

Meyer HC, Chodakewitz MI, Bucci DJ (2016) Nicotine administration enhances negative occasion setting in adolescent rats. Behavioural brain research 302:69-72.

Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. Journal of Neuroscience 16(16):5154-67.

Miller EK (2000) The prefontral cortex and cognitive control. Nature reviews neuroscience 1(1):59-65.

Moreira RD, Bueno JL (2003) Conditional discrimination learning and negative patterning in rats with neonatal hippocampal lesion induced by ionizing radiation. Behavioural brain research 138(1):29-44.

Muller RU, Kubie JL (1987) The effects of changes in the environment on the

spatial firing of hippocampal complex-spike cells. Journal of Neuroscience 7(7):1951-68.

Muller R (1996) A quarter of a century of place cells. Neuron. 17(5):813-22.

Nadel L, Willner J (1980) Context and conditioning: A place for space. Physiological Psychology 8(2):218-28.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain research 34(1):171-5.

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

Pastalkova E, Itskov V, Amarasingham A, Buzsáki G (2008) Internally generated cell assembly sequences in the rat hippocampus. Science 321(5894):1322-7.

Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G (2003) Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. Journal of Neuroscience 23(35):11078-84.

Pickens CL, Saddoris MP, Gallagher M, Holland PC (2005) Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. Behavioral neuroscience 119(1):317.

Pubols Jr BH (1962) Serial reversal learning as a function of the number of trials per reversal. Journal of Comparative and Physiological Psychology 55(1):66.

Redish AD, Jensen S, Johnson A, Kurth-Nelson Z (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. Psychological review 114(3):784.

Rescorla RA (1969) Pavlovian conditioned inhibition. Psychological Bulletin 72(2):77.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2:64-99.

Rescorla RA (1972) Evidence for "unique stimulus" account of configural conditioning. Journal of Comparative and Physiological Psychology 85(2):331.

Ross RT, Orr WB, Holland PC, Berger TW (1984) Hippocampectomy disrupts acquisition and retention of learned conditional responding. Behavioral neuroscience. 98(2):211.

Rudy JW (2009) Context representations, context functions, and the parahippocampal–hippocampal system. Learning and memory 16(10):573-85.

Todd TP, Huszár R, DeAngeli NE, Bucci DJ (2016) Higher-order conditioning and the retrosplenial cortex. Neurobiology of learning and memory 133:257-64.

Schmajuk NA, DiCarlo JJ (1991) A neural network approach to hippocampal function in classical conditioning. Behavioral neuroscience 105(1):82.

Schmajuk NA, DiCarlo JJ (1992) Stimulus configuration, classical conditioning, and hippocampal function. Psychological review (2):268.

Schmajuk NA, Lamoureux JA, Holland PC (1998) Occasion setting: a neural network approach. Psychological review 105(1):3.

Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M (2003) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. Learning & Memory 10(2):129-40.

Schwarz G (1978) Estimating the dimension of a model. The annals of statistics 6(2):461-4.

Sutton RS, Barto AG (1998) Reinforcement learning: An introduction. Cambridge: MIT press.

Urcelay GP, Miller RR (2014) The functions of contexts in associative learning. Behavioural processes 104:2-12.

Wheeler DS, Amundson JC, Miller RR (2006) Generalization decrement in human contingency learning. The Quarterly journal of experimental psychology 59(7):1212-23.

Wiltgen BJ, Sanders MJ, Anagnostaras SG, Sage JR, Fanselow MS (2006) Context fear learning in the absence of the hippocampus. Journal of Neuroscience 26(20):5484-91.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81(2):267-79.

Yoon T, Graham LK, Kim JJ (2011) Hippocampal lesion effects on occasion setting by contextual and discrete stimuli. Neurobiology of learning and memory 95(2):176-84.

Xu C, Krabbe S, Gründemann J, Botta P, Fadok JP, Osakada F, Saur D, Grewe BF, Schnitzer MJ, Callaway EM, Lüthi A (2016) Distinct Hippocampal Pathways Mediate Dissociable Roles of Context in Memory Retrieval. Cell 167(4):961-72.