Title

Unconscious Associative Learning with Decoded Neuroreinforcement

Author

Cherkaoui, Mouslim

Publication Date

2022

UNIVERSITY OF CALIFORNIA

Los Angeles

Unconscious Associative Learning with Decoded Neuroreinforcement

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Psychology

by

Mouslim Cherkaoui

2022

ABSTRACT OF THE DISSERTATION

Unconscious Associative Learning with Decoded Neuroreinforcement

by

Mouslim Cherkaoui

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2022

Professor Jesse Rissman, Chair

The question of what role consciousness plays in memory is as old as modern memory research. While most studies have used techniques that briefly present the stimuli to study unconscious learning, these studies tend to show small effect sizes and are not always successful. This makes it difficult to compare the mechanisms of conscious and unconscious learning. As a result, a debate has ensued about whether learning the relationship between two events is possible without awareness. One thing proponents of both camps with opposite views on this matter agree on, is the need for better methods. This dissertation aims to test whether multi-voxel neuroreinforcement is one such method. Understanding unconscious learning is however only half of the problem, as conscious associative learning also needs investigation. To address these gaps, In Chapter 2, I show that unconscious neuroreinforcement can affect the relative status of a cue by affecting variability in choice behavior. In Chapter 3, I explore a form of associative learning called higher-order conditioning to understand how conscious learning unfolds, and show how two mechanisms that differ in their computational complexity can be

utilized to learn and guide behavior. Lastly, Chapter 4 investigated whether the more computationally complex form of learning can be supported unconsciously and found further evidence in support of variability on choice behavior as a result of neuroreinforcement. These chapters highlight a need for using a range of experimental techniques to elucidate the mechanisms of conscious and unconscious learning.

The dissertation of Mouslim Cherkaoui is approved.

David Clewett

Barbara Knowlton

Megan Peters

Melissa Sharpe

Jesse Rissman, Committee Chair

University of California, Los Angeles

2022

بسم الله الرحمن الرحيم

I dedicate this dissertation to

my mother and father.

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendices

List of Figures

List of Tables

Acknowledgements

Parts of this research were supported by the following:

Paulson Scholarship Fund
Graduate Research Mentorship Award
Graduate Summer Research Mentorship Award

First and foremost, I would like to thank my mother who raised me and my siblings as a single immigrant mother after my father passed away. My mother is the most generous, loving, and caring person I know. To her I am eternally grateful. I live to make her happy and proud. I thank my father for giving me the life that he dreamt for me. Even though my father passed away at a young age, his love for nature documentaries is one of my strongest memories of him and fostered in me a curiosity about the mind. I wish we had more time together. I thank my brother for his encouragement and wisdom, and for teaching me who I am and who I want to become. I also thank my wife Nicole in whom I could not have found a better partner in life. Being her husband is the greatest honor I will ever experience.

I thank my graduate advisor and committee chair Prof. Jesse Rissman for welcoming me to UCLA, and for allowing my curiosity to thrive. Prof. Hakwan Lau for mentoring, supporting, and advocating for me during his time at UCLA and after; his generosity seems to have no bounds. Prof. Barbara Knowlton for setting the groundwork on implicit learning which sparked my love for psychology research. Prof. Melissa Sharpe whose mentorship was necessary and sufficient, especially for Chapter 3 of this dissertation. Prof. David Clewett for his encouragement and support. Prof. Megan Peters, meeting her made me want to try to become just like her. Prof. Brian Odegaard for picking me up when I was almost ready to give up. Thank you also to Prof. Mitsuo Kawato, Prof. Sam Gilbert and Prof. Anette Ragó. Thank you to Dr. Cody Cushing for making it a joy to be his lab mate and for always helping when I got stuck; those days in Japan and leading up are my fondest memories of grad school. Prof. Vincent Taschereau-Dumouchel, Prof. Aurelio Cortese, Prof. Kiyofumi Miyoshi, Prof. Yujia Peng, Dr. Jeffrey D. Knotts. Dr. Taylor Webb, Raihyung Lee, Dr. Ben Rosenberg, Dr. Cathie Tsz So, Catherine Walsh, Mary Vitello, Samantha Walters, Fleming Peck, Dr. Seong Hah Cho, Dr. Piercesare Grimaldi, Prof. Jorge Morales, Sivananda Rajananda, Andrew Eastwick, Angela Clague, Shiyun Wang, Dr. Matthias Michel, Lisa Lee, Freya Vaughn, Mieko Hirata and Kaori Nakamura.

Vita

EDUCATION

MSc. Cognitive Neuroscience, University College London, London, UK, 2015

- Advisor: Sam Gilbert
- Grade: Distinction

BA. Psychology. Eötvös Loránd University, Budapest, Hungary, 2014

- Advisor: Anette Ragó
- Grade: Summa cum laude

PUBLICATIONS

Orouji, S., Taschereau-Dumouchel, V., Cortese, A., Odegaard, B., Cushing, C., **Cherkaoui, M.**, ... & Peters, M. A. (2022). "Task-relevant autoencoding" enhances machine learning for human neuroscience. *arXiv:2208.08478*.

Cherkaoui, M., Rissman, J., Lau, H., & Hampson, M. (2021). Ethical considerations for fMRI neurofeedback. In *fMRI Neurofeedback* (pp. 315-331). Academic Press.

Cherkaoui, M., & Gilbert, S. J. (2017). Strategic use of reminders in an 'intention offloading' task: Do individuals with autism spectrum conditions compensate for memory difficulties?. *Neuropsychologia*, *97*, 140-151.

FELLOWSHIPS & AWARDS

2022	Paulson Fellowship	($9,492)
2021	Paulson Fellowship	($3,512)
2020	Paulson Fellowship	($8,884)
2019	Paulson Fellowship	($6,000)
2019	NAS Colloquium, Travel Award	($800)
2018	Graduate Research Mentorship	($20,000)
2018	Graduate Summer Research Mentorship	($6,000)

Chapter 1

Background

Memory is distinct from consciousness (Lau and Rosenthal 2011; Block et al. 2014). But is conscious awareness required for creating memories? And if so, what role does consciousness play? These questions are not without merit. They are as old as modern memory research (Thorndike 1898; Lashley 1923). Identifying the computations that support conscious and unconscious learning is essential not only for cognitive neuroscience of learning and memory (Henke 2010; Frith 2019; Kouider and Dehaene 2007), but also for highlighting differences across populations, including differences associated with psychiatric and neurological disease (Adams, Huys, and Roiser 2016; LeDoux and Daw 2018; LeDoux and Lau 2020; LeDoux, Michel, and Lau 2020; Fanselow and Pennington 2018; Michel et al. 2019).

Neuropsychological evidence has uncovered that humans have several types of memory systems that are partially independent from each other and in which consciousness differs (Squire 1992). For example, procedural memory has a strong implicit (i.e., unconscious) component and episodic memory is accompanied by a conscious component (i.e., recollection). If our goal is to understand the role of consciousness in learning and memory, then it would not be always fruitful to use consciousness as a criterion to differentiate between these systems (Henke 2010).

Instead, one can acknowledge the difference between learning from information that one is not consciously aware of, from learning from information one is aware of but without explicit awareness of what has been learned (Block 2007). An example of the former is *unconscious* *learning* which involves encoding of unconscious stimuli that remain unconscious due to experimental (e.g., visual masking) or clinical conditions (e.g., blindsight). That is, the stimuli are never consciously experienced. Unconscious learning is different from *implicit learning*, which mostly concerns learning from consciously experienced stimuli, but without being explicitly aware of what has been learned. Examples of implicit learning include but are not limited to motor skill learning or artificial grammar learning. The distinction between unconscious learning and implicit learning is important because it implies that the *process of learning* can certainly occur unconsciously. To fully understand memory, we need to also understand the mechanisms of memory.

A source of much theoretical debate is to which extent (if at all) the *mechanisms of learning* can occur unconsciously. That is, how does one event inform another event? Is it through the automatic creation of an unconscious link between the events? Or does it require some conscious reasoning process?

Broadly, this debate is exemplified in two strong views that purport to explain the mechanisms of learning. The strong *associative view* argues that learning can be automatically acquired (i.e., without consciousness, effort or intention). For example, according to this view, fear or reward learning can be acquired through nonconscious associative representations that may automatically elicit an avoidant or approach response for fear or reward, depending on the circumstance (LeDoux and Pine 2016). The associative view implies that humans are susceptible to environmental influences that can operate unconsciously. The view is sometimes referred to as the dual-systems view because according to this view behavior is determined both by an automatic link-formation mechanism and a system that is based on propositional reasoning. The *propositional view* posits that associative learning does not result from automatic formation of unconscious links, but from the operation of a controlled conscious reasoning process. According to this view, it is not possible to have learned a relationship between two events without awareness of that relationship (Mitchell, De Houwer, and Lovibond 2009). Indeed, proponents of the propositional view write: “[...] a demonstration of unaware conditioning would be highly damaging to the propositional approach and would provide strong evidence for a second (automatic) learning mechanism” (Mitchell, De Houwer, and Lovibond 2009, p 189). What this debate shows is that delineating between conscious and unconscious components of learning is not trivial and requires careful methodological and theoretical discussion.

The existence of this debate might be surprising for scientists in related fields. Indeed, there are many examples of studies that suggest new associations can be learned unconsciously. Such examples include learning associations between subliminal stimuli (i.e., visually masked) and supraliminal (i.e., consciously experienced) reward/punishment (Seitz, Kim, and Watanabe 2009; Pessiglione et al. 2008; Raio et al. 2012; Scott et al. 2018), as well as between tones and odors during sleep (Arzi et al. 2012) in healthy human adults. Further support of the existence of a dual-system is the demonstration of robust fear conditioning to unseen visual stimuli in cortically blind patients with damage to the primary visual cortex (i.e., blindsight) (Hamm et al. 2003), as well as the demonstration of intact reward-based associative learning in an animal model of blindsight (Takakuwa et al. 2017; Kato et al. 2021).

And yet, despite this evidence, emerging studies suggest that unconscious learning may not be possible. For example, a recent systematic review concluded that unconscious fear conditioning studies are affected by methodological problems and publication bias (Mertens and Engelhard 2019; Shanks 2017). One such methodological problem is the need for sufficient awareness checks which many studies have failed to use (Vadillo, Konstantinidis, and Shanks 2016; Vadillo et al. 2021). Indeed, when another study used appropriate awareness checks they found evidence *against* the existence of unconscious instrumental conditioning (Skora et al. 2021). This leads to the question: if unconscious learning is so hard to demonstrate in healthy adults, why does it seem to be such a robust phenomenon in neuropsychological patients? The answer to this question might be explained by a difference in representational content, signal strength, and sensitivity of awareness checks.

To compare conscious conditioning with unconscious conditioning, we could compare a condition in which participants experience the stimulus consciously to a condition in which participants represent the stimulus unconsciously. The difficulty of such an approach is that the unconscious condition should not only have a very similar stimulus but should also register the same perceptual content in the brain. If the unconscious stimulus is weak, or non-existent, then the comparison would be representation vs. no representation. Instead, ideally, the comparison should be between the same representational content. To match representational content between conscious and unconscious conditions we need to match signal strength as much as possible (Lau 2008). Otherwise, the difference between conditions would be the difference between strong and weak sensory representations.

Visual masking techniques aim to suppress subjective awareness of a stimulus by for example flashing the stimulus extremely briefly in a procedure in which the stimulus is sandwiched in between two masks that serve to reduce the visibility of the stimulus. Modern visual masking techniques may not provide *sufficiently strong* signals to facilitate strong unconscious conditioning (Peters and Lau 2015; Knotts, Lau, and Peters 2018); even masking techniques that operate at higher levels of visual processing may not do so (Breitmeyer 2015; Kouider and Dehaene 2007). When masking techniques are used, conscious visibility may be abolished, but the internal signal may be abolished with it. That may mean that an awareness check on a truly unaware trial might reflect an abolished internal signal, not just abolished conscious visibility.

Even if there is an internal signal, then it is very likely that it will be weak (Knotts, Lau, and Peters 2018). Weak residual signals may cause small effect sizes which may be sufficient for simple associative learning (Greenwald and De Houwer 2017; Scott et al. 2018), but which may make it difficult to replicate and interpret more complex forms of learning such as those that require learning contingencies between stimuli and their outcomes and acting upon them (Mertens and Engelhard 2019; Skora et al. 2021).

Of course, there is no single perfect experiment that could be ethically conducted in humans that would be able to bypass the signal strength confound. However, blindsight suggests that given the right technological advancements such an experiment should be possible (Persaud et al. 2011), at least in terms of supporting that a sufficiently strong but unconscious signal can facilitate high levels of performance. Blindsight patients have damage to one hemisphere of the primary visual cortex. They claim to be blind in that visual field but nonetheless perform significantly above chance at discrimination tasks.

It turns out that facilitating a strong signal is not only important for understanding the role of consciousness in memory but also for understanding the brain mechanisms of how human memory works. If it was possible to incept a specific and sufficiently strong artificial brain signal without conscious experience, then such an experiment would satisfy a highly influential criteria for memory and would thereby constitute a significant leap toward understanding the brain mechanisms that support memory formation in humans (Martin and Morris 2002). Such tests require the deployment of appropriate conditioning procedures to sufficiently distinguish between different memory mechanisms. A key example of such procedures are the ones used in higher-order conditioning. These have mostly been used in computational neuroscience and associative learning literature to delineate between two mechanisms postulated in behavioral neuroscience theories. These theories and the procedures used to test them will be discussed further below.

In regards to the technical advancements necessary to demonstrate sufficiently strong unconscious signals, one technique stands out as promising. Real-time functional magnetic resonance imaging (fMRI) neurofeedback may provide a non-invasive means of establishing unconscious learning and may provide insight into where in the brain learning takes place. Real-time fMRI neurofeedback is a closed-loop procedure that merges data collection, analysis and stimulus presentation. The fMRI data is processed as they arrive and compared against a trained model (i.e., target brain activity). Feedback about the fMRI measurement is then communicated to the participant by updating the display.

One such method is decoded multi-voxel neuro-reinforcement (hereby referred to as neuroreinforcement). Neuroreinforcement applies multivoxel pattern analysis (MVPA) to fMRI data online (i.e., in real-time). Importantly, while in most fMRI neurofeedback studies participants are informed of the nature of feedback and are instructed about how to maximize it, neuroreinforcement can occur without instruction (Muñoz-Moldes and Cleeremans 2020). This renders awareness checks unnecessary. Unbeknownst to neuroreinforcement participants, the feedback corresponds to how well their brain pattern (in the present) matched their target brain pattern.

This protocol has been used to demonstrate strong unconscious perceptual learning and the brain mechanisms supporting it (Shibata et al. 2011; Amano et al. 2016). Furthermore, it has been suggested that neuroreinforcement might allow pairing of specific target brain patterns with value (Taschereau-Dumouchel et al. 2018; Chiba et al. 2019; Cortese et al. 2021). This may allow causal insights into the relationship between neural changes and modified behavior

(Watanabe et al. 2017). Because neuroreinforcement can proceed without instruction, the participant *and the experimenter* can be kept unaware of the nature of training, thereby allowing double-blind protocols.

Neuroreinforcement therefore represents a unique opportunity not only to test the functional (i.e., behavioral and neural) consequences of forming unconscious associations, but also allows doing so with a high degree of neural and behavioral specificity, and experimental rigor.

The idea that behavior can be determined by two independent systems is long held (Evans 2008). Many frameworks argue that one system is unconscious, reactive and fast (i.e., associative), while another system is conscious, prospective but slow (i.e., propositional). This idea has failed to constrain the mechanisms that it purports to explain (Krajbich et al. 2015). Instead, cognitive neuroscience looked to machine learning in order to draw more precise quantitative formalizations of behavior (Daw, Niv, and Dayan 2005), and to associative learning procedures to test the behaviors predicted by these quantitative formulations. As we shall see below, this computational framework has been useful to constrain the mechanisms of learning, but it remains silent on the role of consciousness in learning. Here, I seek to draw from key concepts within this formalization to map out the boundaries of the mechanisms of unconscious learning.

The main benefit of this framework is that it, to a very useful extent (Daw 2018; Collins and Cockburn 2020) formalized a well-known theory from behavioral neuroscience (Balleine and Dickinson 1998), namely the distinction between habitual and goal-directed behaviors. Habitual behaviors are acquired through automatic links between stimuli and responses so that presentation of the stimulus automatically elicits a habitual response. Computationally, habitual behaviors arise through *model-free* learning in which actions are evaluated through trial and error. Model-free behaviors are considered habitual because they persist even if the desirability of the outcome has changed (i.e., through devaluation of the outcome). In contrast, *model-based* learning produces flexible behavior that adapts to devaluation by using models of the world to evaluate possible actions and anticipate future outcomes (Doll, Simon, and Daw 2012). Therefore, model-based learning makes complex demands which are computationally more expensive than model-free learning.

One important note is that while model-free and model-based learning might sit at opposite ends of the efficiency-flexibility trade-off (Gershman 2018), they are not synonymous with habitual and goal-directed actions. Furthermore, learning is likely governed by more than two systems (e.g., Pavlovian model-based learning or successor-representations) (Dayan and Berridge 2014; Ida Momennejad 2020), that can operate in parallel (Collins and Cockburn 2020). Nevertheless, this dichotomy can be relied on as a quantitative framework to map out the boundaries of unconscious learning (Frith 2019), which is what this dissertation investigates.

I argue here that any conclusions about unconscious learning will require clear and thorough investigation of the associative structure of higher-order conditioning, namely second-order conditioning, and sensory preconditioning. This is because while there are many demonstrations of unconscious learning, albeit weak, those procedures have been confined to more simple Pavlovian and instrumental conditioning. To properly understand the role of consciousness in learning, we need to also investigate the associative structure of higher-order conditioning and then apply a computational framework to its mechanisms.

Dissertation overview

This dissertation is divided into 5 chapters. The second chapter is a double-blind study aimed at establishing the efficacy of multi-voxel neuroreinforcement as a tool for unconscious learning. In this chapter, participants underwent real-time fMRI to test whether activations of nonconscious visual representations (e.g., chairs, fish) could acquire a preference as demonstrated in a forced choice task. Participants preferences were assessed after neuroreinforcement using a two-alternative-choice task. The main hypothesis was that participants will show a specific and significant preference for the category associated with the brain activity pattern induced during neuroreinforcement. This category will be referred to as the Target category. It is hypothesized that preference will manifest at the group-level as a tendency to choose the Target category more often than the control categories in the forced-choice task.

Chapter 3 investigates the associative structure of higher-order conditioning in a large sample of online participants. This chapter aimed to better segue the results of animal and human studies of higher-order conditioning in order to ultimately understand the role of consciousness (Lovibond and Shanks 2002). Chapter 3 is divided into three experiments. In the first experiment, extinction was used after acquisition to selectively extinguish a previously rewarded stimulus to test whether that affected responding to its associate stimulus (i.e., second-order cue or preconditioned cue). In the second experiment devaluation was used to devalue the outcome associated with one of two previously rewarded stimuli. Like in the extinction experiment, this experiment used a forced-choice task. The second experiment aimed to test whether a cue indirectly associated (i.e., second-order cue or preconditioned cue) with the outcome was affected by devaluation. In the last experiment of Chapter 3, the value of the indirectly associated cues were investigated using a conditioned reinforcement test to better understand higher-order conditioning within the computational framework discussed above.

Chapter 4 tested whether unconscious learning through multi-voxel neuroreinforcement can support inference in sensory preconditioning. Here participants underwent a task aimed at learning to associate a target category with an associated cue to form a stimulus-stimulus representation that can support model-based inferences.

Chapter 2

Establishing the efficacy of decoded neuroreinforcement for unconscious learning

Introduction

Understanding the potential functions of consciousness will require delineating the mental acts that depend on it from those that do not (Kouider and Dehaene 2007; Dehaene, Lau, and Kouider 2017). While it is largely undisputed that unconscious *processing* exists, it is unclear whether unconscious learning can succeed at higher levels of complexity, such as learning how actions are shaped on the basis of the history of past rewards or on higher-order associations. Remarkably, while there is evidence that such learning appears to proceed in healthy adults when the cues are presented below consciousness threshold (Pessiglione et al. 2008), mounting evidence has recently indicated otherwise (Skora et al. 2021; Mertens and Engelhard 2019).

Unconscious learning is often investigated with methods that hide visual stimuli from awareness (i.e., masking or flash suppression) (Kouider and Dehaene 2007). Masked presentations are considered unconscious when participants report complete unawareness or perform at chance on tasks that require them to correctly identify or discriminate the masked stimulus (A. G. Greenwald, Klinger, and Schuh 1995). If participants report complete unawareness of the masked stimuli (i.e., d’ = 0) and the stimulus nevertheless affects behavior, then unconscious learning is thought to have occurred.

There are two main issues with this approach. First, rendering a stimulus unconscious with masking techniques will inevitably cause the stimulus strength to be weak (H. C. Lau 2008). Therefore, any behavior that may be tested will also be weak as a consequence, which may lead to very small effect sizes. Second, appropriate awareness checks require that discrimination is at chance (i.e., d'=0), non-significant from chance is not enough (Vadillo, Konstantinidis, and Shanks 2016). This makes any unconscious learning study employing masking techniques difficult to interpret and extend.

Here, I will attempt to rectify these issues using an emerging innovative fMRI technique called multi-voxel neuroreinforcement (neuroreinforcement). This technique pairs an implicit activation of a distributed (multi-voxel) brain response pattern with reward in real-time closed-loop neuroreinforcement (Figure 1.1). In this case, I will use category-specific targets (i.e., animals and objects) as stimuli. One of these categories (e.g., cars) will be designated the target for intervention. On each trial of neuroreinforcement induction training, the target category will be decoded from the ventral temporal cortex. Importantly, participants are kept unaware of the purpose of training and are thinking about any variety of non-related things during training. This allows us to subvert the issue of awareness checks altogether (Vadillo, Konstantinidis, and Shanks 2016; Muñoz-Moldes and Cleeremans 2020). In fact, even the experimenter can be kept unaware of the target of training, enabling a double-blind protocol, thereby meeting high levels of rigor.

The target brain activity pattern is determined with MVPA to identify a distributed pattern of neural representations (Haxby et al. 2001; Cohen et al. 2017). MVPA looks at patterns of activation across a set of predefined voxels to train weights for each voxel in order to estimate a 'classifier' (e.g., likelihood that the participant was looking at cars). During each trial of neuroreinforcement training, the participants' real-time fMRI activity is passed through this classifier. The classifier then estimates a likelihood that the fMRI activity represents the target (i.e., is the participant's brain 'representing' cars?). If the classifier estimates that the likelihood is high, the participant receives high feedback and reward.

While the mechanisms of neuroreinforcement are not yet fully understood (Shibata et al. 2019; Chiba et al. 2019), early results suggest that injecting value into neural representations of specific relevant features can help extract those features to guide inference-based decision-making (Cortese et al. 2021). It is therefore a promising alternative to test the extent of unconscious learning. Here, I will test whether unconscious learning can support model-free actions.

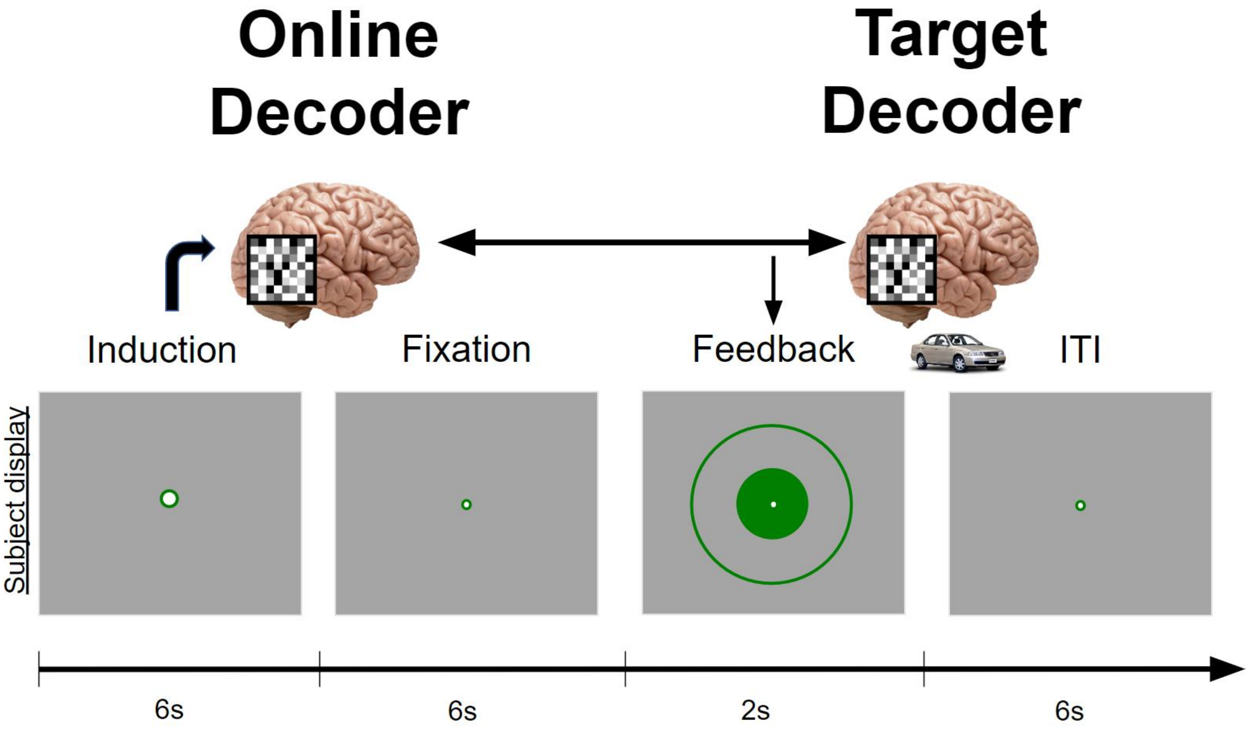

Figure 1.1 **Neuroreinforcement procedure**. Online decoding is used to reinforce the target (i.e., cars in this example). The size of the disc is proportional to the likelihood of the target being represented, and to monetary gain. Brain activity pattern representing cars is endowed with value.

Methods

Participants: Five participants were recruited for the study (***Mean Age*** = 20.8, ***Female*** = 4). The expected effect size based on a review of the decoded neurofeedback literature indicated that the average effect was 1.02. The experiment that most matches our design had a behavioral effect of 1.07 (Amano et al. 2016). We therefore expected an effect size of around 1.00. 16 participants are needed to obtain such an effect with a two-tailed paired samples t-test. However, due to the pilot experiment being halted, we only obtained data from five participants. The research was approved by the Institutional Review Board at University of Florida and all participants provided written informed consent.

Participants underwent a phone screening to determine eligibility. Participants were selected for recruitment if they reported more than four categories as “neutral” on a 7-point Likert preference scale, ranging from “strongly dislike” to “strongly like”. No participants were excluded based on these criteria. The mean preference rating was 4.2.

Blinding: The experiment was conducted in double-blinded fashion, i.e., neither the participants nor the experimenters were aware of the target category of the neuroreinforcement procedure. Blinding was obtained by spatially separating the experimenters conducting data collection from experimenters building the decoders; data-collection was conducted at the University of Florida in Gainesville, Florida, US, and the experimenters building the decoders were in Los Angeles, CA, US. Only the decoder builder knew what the target and control categories were for each subject, and only the data collection experimenter collected the data. Blinding was confirmed at debriefing.

Materials: The stimuli are images of 10 object categories (e.g., chairs, cars) and 30 animal categories (e.g., horses, fish, snakes). These stimuli were collected from various databases on the internet and have been demonstrated in previous research to elicit a strong signal in the ventral temporal cortex (Taschereau-Dumouchel et al. 2018). The images are naturalistic photos depicting the stimuli categories.

The 3,600 images (90 images per category) were collected from sources including from the Creative Common initiative (https://creativecommons.org), images from Pixabay marked for commercial use and modifications (http://pixabay.com), images from Flickr allowing commercial use and modifications (http://www.flickr.com), and images from Shutterstock (http://shutterstock.com). The images were selected if they presented a full frontal view of the object or animal and if no other objects were clearly identifiable in the background. Images were cropped so that they would frame the object. The final images were 533 X 533 pixels and covered 13.33 degrees of visual angles during the procedure. The average contrast and luminance of images were not different between categories (Taschereau-Dumouchel et al. 2018).

Decoder Construction: Before neuroreinforcement, participants underwent a one hr fMRI decoder construction session during which they were presented with 3,600 pictures in chunks of two, three, four or six images of the same category (Figure 1.2). The sequence was pseudo-randomized and fixed across participants. To optimize attention, participants were asked to perform a 1-back task in which they reported when the image category changed with a button press.

In order to allow high-pass filtering of the fMRI data, chunks within each category were organized so that their period was smaller than 120 seconds. The 30 different animal categories included reptiles (snake and gecko), amphibians (frog and turtle), insects (cockroach, beetle, ant, spider, grasshopper, caterpillar, bee, butterfly, and fly), birds (robin, peacock, and chicken), annelids (earthworm), mammals (mouse, guinea pig, bat, dog, sheep, cat, rabbit, horse, and giraffe) and aquatic animals (shark, whale, common fish, and dolphin). The database also included 10 categories of human-made objects (airplane, car, bicycle, scissor, hammer, key, guitar, cellphone, umbrella, and chair).

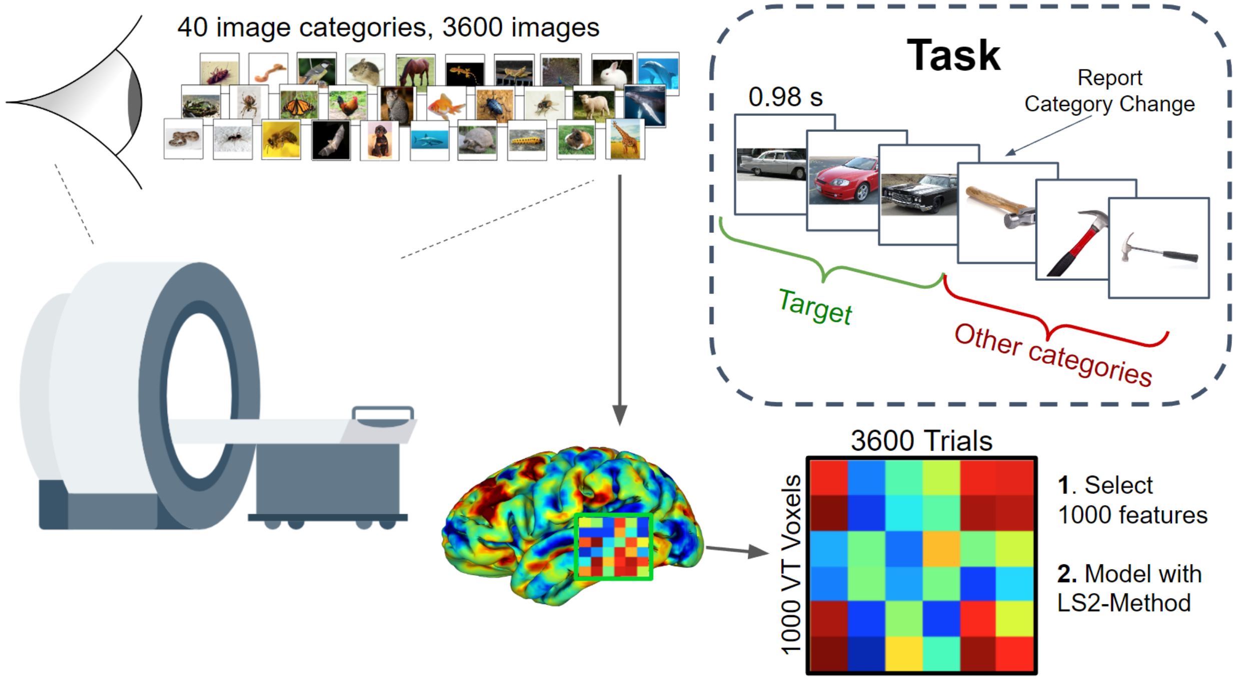

Figure 1.2 **Decoder construction**. Task: All participants viewed 3,600 images in the scanner and were instructed to press a button each time the category changed. Each image was presented for .98 seconds. An F-test selected the 1,000 most relevant voxels in the ventrotemporal cortex that accounted for the most variance elicited by each of all the image categories. Single trial estimates were generated using the least-square 2 method (Turner et al. 2012) to model the rapid-event related design.

Trials were organized into six runs of 600 trials each, interleaved with short breaks. Each trial lasted for .98 seconds during which an image was presented. There was no inter-trial-interval (ITI). This rapid event-related design was used to maximize the number of images each participant viewed.

The fMRI data for decoder construction were processed using SPM (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm) and custom python scripts using pyMVPA and sklearn packages (Hanke et al. 2009; Pedregosa, 2011.) The six runs of the task were concatenated and preprocessed in SPM using default parameters. The first image of the first run was used as reference for realignment and segmented into tissue classes. Anatomical and functional data were coregistered using the gray matter image from segmentation as a reference. Motion was then regressed out of the functional data using the parameters from realignment. Single-trial estimates were then generated with pyMVPA using the least-squares 2 method (LS-2) (Turner et al. 2012) in which a separate GLM is computed for each trial where the current trial is assigned to one regressor while the remaining trials are equally split between two “rest” nuisance regressors.

A 6-fold cross-validation (CV) procedure was used to determine area under the receiver operating characteristic curve (AUC) for binary classification. Each participant’s fMRI data were loaded and masked to the ventral temporal (VT) area in their native space using anatomical masks from Freesurfer parcellations of the fusiform, lingual, parahippocampal, and inferior temporal areas (Fischl et al. 2004). Single-trial parameter estimates were standardized by feature within subject and within each of the six task runs. Since each participant completed six runs, the data were split into six folds for training and testing. For each CV split, the withheld testing set consisted of all the data for one of the six task runs. To avoid overfitting, the remaining preprocessing was done using only the training data. Based on previous work (Taschereau-Dumouchel et al. 2018), 1,000 VT voxels were selected via F-test to select which voxels accounted for most variance elicited by all image categories across all training trials, in order to increase the number of training samples to features. A Sparse Multinomial Logistic Regression was then trained to perform binary (target vs. all-non-target) classification between the target category and all categories rated as “4; Neutral” on the pre-scan preference rating scale (Krishnapuram et al. 2005). The weights were transformed to match Nifti format. AUC scores were calculated for each CV split based on classifier estimates.

The target category was selected as the category with the highest AUC scores across all six CV splits. This choice was made to optimize the potential for learning and was the same for all participants. The squared Euclidean distance between the target and potential control categories was calculated; all the trials of the target category were averaged together, then all the trials for the potential control categories were averaged together, and the distance between those patterns were computed. Two control categories were chosen based on two criteria 1) the categories always were within the same animate/inanimate distinction as the target category and 2) the control categories were chosen as the two with the highest ranked distance from the target category. This was to further limit the potential for overlap between the target and the two control categories.

Procedure*:* The experimental timeline is presented in Figure 1.3. The experiment involved two phases. 1) Induction phase and 2) Test phase. Pre and post-task preference ratings for the categories were collected before and after the experiment in addition to the phone-screening preference ratings. Participants were presented with a sheet to rate their preference for each of the categories (responses were scaled from 1 = “Extremely/strongly dislike” to 7 = “Extremely/strongly like”). The categories rated ‘4’ (“Neutral”) were chosen to be used as stimuli for the Choice task, for which the final selection criteria was based on farthest Euclidian distance from the Target, specified under “*Decoder Construction”*.

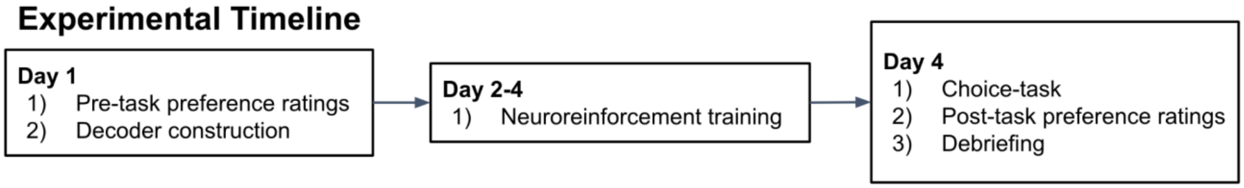

Figure 1.3 **Experimental timeline** for *Establishing the efficacy of decoded neuroreinforcement for unconscious learning*.

Induction phase (Neuroreinforcement): Following Decoder Construction, participants came back on a later day for neuroreinforcement sessions. There were 2-3 days between decoder construction and neuroreinforcement, determined by scanner and participant availability. Using multi-voxel neuroreinforcement, activation of the Target category was paired with reward. While in the scanner, participants were told to “do whatever you can" to maximize the size of a subsequently presented feedback disc - the larger the disc, the greater the reward on that trial. Each trial began with six seconds or rest, followed by a six second induction period, during which participants were instructed to “try to activate a pattern in their brain” to maximize the size of a subsequently presented feedback disc (Figure 1.4; Induction phase & Neuroreinforcement) (i.e., the diameter of the inner green circle). Participants were informed that their monetary gain was a function of the overall success in correctly activating brain patterns during each session. Critically, they were not told what the target multivoxel pattern represents.

This method has previously been used to reinforce nonconsciously presented image categories (e.g., snake) (Taschereau-Dumouchel et al. 2018). The feedback was calculated in real-time and was based on the output of the individual decoder constructed on Day 1. Participants went through three days of neuroreinforcement.

There was a 20 second extended rest period at the start of each neuroreinforcement run while image reconstruction processing caught up to real-time. Then an additional 10 seconds were

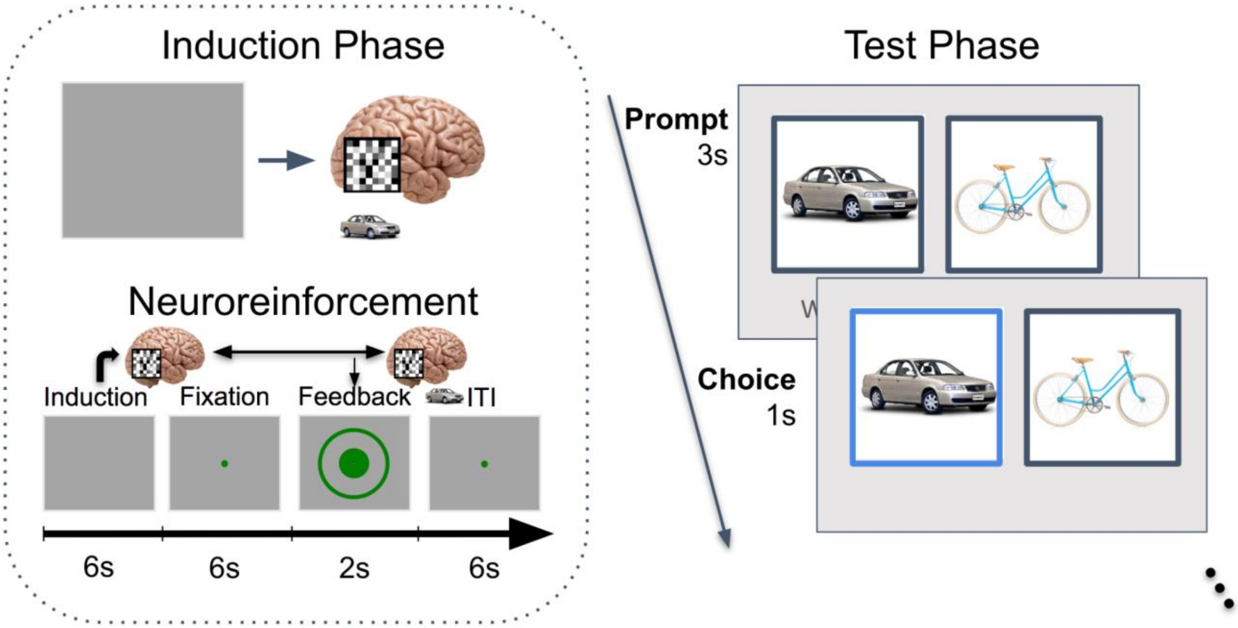

Figure 1.4 **Procedure** for *Establishing the efficacy of decoded neuroreinforcement for unconscious learning*. Participants underwent three days of neuroreinforcement which aimed to pair a Target category (cars in this example) with reward. On the last day, participants completed the Test phase in which they played the choice task where they had to pick the stimulus (or stimuli) they thought would lead to a reward. Participants used the buttons associated with the right and left stimulus on the screen. A blue square surrounded the chosen option but no feedback as to the outcome of their choice was provided.

collected to determine baseline BOLD activity levels. This was followed by 16 trials of neuroreinforcement per run.

Induction and Feedback phases were separated by a six second period that allows accounting for hemodynamic delay and to perform online decoding on data from the time window corresponding to the Induction phase. During this period, real-time decoder output was calculated during a six second period and the decoder output was displayed to the participant as a green disc for two seconds. The size of the disc corresponded to the decoder’s likelihood estimate. That is, a likelihood of 100% was associated with a maximum size disc and a likelihood of 0% associated with no disc displayed. There was a visual boundary indicating the maximum likelihood.

The average size of the disc during a run determined the amount of bonus reward participants received. Participants could earn a potential bonus of up to $15 on each day. For example, if a run finished with an average score of 50% then the bonus for that day would be $7.50. There was also an additional bonus to motivate participants further which was given when participants were generated a feedback score of 80% or more for three trials in a row. These high-score streak bonuses were an additional $1 and were visually indicated by the feedback disc turning blue with a written message (i.e., “High-Score!”) alerting participants to the high-score streak bonus.

During each trial, real-time data was realigned to match the coordinates of the Decoder Construction session by realigning to the first EPI image acquired during this session. Motion correction was then applied using an in-house real-time neurofeedback toolbox. Polynomial trends were removed from the activity of individual voxels and the signals were standardized voxel-wise using a 20 second waiting period at the beginning of each block. The induction likelihood of the Target was then computed based on the average multivoxel pattern during the 6 seconds corresponding to the Induction phase. This value was displayed back to the participants using a disc with a radius that is a function of the computed likelihood. The feedback disc was presented inside a ring indicating the disc’s maximum possible size. At the end of each block, participants were informed of the monetary gain associated with their performance during the neuroreinforcement run.

Test phase: Following induction, participants underwent the Test phase which involved a forced choice task. Each trial represents a two-alternative forced choice between two category stimuli presented to the right and left of the display. The participants were instructed to “pick the category or categories” they think are most likely to be rewarded. Each trial lasted for four seconds during which the stimuli were immediately presented and participants had three seconds to record their response. A blue square surrounded the chosen category after participants made their response as visual feedback to their choice. No feedback was given in regards to whether the chosen option was a winning one or not after each trial. The trial order and location of the categories on the screen were randomized and counterbalanced for each participant.

To account for hemodynamic delay, an ITI ranging from three to six seconds and averaging four seconds was introduced. A fixation cross was presented at the center of the screen during the ITI. An array of ITIs of equal length to the number of trials was shuffled between-participants and each ITI was assigned to each trial. Each ITI was sampled from an exponential-like distribution. There were 32 trials per condition and 96 trials in total. The Test phase run lasted for 12 minutes and 48 seconds.

Conditions: The choices in the Test phase were between A) Target Category (i.e., category stimuli that represents the target brain activity pattern previously trained during induction) vs. Control Category 1, B) Target Category vs. Control Category 2, and C) Control Category 1 vs. Control Category 2. Two control categories were chosen in order to reduce the possibility of false positive results in the comparisons between the target and each of the controls, and to test whether the two control conditions would be picked equally when pitted against each other in order to establish an active baseline; each control should be chosen around 50% of the time on average across participants.

MRI Parameters: Participants were scanned using a 3T MRI scanner (Prisma Siemens) with a 32-channel head coil at the University of Florida McKnight Brain Institute. 60 interleaved slices were obtained (TR = 1500 ms, TE = 30 ms, voxel size = 2.5 × 2.5 × 2.5 mm 3, field-of-view = 240 x 240 mm, slice thickness = 2.5 mm, 0 mm slice gap, flip angle = 70 deg) oriented parallel to the AC-PC plane, which covers the entire brain. Also, T1-weighted MR images were obtained (MP-RAGE; 256 slices, TR = 2300 ms, TE = 2.29 ms, voxel size = 0.9 × 0.9 × 0.9 mm, field-of-view= 240 x 240 mm, slice thickness = 0.94 mm, 0 mm slice gap, TI = 900 ms, flip angle = 8 deg).

Behavioral analysis*:* Choices were extracted using custom scripts in MATLAB for the following categories: Target, Control 1, and Control 2. The analysis was conducted using JASP. One sample t-tests with 50 as the criteria were used to test for choice bias between the Target vs. Control 1, and the Target vs. Control 2, and between the two controls. The pooled average was also computed by calculating the ratio of times the Target was chosen over the two controls in the two conditions with the Target present. Paired samples t-tests were conducted between the pooled condition and the controls condition to determine whether neuroreinforcement led to an increased tendency to choose the target. To test whether there was a specific increase in self-reported preference following neuroreinforcement, paired t-tests were performed between the averaged score of the phone-screening preference questionnaire combined with the pre-experiment preference questionnaire and the post-experiment preference questionnaire on the Target category and the neutral controls. The results from the preference questionnaires were inconclusive and will therefore not be reported.

Results and Discussion

The results are presented in Figure 1.5. On average, participants chose the Target 52.2% (SD = 9.7) of the time compared to the pooled average of the two control stimuli (one-sample t-test; $t(4) = 0.50$, $p = 0.32$; i.e., Target vs. Controls). Critically, a paired-samples t-test revealed no significant difference between Target vs. Controls and Controls ($t(4)$ .51, $p = .64$). These results suggest that the neuroreinforcement procedure did not lead to a specific and significant preference for the stimuli that represent the target pattern induced during neuroreinforcement at the group-level. With these results and due to an ongoing pandemic, it was determined to halt further data collection. The experiment was therefore highly underpowered, all results are mainly exploratory analysis, and the small sample size should be taken into account.

Although the number of trials were the same between all three experimental conditions, compared to the control condition there was more variation in choice consistency in the conditions with the Target. The presence of the Target category during a trial led to numerically larger standard deviations (*Target vs. Control 1:* SD = 38.95; *Target vs. Control 2*: SD = 31.98) compared to the Controls condition (SD = 4.51). This suggests that the Control condition was appropriately balanced. Importantly it also suggests that although the neuroreinforcement procedure did not instill preference bias in the Target category in such a way as to manifest as an increased tendency to choose it over the controls during the choice task, it had the effect of changing the *relative status* of the Target category compared to the control categories.

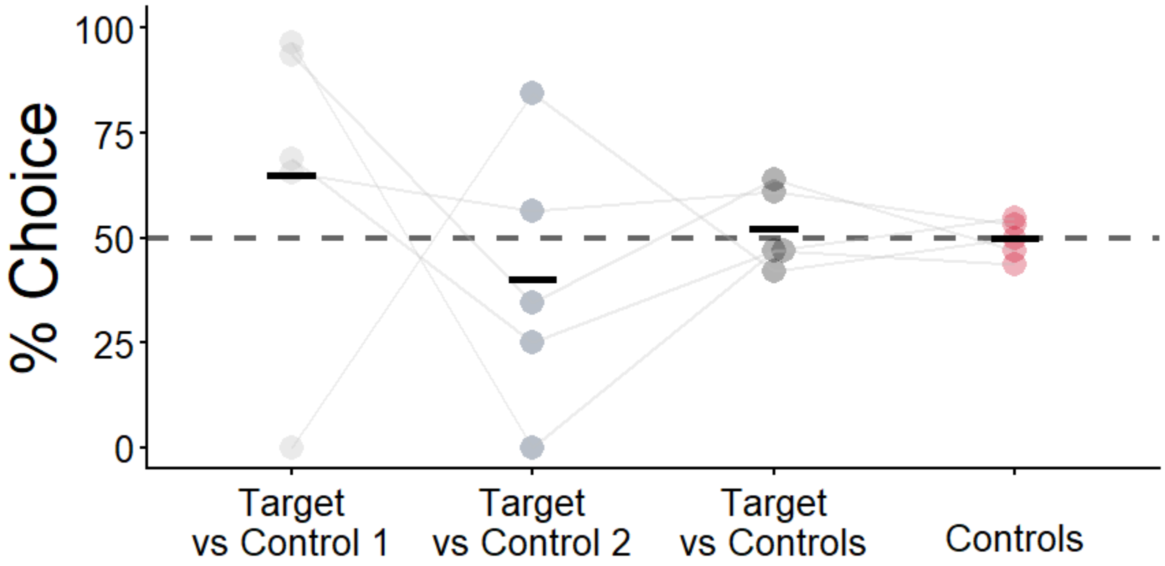

Figure 1.5 **Choice task results.** Dots indicate individual participants. Center lines represent means for each condition. Note: small sample size.

The average induction likelihood increased numerically across days however the increase was not significant ($F_{(2,12)} = 0.125$, $p = .883$, $\eta^2 = .020$). Planned contrast tests also confirmed that the induction likelihood did not increase significantly between individual days (ts(12) < .498, ps > .628). Individual day induction likelihood was not significantly above 50% on Day 2, significantly above 50% on Day 3, and marginally significant on Day 4 (***Day 2:*** M = 59.9, SD = 16.6, t(4) = 0.126, p = .126; ***Day 3***: M = 63.2, SD = 13.8, t(4) = 0.213, p = 0.05; ***Day 4*****:** M = 65.3, SD = 19.9, t(4) = 1.714, p = .08) suggesting participants learned to activate the target pattern but that they did not significantly improve across days (Figure 1.6). None of the participants reported that they had the target color in mind during induction at debriefing (Supplementary Table 1.1). Participants were also not above chance when guessing which of the three stimuli was the target.

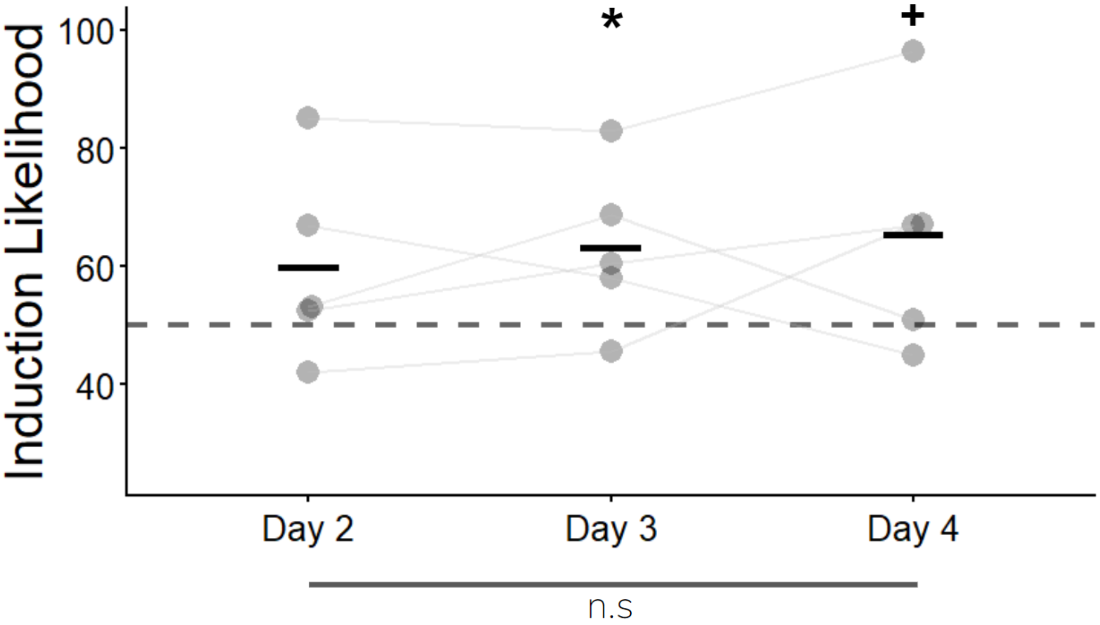

Figure 1.6 **Induction likelihood results.** Each dot represents one participant. The lines represent mean induction likelihood per day. * indicates significant at .05. **+** indicates marginally significant (p < .10). Note: small sample size.

While the mechanisms of neuroreinforcement are not well understood it is reasonable to assume that an increase in induction likelihood across days should accompany an increased tendency to choose the target (Chiba et al. 2019; Oblak, Lewis-Peacock, and Sulzer 2017). The relationship between the induction likelihood and choice was therefore investigated. This analysis revealed that at least numerically, there does not appear to be such a relation (Figure 1.7). Across all three days of training, the relationship between the tendency to choose the target over the two controls seems *negatively* related to induction likelihood, in this small sample size (***Day 2:*** r = -528, p = .360; ***Day 3:*** r = -.062, p = .921; ***Day 4:*** r = -.476, p = .418).

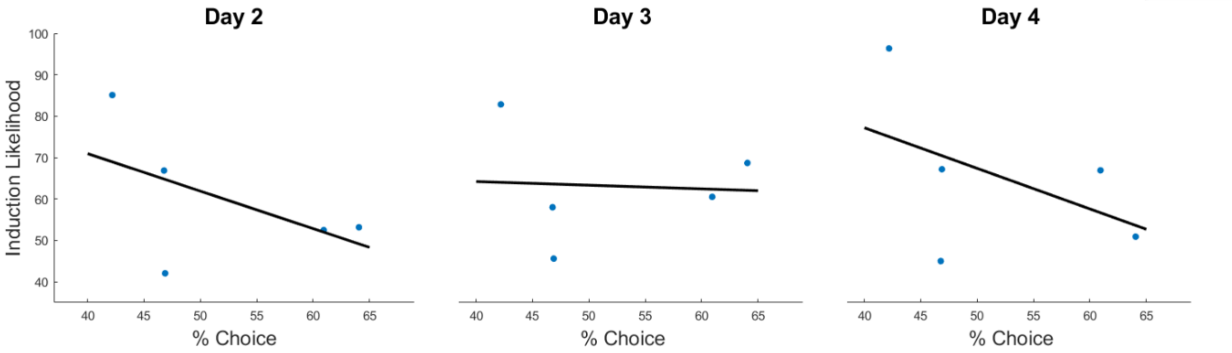

Figure 1.7 **Choice-induction likelihood correlations.** Each dot represents one participant. The lines represent the fitted regression line. Note: small sample size.

This experiment attempted to establish the efficacy of unconscious learning through neuroreinforcement. Although the sample size was very small, it appears that the presence of the Target category in a trial caused substantially more variance in choice compared to when only the two control categories were present. A pattern emerged which suggests that participants picked the Target in one condition but then switched their choices in the other condition and instead picked the Control category and vice versa. That is, the correlation between Target vs. Control 1 and Target vs Control 2 was marginally significant (r = -.875, p =

.052). Neither condition was correlated with the Controls' condition (r > .059, p > .924) Importantly, this pattern did not emerge in the Controls condition. This suggests that neuroreinforcement has an effect that changes the relative status of the Target category to the Control categories.

Previous findings indicated that nonconscious neuroreinforcement does not alter subjective fear experiences, despite successful reduction in physiological threat response (Taschereau-Dumouchel et al. 2018). A simulation study investigating Taschereau-Dumouchel et al.'s results found that the effect can be primarily attributed to exposure rather than reinforcement learning (Chiba et al. 2019). Hence, it may be that the kind of learning that occurred in this experiment was more effective at targeting physiological functions than behavior, which may explain why the behavioral effect did not materialize as hypothesized. This is consistent with higher-order theories of consciousness in which subjective mental experience, behavioral responses, and physiological reactivity operate via different mechanisms (Taschereau-Dumouchel et al. 2022; Brown, Lau, and LeDoux 2019). This discordance and its potential desynchrony can be probed within contexts similar to the present study by systematically measuring pupil dilation, eye-gaze and behavior to track learning over time. Indeed, eye-tracking studies reveal that Pavlovian conditioning in humans may be underpinned by at least two separate learning processes that operate in parallel, one controlling eye-gaze, and another controlling pupil dilation (Pool et al. 2019).

Along those lines, computational modeling can be introduced as a powerful tool to provide a detailed account of how learning processes contribute to multiple forms of Pavlovian processes (O'Doherty et al. 2021), and to understanding how those processes may differ in conscious and unconscious learning. A recent study showed that neuroreinforcement could enhance abstraction (Cortese et al. 2021). Abstraction is the process of extracting relevant lowdimensional representations from high-dimensional instances and is critical for generalization in inference-based decision-making. In that study, participants were reinforced to evoke neural representations of specific stimulus features which increased the tendency to use corresponding abstract representations for decision making. As the authors of the study and others have argued, value itself may subserve a critical function in constructing abstractions (Hayden and Niv 2021).

The ability to choose between options implies that information regarding the options' values must be available, consciously or unconsciously in the brain. The question of how this value is scaled and realized in the brain has been central to the field of neuroeconomics. The goal of the present study was not to identify where in the brain this value computation occurs, but to establish whether unconscious representations can be imbued with value. Indeed, it is not controversial whether people know if they prefer apples to oranges; that decision may be made without consulting an internal "common-currency" value signal. Many alternative explanations for these behaviors have been made that do not require an internal, neural, value scale in the brain. The present experiment used a decoding procedure that does not assume a common-currency valuation stage. That is, the decoder on each trial of neuroreinforcement asked: "is the target more likely compared to the other categories?". Alternatively, it could have also asked "is the target more likely compared to a specific alternative category?". This distinction may be subtle but potentially important as it might give clues about why there was more choice-variability in the Target conditions than in the Controls condition.

That is, after neuroreinforcement, on each trial participants chose between the Target or one of the two Controls. That choice may be fueled by value *per se*, and/or it may be fueled by, for example, the relative value between the options (Klein, Ullsperger, and Jocham 2017), which is more in line with the decoder in previous neuroreinforcement work (Cortese et al. 2021). The latter alternative would perhaps best be expressed by a decoder that compares the Target vs a Control and does not necessarily require values about the target category to be computed at all, instead behavior could be driven by a behavioral policy (Hayden and Niv 2021; O'Doherty et al. 2004). The decision-maker may apply a heuristic approach at the moment of choice in which they first identify a dimension along which the two options vary and then compare the options along that dimension (Niv 2019). Value may have helped establish what that dimension should be, and done so differently between participants.

Critically, value as described here seems to have been responsible for the observed variability in choice – it is possible that the variability in choices was driven by neuroreinforcement effects, even if that variability did not consistently point towards the intended direction. It is therefore important to note that the learning assessed in the present experiment is not unambiguously, purely model-free - instead, in a dual-system model this type of learning could be multiply determined. Indeed, a *pure* behavioral test of model-free learning does not seem to exist, nor might a pure value correlate in the brain (Hayden and Niv 2021). While this may be true, it is reasonable to assume that if the Target accrued value (model-free or otherwise), that learning should have been expressed given the right measurement.

While the use of animate and inanimate object categories as stimuli were chosen for the purpose of eliciting a strong fMRI-based signal for decoding, one might argue that the use of categories as stimuli might have been suboptimal as pre-existing preferences or conceptual links between the categories may influence the choice task. To limit this possibility, the categories used in the experiment were only the ones rated "Neutral" on the pre-scan preference. The stimuli were also always within the same animate/inanimate distinction. However, pre-existing preferences and conceptual links may still have biased choice behavior. That is, participants may have based their choices in the forced-choice task on a momentary relative preference of one category over the other or based on a conceptual link between them. This is unlikely as the variability in choice behavior only emerged in the two critical conditions with the Target category present, and consistently not, in the Controls category. If pre-existing preferences or conceptual links influenced choice behavior then there should have also been the same level of variability in the Controls condition as in the critical conditions, which was not the case. Instead, it is likely that the neuroreinforcement procedure itself was responsible for the variability in choosing the Target. Future studies should incorporate abstract stimuli such as fractals or symbols to directly test this as a tighter control for this potential confound.

Here I attempted to establish the efficacy of neuroreinforcement for unconscious learning. Results showed that while participants did not choose the Target category more often than the Control categories, there was more variability in the conditions with the Target category present. This effect was not a result of imbalance conditions. Neither was it an effect of pre-existing preferences or conceptual links. Instead, the variability seems to be a result of the neuroreinforcement procedure itself. Why this pattern emerged should be investigated in future studies using physiological, behavioral and subjective responses. Furthermore, more research is needed to predict neuroreinforcement success (Haugg et al. 2020), and to understand the mechanism of neuroreinforcement (Shibata et al. 2019; Chiba et al. 2019).

Chapter 3

The associative structure of higher-order conditioning in humans

General introduction

Say I make you like A, then I pair A with B, then I make you dislike A, what will you think of B? How about this then: Say I first teach you to pair X with Y, then make you like X, right after that I make you dislike X, what will you think of Y? Would your preference for B and Y be any different?

Intuitively, the answer might be no; the only difference between the two scenarios is the order in which you learned to prefer A and X; in the first scenario A is liked before the A-B pairing, and in the second scenario is X is liked after the X-Y pairing. However, it turns out that the order in which you learn *does* matter and that the two scenarios likely involve different types of inferences. The ability to make the inferences required for the scenarios above involves integrating experiences that occur at different times. Inferences are required when there is limited experience. In the scenario above, you have not directly learned anything about B and Y, you have to infer their status.

Animals and people are able to solve novel problems in the absence of experience using this type of complex reasoning system. Understanding this ability is crucial for understanding reasoning and memory formation. Learning to predict outcomes increases the chances of encountering or avoiding them, depending on current needs. Hence it is foundational for adaptive behavior. For example, while driving you might see an arch by the side of the road. This may make you think of burgers which in turn predicts calories. Similarly, being bitten by a dog may develop a fear of dogs; which in turn might cause one to avoid places where dogs frequent.

This learning process is referred to as higher-order conditioning and involves integrating information that might allow one to obtain desirable outcomes and avoid undesirable ones. Higher-order conditioning accounts for many of our learned experiences. Therefore, studying higher-order conditioning is important for understanding how memories are structured. Most importantly, it has clinical implications. For example, treatment against dog phobia may require reducing fear of the dog *and* the place where the dog bite was encountered.

Higher-order conditioning is studied in the lab through second-order conditioning and sensory preconditioning procedures (Figure 2.1). Second-order conditioning refers to instances in which a neutral stimulus (S1) is paired with an outcome (e.g., food or money). After this, a second stimulus (S2) is paired with S1, but without the outcome. Second-order conditioning occurs when S2 elicits a response at Test despite the absence of direct pairing with the outcome. The pairing procedure in sensory preconditioning is similar to second-order conditioning, except that the order of phase 1 and 2 is reversed.

	Phase 1	**Phase 2**	**Test**
Second-order Conditioning	S1 → $	S2-S1	S2?
Sensory Preconditioning	S2-S1	S1 → $	S2?

Figure 2.1 **Higher-order conditioning procedures.** $ indicates money. - indicates associations. Arrows indicate that S1 is followed by money.

Researchers have used extinction or devaluation of the outcome as a tool to study the associative structure of higher-order conditioning (Rizley and Rescorla 1972). Extinction involves repeatedly presenting S1 without the outcome before Test which causes the response to S1 to decrease (i.e., after phase 2 but before Test). Devaluation makes the outcome less desirable by for example pairing food with illness or making the outcome less valuable (e.g., less money) (Allman et al. 2010). If S1 extinction or outcome-devaluation also reduces S2 responding that is taken to mean that the associative structure is dependent on S1. This associative structure is referred to as a chain structure (S2–>S1–>outcome). Alternatively, if S2 responding is relatively intact, the associative structure is referred to as a direct-link (S2–>outcome) (Figure 2.2).

Rescorla has proposed that the manner in which stimuli are paired determine the type of association that is formed. Rescorla found evidence supporting the chained-link account in second-order conditioning when the S2-S1 pairings are presented simultaneously, while the direct-link account was supported when the S2-S1 pairings are presented sequentially (Rescorla 1982; Rescorla 1980).

Rescorla reasoned that ‘between-event’ associations serve as predictions about future events which are formed by sequential presentations. The presence of S2 at test elicits an expectation that S1 will occur and thereby its appropriate response is generated. In contrast, ‘within-event’ associations can be retrieved following exposure to partial event information. Memories formed through simultaneous protocols are more likely to reflect mediated conditioning (discussed below). According to Rescorla, S1 extinction following sequential second-order conditioning should keep responding to S2 intact, while S1 extinction following simultaneous second-order conditioning should decrease responding to S2 (Rescorla 1981).

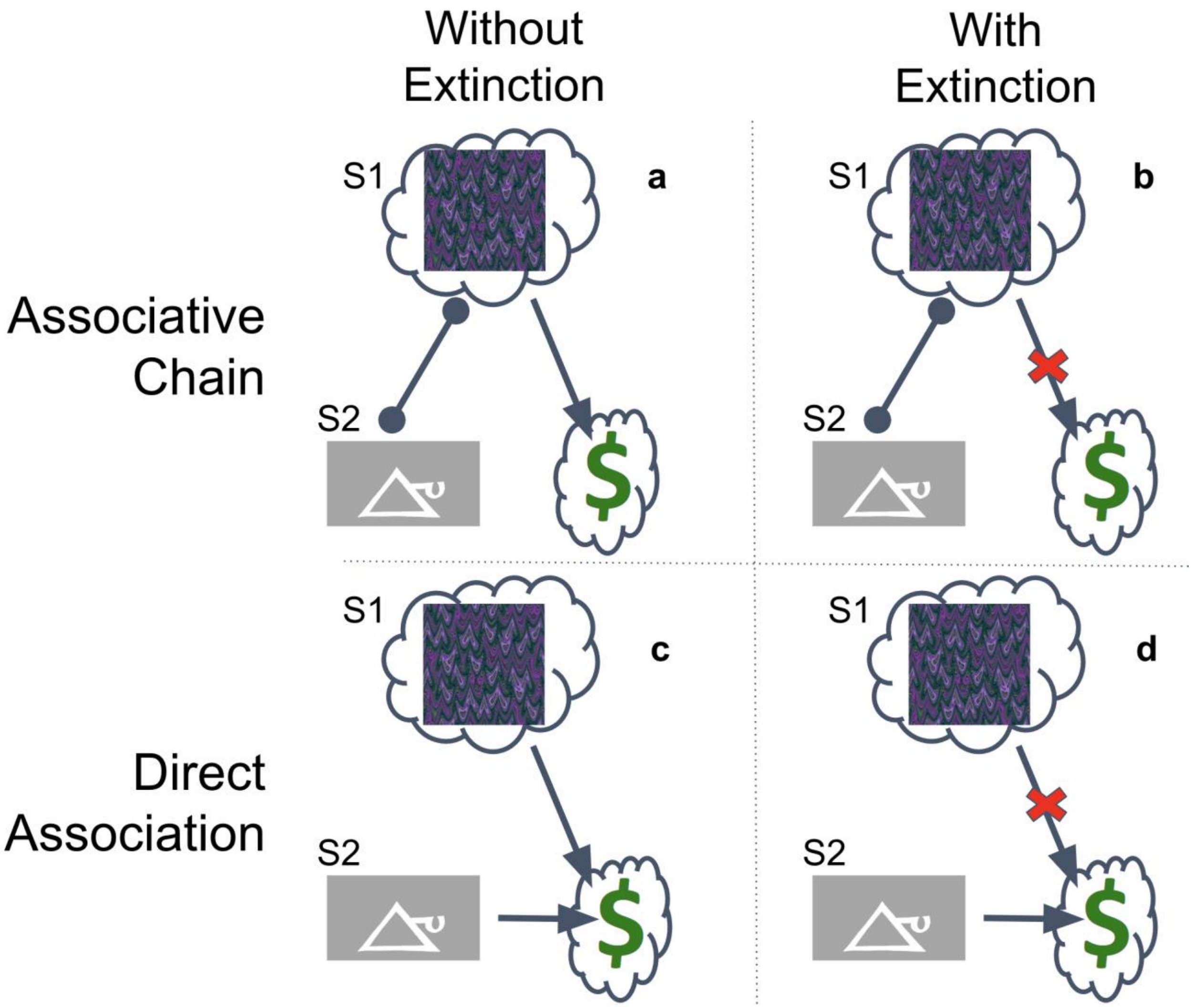

Figure 2.2 **Hypothetical associative structures underlying inference about the second-order cue (i.e., S2) in second-order conditioning.** a) According to the chain-link account, S2 is linked to S1 which in turn predicts money. b) According to this view, extinguishing S1 also affects responding to S2. c) According to the direct-link account, S2 becomes directly associated with money. d) Therefore, extinguishing S1 does not affect responding to S2. Arrows = predictions, Links = associations, Cloud = evoked representation, Cross = extinction.

Surprisingly, most studies have used sequential presentations for higher-order conditioning. Simultaneous presentations have largely been ignored. These studies, mostly conducted with rats, have found that the most likely structure for sequential second-order conditioning is one where S2 becomes directly associated with the US during phase 2 (i.e., direct-link) (Seitz, Blaisdell, and Sharpe 2021; Lee 2021). Alternatively S2 could evoke a more complex internal model at Test between all three elements that does not involve a direct-link between S2 and the US (i.e., chaining) (Holmes et al. 2021; Prével and Krebs 2021). There is now rather convincing evidence in rodents that sequential sensory preconditioning is supported by a chaining mechanism, and that sequential second-order conditioning is supported by a direct-link (Sharpe et al. 2017; Sharpe, Batchelor, and Schoenbaum 2017; Seitz, Blaisdell, and Sharpe 2021; Holmes et al. 2021; Barron et al. 2020).

Notably, it has been proposed that the associative structure of sensory preconditioning could be supported through mediated conditioning during phase 2 in which S1 evokes a representation of S2 which allows S2 to enter an association with the outcome (Doll and Daw 2016). That is, the evoked representation of S2 during phase 2 is enough for S2 to become directly associated with the reward. A similar mediated conditioning mechanism has been proposed for second-order conditioning during phase 2 since here S1 already predicts the reward (which is evoked when S1 is presented), thereby allowing S2 to enter an association with the reward (Holland 1981; Holland and Forbes 1982).This association is not mutually exclusive from a chain-link association and both may exist in parallel.

Evidence for the associative structure of higher-order conditioning in humans is rather scarce and has shown mixed results (Lee 2021; Holmes et al. 2021). Here, I aim to address this gap in three experiments using simultaneous higher-order conditioning.

The first experiment used an extinction procedure to selectively extinguish one first-order cue (S1Ext) and leave another first-order cue intact (S1+). Extinction occurred after phase 2 but before Test. Following Rescorla's hypothesis, it was expected that simultaneous presentations should produce results more conducive to chain-link (or mediated conditioning). Each first-order cue was paired with an associate second-order cue (i.e., S2Ext and S2+). It is expected that S2+ should be chosen more often than S2Ext when each is separately paired within a choice against Neutral (2).

Importantly, since sensory preconditioning is thought to be supported by a chain mechanism (or mediated conditioning), it is expected that responding to the preconditioned cues will mirror responding to the second-order cues. The second experiment used a devaluation procedure to test the same question but within the context of an influential behavioral neuroscience theory about two different modes of behavior and the actions they guide (Dickinson and Balleine 1994). The last experiment asked whether preconditioned cues and second-order cues access model-free value using conditioned reinforcement tests in order to answer the questions asked in Exp. 1. and 2 within a computational framework (Daw, Niv, and Dayan 2005).

To anticipate, in contrast to Rescorla’s hypothesis but in agreement with most sequential higher-order conditioning studies in non-human animals, results support the accounts that second-order conditioning is supported by a direct-link mechanism, and that sensory preconditioning is supported by a chained mechanism; even when simultaneous presentations are used.

Experiment 1

While most evidence in humans point to the direct-link account being the most likely structure for second-order conditioning (Luettgau et al. 2021; Davey and McKenna 1983; Davey and Arulampalam 1982), and the chained-link account the most likely structure for sensory preconditioning, almost all have used sequential presentation (Wang, Schoenbaum, and Kahnt 2020). With the exception of one study which found correlational evidence for mediated conditioning in simultaneous sensory preconditioning in humans (Wimmer and Shohamy 2012), no other study has tested the associative structure of simultaneously presented higher-order conditioning. To address this gap, this experiment used extinction in simultaneous second-order conditioning and simultaneous sensory preconditioning procedures.

Participants underwent extinction of the first-order cues before Test. The extinction phase involved a two-alternative forced-choice where participants learned through trial and error that a previously rewarded first-order cue was no longer rewarded (and hence extinguished). The forced-choice task was used because second-order conditioning in humans can often be relatively weak when predictions have been used as an outcome measure (Lee 2021). This is because it is not always rational to expect that the second-order cue will lead to reward. In fact, the second-order cue should reasonably predict the *absence* of reward since no reward is actually received during the second-order conditioning phase. Indeed, increasing second-order conditioning trials or deliberation time can generate conditioned inhibition. If second-order conditioning is weak, then it is expected that a prediction will also generate a weak effect. Previous research found that a forced-choice task is more likely to overcome these shortcomings by making the expression of an excitatory relationship with the outcome more likely (Mitchell, Livesey, and Lovibond 2007). Crucially, participants are not required to also make predictions during second-order conditioning about the outcome, as this could also lead to the development of conditioned inhibition.

Sensory preconditioning was used as a comparison to second-order conditioning because no account predicts that the preconditioned cue should not be affected by extinction of the first-order cue. The chained mechanism account argues that the preconditioned cue’s status depends on its first-order associate. Even a mediated conditioning account for sensory preconditioning predicts reduced responding to the preconditioned cue, but this may occur in addition to, or absence of a chained mechanism. Hence it is hypothesized that extinguishing the first-order cue should also affect responding to the preconditioned cue as it is “indirectly” extinguished.

Methods

Participants: 334 participants were recruited through the Prolific online platform. Of these, 169 were recruited for the second-order conditioning study, and 165 were recruited for the sensory preconditioning study. Power analysis from piloting determined that 95 participants were needed in each study for power of 0.80. After applying the exclusion criteria (Supplementary Table 2.1) the final sample size of the second-order conditioning experiment was 99 (***Mean age*****:** 40.03; ***Range*****:** 19-70; ***Gender*****:** 46 Female, 2 Prefer not to say; ***Ethnicity***: 6 Asian, 4 Black, 82 White, 4 Other, 3 Mixed, 3 Prefer not to say). The final sample size for sensory preconditioning was also 99 after applying exclusion criteria (***Mean age*****:** 31.9; ***Range*****:** 19-65; ***Gender*****:** 42 Female; ***Ethnicity:*** 8 Asian, 6 Black, 71 White, 4 Other, 7 Mixed, 3 Prefer not to say). The experiment was approved by University of California, Los Angeles Institutional Review Board and all participants provided informed consent.

Materials: At the beginning of the experiment a randomization procedure determined the stimulus to be used for each participant. The stimuli were fractals and abstract symbols (agatho font). Randomization designated whether abstract symbols or fractals were first-order or second-order (or preconditioned) stimulus. For each participant, it was randomly assigned whether the first-order or second-order (or preconditioned) stimulus were abstract symbols or fractals. For example, if the first-order stimuli were fractals, that meant the second-order stimuli were abstract symbols. Each of the abstract symbols and fractals were also randomly assigned to each of the stimulus types (i.e., first-order cue, first-order extinction, Neutral (1); second-order cue, second-order extinction etc.). All cues were presented on a light gray background and presented on the left and right side of the screen. The outcomes were the word “*WIN”,* and “*No* *outcome”* which were presented on the center of the screen after each choice. The design of the experiment is summarized in Supplementary Table 2.2 and 2.3. The experiment was coded in JavaScript using custom jsPsych plugins (de Leeuw 2015).

Procedure: The procedure is presented in Figure 2.3. The experiment included the following phases: A) *Conditioning phase,* B) *Second-order conditioning phase*, and *Preconditioning phase* (referred to as “Association phase” when both experiments are discussed), C) *Extinction phase,* D) *Test phase.* The conditioning phase was before the second-order conditioning phase in the second-order conditioning experiment and after the preconditioning phase in the sensory preconditioning experiment. The extinction and Test phase order was the same across experiments. Participants were told they were going to play a choice task in which they could earn extra money in bonus.

Conditioning phase: Participants were told to learn which stimulus led to the outcome and which did not. They were told that they needed to use the left and right key to choose between the stimulus presented on the left and right side of the screen to win five cents on each trial. The two stimuli were separated by a gap in the middle. Each trial lasted for four seconds and participants had three seconds to make their choice. There was no ITI. After their choice, a blue square surrounded the chosen stimulus to indicate the selected option followed by feedback on whether the choice was a win or not. Participants therefore learned by trial and error which choice was winning and which was not. If participants missed a trial, the text “Too late!” appeared at the center of the screen and 15 cents was deducted from their total bonus. This was done to ensure compliance with the task instructions. The total accumulated reward was constantly presented on the top right of the screen. All the stimulus-types were paired against each other. The filler was added to equate the number of reinforced cues and nonreinforced

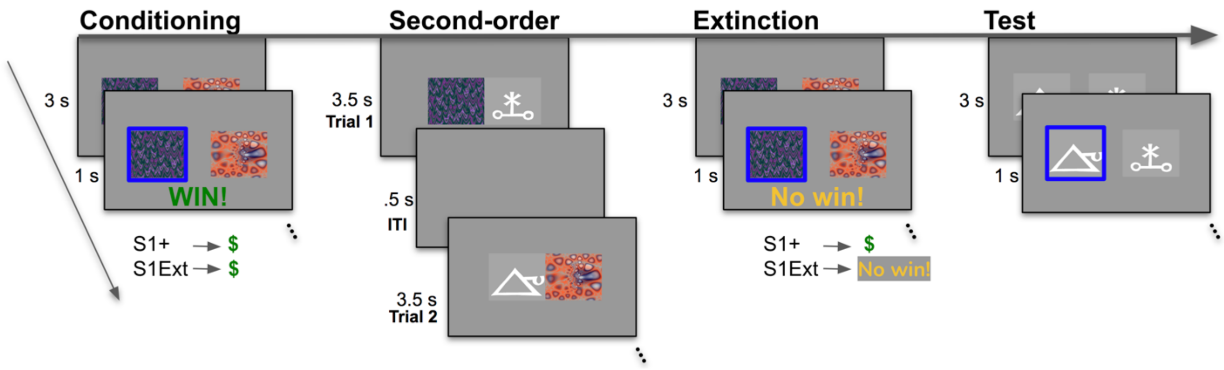

Figure 2.3 **Second-order conditioning extinction task.** During conditioning, participants learned associations between each first-order cue (S1+, S1Ext), and a monetary outcome ($). On each trial, participants had to choose between the presented cues by pressing the left or right arrow key corresponding to the cues presented on the screen. Then, feedback was presented with a blue square surrounding the chosen option. In second-order conditioning, participants learned associations between cue pairs. During extinction, one of the first-order cues was no longer associated with a monetary outcome. In sensory preconditioning, the order of conditioning and second-order conditioning was flipped such that conditioning happened after the association phase (i.e., equivalent of the second-order phase). During Test, participants were asked to choose the stimuli they thought led to the monetary outcome, but no feedback was provided. ITI, inter-trial interval. S1+, rewarded first-order cue. S1Ext, extinguished first-order cue.

cues as well as to encourage discrimination learning as opposed to simply responding on all trials. Which stimulus was presented on which side of the screen was randomized on each trial. There were six different conditions presented in a randomized and counterbalanced order for 12 trials for each condition. Hence, there were 72 trials total.

Association phase: Prior to the association phase, participants were instructed that they would see abstract symbols and fractals getting paired together, and that they should try to learn which stimuli make up a pair. The stimuli pairs were presented side by side with no gap in between. The trial order, and stimulus location on the screen was randomized. To ensure attention and compliance with the task, participants were told to complete a task while learning the pairs. Namely, participants were told to press the letter ‘S’ on the keyboard if the current pair was the same as the pair on the previous trial and to press the letter ‘D’ on the keyboard if the current pair was different from the previous trial. Each trial lasted for 3.5 seconds and there was a 500 millisecond ITI. There were 10 trials per condition and 30 trials total.

Extinction phase: Before the extinction phase, participants were instructed that they would once again be choosing between stimuli but that this time the outcome might change from what it was in the conditioning phase. This time, if participants chose the extinguished first-order cue (i.e., S1Ext), they would not win any money. Participants therefore had to learn through trial and error whether which stimuli was extinguished. The order of the trials and the left/right location on the screen was randomized across trials. The neutral filler was presented again in the extinction phase, and a new stimulus, rewarded filler was introduced. Choosing the rewarded filler led to a 5 cent bonus. Fillers were added to the task to encourage discrimination learning and to discourage the learning of nonspecific rules.

Test phase: Finally, participants were instructed that they would now have to play the choice task but that this time there would be no feedback. They were instructed to try their best to pick the winning stimuli. As before, the stimuli were presented side by side on the screen against a gray background and participants made their choice using the left and right arrow keys on their keyboard. Each trial lasted for four seconds and participants had three seconds to respond. A blue square surrounded the chosen option after each choice. There was no counter on the top right of the screen.

Planned analysis: The data was extracted using custom MATLAB scripts, and JASP and R were used for data-analysis. Choice behavior was analyzed by computing the average number of times each participant picked one stimulus over the other in each condition. Support for the direct-link account would take the form of choosing S2+ in the S2+ vs. Neutral condition equally as often as choosing S2Ext in the S2Ext vs. Neutral, despite the extinction of S1Ext and S1+ left intact. Support for the chained-link account would take the form of choosing S2+ in the S2+ vs. Neutral condition more often than choosing S2Ext in the S2Ext vs. Neutral condition. This is because if S2Ext's status is dependent on S1Ext, then extinguishing S1Ext should affect S2Ext.

This was tested formally with a paired samples t-test between the conditions S2+ vs. Neutral and S2Ext vs. Neutral. Sensory preconditioning was used as a control to second-order conditioning. This is because the direct-link account cannot explain sensory preconditioning. Therefore, in sensory preconditioning S2Ext vs. Neutral (i.e., the indirectly extinguished preconditioned cue) is picked less than in the S2+ vs. Neutral condition (i.e., the indirect non-extinguished preconditioned cue). This was also tested formally with a paired-samples t-test. Evidence for first-order extinction, second-order and preconditioning was tested using one-samples t-tests against 50% as the criterion.

Results and Discussion

The second-order conditioning and sensory preconditioning results are presented in Figure 2.4.

Second order conditioning: Extinction of the first-order cue was robust and significantly above 50% (t(98) = 36.809, p < .001, d = 3.699). Participants picked S1+ on average 92.1% of the time (SD = 11.375). Although second-order conditioning is typically a relatively weak effect, in this sample second-order conditioning was also robust and significantly above 50% (*S2+ vs. Neutral*: t(98) = 5.501, p < .001, d = .553, M = 69.6, SD = 35.455). S2Ext was chosen more often than the neutral control (*S2Ext vs. Neutral*: t(98) = 3.812, p < .001, d = .383, M = 64.6, SD = 38.2), but less often than S2+ (i.e., *S2+ vs. S2Ext*: t(98) = 4.664, p < 0.001, d = 0.469, M = 66.5, SD = 36.2). Critically, there was no significant difference between the S2+ vs. Neutral condition and the S2Ext vs. Neutral condition (t(98) = 1.343, p > 0.180), even though the experiment was optimized to find such a difference. Extinguishing S1Ext did not affect S2Ext, thereby supporting the direct-link account for the associative structure of second-order conditioning, even when simultaneous presentations are used.

Sensory preconditioning: The critical test for the second-order conditioning results described above is whether there is a difference between the S2Ext vs. Neutral and the S2+ vs. Neutral conditions in sensory preconditioning. The chained-link account (and the mediated learning account) predicts that extinguishing S1Ext should affect S2Ext. Therefore, it was hypothesized that the response to S2Ext would be less than responding to S2+ in the conditions where they each are paired against the Neutral cue. A paired sample t-test confirmed this prediction, there was a significant difference between the S2+ vs. Neutral condition and the S2Ext vs. Neutral condition (t(98) = 2.715, p = .008, d = .273). There was also robust sensory preconditioning (*S2+ vs. Neutral*: t(98) = 5.392, p < .001, d = .542, M = 69.7, SD = 36.3), and the non-extinguished preconditioned cue was chosen more often compared to S2Ext (*S2+ vs. S2Ext*: t(98) = 5.652, p > .001, d = .542; M = 69.7, SD = 34.7). Furthermore, S2Ext was chosen more than Neutral (*S2Ext vs. Neutral:* t(98) = 2.442, p = 0.02, d = .245, M = 59.2, SD = 37.59).

Experiment 1 yielded results that differed from what was expected based on Rescorla's predictions on simultaneous higher-order conditioning (Rescorla 1981). Rescorla reasoned that simultaneous presentations may evoke a configural representation of S1-S2 which could explain why simultaneous but not sequential second-order conditioning results in chain-link representations. That is, in simultaneous but not sequential second-order conditioning, participants form a configural representation of a S1-S2 complex as opposed to integrating and associating two stimuli as two elements. To my knowledge, the present experiment is the first to investigate the associative structure of higher-order conditioning with simultaneous presentations in humans using extinction. Instead, results agree with the existing literature concerning sequential second-order conditioning in that S2Ext responding was equal to S2+ responding when compared to Neutral. These results support the direct-link account for

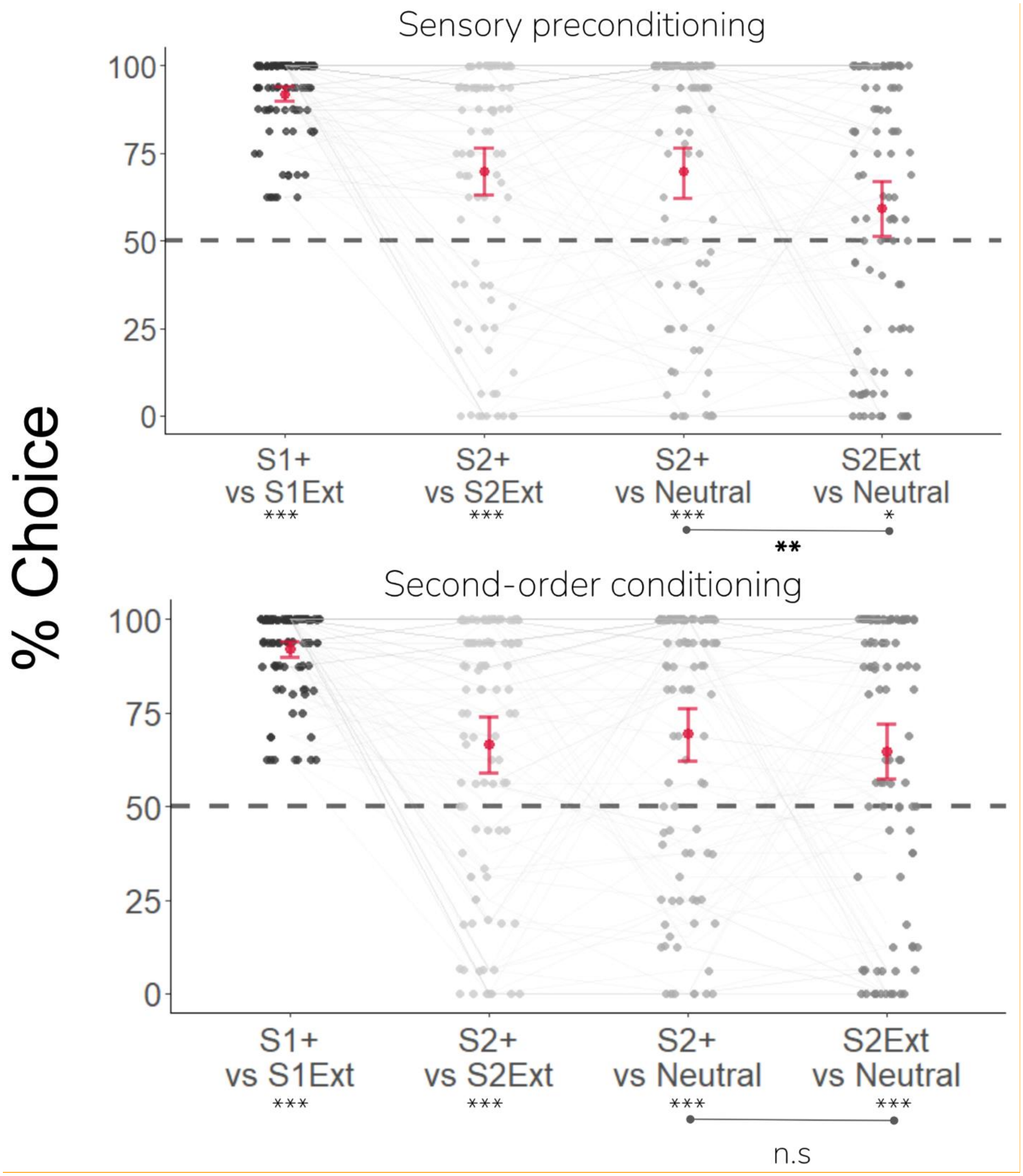

Figure 2.4 **Second-order cues are not affected by extinction of first-order cues, but preconditioned cues are.** S1+, rewarded first-order cue. S1Ext, extinguished first-order cue, S2+, second-order cue associated with S1+, S2Ext, second-order cue associated with S1Ext. Neutral, neutral cue. n.s = not significant. * p < .05, ** p < .01, *** p < .001

second-order conditioning. Importantly, the sensory preconditioning study did find that S2Ext responding was smaller than S2+ responding when compared to Neutral. Sensory preconditioning cannot be explained by the direct-link account and was therefore used as a control.

These results seem inconsistent with Craddock et al. (2018) who found evidence for the chained-link account in sequential second-order conditioning. However, there are a number of differences between studies. Of note is that Craddock et al. (2018) used predictions as their outcome measure along with an unusual dependent variable. As discussed earlier, prediction may not always be optimal for second-order conditioning because they assume that participants predict a reward when the reward is actually absent. Predictions are therefore highly sensitive to the number of trials during the second-order phase. More trials can lead to conditioned inhibition in which participants will expect the *absence* of reward. Instead, forced-choice prediction tasks find evidence that asking participants to make predictions about the outcome may lead to the expression of an inhibitory causal relationship between S2 and the outcome. On the other hand, participants were still able to identify which outcome went with S2 (Mitchell, Livesey, and Lovibond 2007).

What seems to matter is how learning is expressed. It may even be so that inhibitory and excitatory relationships can exist in parallel (Muñiz-Diez, Muñiz-Moreno, and Loy 2021; Holland and Rescorla 1975). Here, the expression of inhibitory associations was likely avoided by using a forced-choice task and by encouraging participants to encode the association between S1 and S2. However, it is not clear from this study whether this is a performance effect or a learning effect. That is, an inhibitory association may have been developed in parallel that is not expressed in the forced choice task. This intriguing possibility could be assessed using a retardation test in future studies.

Many years of research have shown that extinction weakens responding but does not erase original learning. Extinction involves a form of context-specific inhibition, rather than unlearning. If the context changes after extinction, the original learning can return. This effect is referred to as the renewal effect. Based on this, it is possible that if the preconditioning participants' response to S1Ext was extinguished in one context and tested in the original conditioning context, then they might have nevertheless responded to S2Ext at the same level as S2+ due to the renewal effect. These results would look similar to the present second-order conditioning results, possibly leading one to incorrectly conclude that sensory preconditioning is supported by a mechanism akin to a direct-link. What this example highlights is that extinction involves the creation of a new, competing memory. Hence, it is warranted to conduct another test of the associative structure of higher-order conditioning using an additional manipulation, namely devaluation (see also Bouton 2021). This procedure helps determine whether human second-order conditioning is affected by motivational value. That is, devaluation tests whether the representation of the outcome is linked to S2 (i.e., its associative motivational state or specific sensory features; Holland 1977).

Experiment 2

As discussed in the introduction, the behavioral neuroscience theory of goal-directed and habitual behavior (Dickinson and Balleine 1994) is formalized to a useful extent within the model-free/model-based reinforcement learning framework (Daw, Niv, and Dayan 2005; Dolan and Dayan 2013). Although not all the exact details overlap, this distinction can be extended to that of chained associations and direct-links (Collins and Cockburn 2020). Chained representations can be thought of as solving higher-order conditioning using an associative map of the task at hand (Seitz, Blaisdell, and Sharpe 2021). These maps represent a model of the world that is acquired flexibly which allows the organism to anticipate likely outcomes. On the other hand, direct-links can be acquired through a habitual system that arises through a process akin to *model-free* learning in which actions are evaluated through trial and error. Model-free behaviors are considered habitual because they persist even if the desirability of the outcome has changed (i.e., devaluation of the outcome) (Watson et al. 2022). In contrast, *model-based* learning produces flexible behavior that adapts to devaluation by using models of the world to evaluate possible actions and anticipate future outcomes (Doll, Simon, and Daw 2012). Model-free learning has been formalized with the reward-prediction-error hypothesis. This hypothesis states that the change in the value of a stimulus is driven by a discrepancy between the expected outcome and the obtained outcome. Temporal difference learning captures these dynamics (Sutton and Barto 2018).

Temporal difference allows value to spread from S1 to S2 during second-order conditioning but fails to account for sensory preconditioning. This is because during sensory preconditioning there is no outcome, hence there is no value to be spread. In contrast to temporal-difference models, model-based reinforcement models can account for sensory preconditioning (Daw et al. 2006; Mattar and Daw 2018; Barron et al. 2020). These models include the states and the transition probabilities between states and the available reward. During the first phase of sensory preconditioning participants learn the relationship between S1 and S2, before learning about the relationship between S1 and the outcome during the conditioning phase. Then, the updated value of S2 can be computed through a simulation mechanism based on the learned transition between S1 and S2, and the expected value of S1.

One way to test these predictions within the goal-directed vs. habitual actions theory framework is through using devaluation (Balleine and Dickinson 1998). For instance, a study using sequential sensory preconditioning in rats showed that responding to the preconditioned cue is sensitive to devaluation of the reward at the time of test (Hart et al. 2020). The direct-link account assumes that the second-order cue is directly associated with the outcome during conditioning in a manner that is independent of S1 and of subsequent manipulation of the value of the outcome.

To extend these findings to humans, Experiment 2 tested the associative structure of higher-order conditioning using devaluation of a monetary reward. It is hypothesized based on the results from Experiment 1, that the second-order cue should not be sensitive to devaluation but that the preconditioned cue will be.

Methods

Participants: 143 participants in total were recruited through the Prolific online platform. Of these, 84 were recruited for the second-order conditioning study, and 66 were recruited for the sensory preconditioning study. Power analysis from piloting determined that 53 participants were needed in each study for power of 0.80. After applying the exclusion criteria (Supplementary Table 2.4), the final sample size of the second-order conditioning experiment was 55 (***Mean age*****:** 44.53; ***Range*****:** 19-76; ***Gender*****:** 19 Female, 1 Prefer not to say; ***Ethnicity***: 3 Asian, 4 Black, 41 White, 2 Other, 2 Mixed, 3 Prefer not to say). The final sample size for sensory preconditioning was 57 after applying exclusion criteria (***Mean age*****:** 36.2; ***Range*****:** 19-70; ***Gender*****:** 33 Female; ***Ethnicity:*** 5 Asian, 1 Black, 48 White, 1 Other, 1 Mixed, 1 Prefer not to say). Two multiple-choice tasks, the first delivered after the conditioning phase, and the second after the devaluation phase were used to exclude participants that did not learn from instructions which outcome was devalued. Participants were excluded if they got one of the questions wrong. The experiment was approved by University of California, Los Angeles Institutional Review Board and all participants provided informed consent.

Materials: The stimuli were the same as in Exp. 1, with the exception that there was no extinguished first-order cue since Exp. 2 used devaluation. The design is summarized in Supplementary Tables 2.5 and 2.6.

Procedure: The phases of Exp. 2 were exactly the same as Exp. 1 except that the Extinction phase was replaced with the Devaluation phase (Figure 2.5). There was also a difference between Exp. 1 and Exp. 2 in the Conditioning phase, as described below.

Conditioning phase: Like in Exp. 1, participants completed the choice task in which they learned which stimulus led to a favorable outcome. This time however, the outcome of a winning choice was followed by a blue or orange triangle indicating that the participant had won 50 points on that trial. Participants were told to learn to choose the stimulus associated with points in order to collect as many points as they can to maximize their bonus. S1+ and S1Deval were rewarded, while the other stimuli were not. For example, choosing the S1+ stimulus leads to a presentation of the blue triangle (or orange triangle; counterbalanced between subjects). Except for this difference, all other parameters were exactly the same as for Exp. 1.

Devaluation phase: One of the triangles was selectively devalued through instructions before the Test phase (i.e., the triangle associated with S1Deval). Participants were instructed that instead of both triangles being worth 50 cents, now only one is worth 50 cents, and the other one worth 5 cents. This was also confirmed in a subsequent multiple-choice task in which they had to indicate the value of both triangles. Participants were excluded from the experiment if they answered incorrectly. Therefore, all the participants in the analysis passed the devaluation manipulation check.

Analysis: Choice behavior was investigated with the same analysis as in Exp. 1. To confirm the main hypothesis of the experiment, exploratory analysis was conducted using the same approach as the planned analysis on participants that had chosen S1+ over S1Deval more than 60% of the time in both experiments. An independent samples t-test was used to explore the difference between the S2Deval vs. Neutral condition and the S2Deval vs. Neutral condition across second-order conditioning and sensory preconditioning in the full sample. It was hypothesized based on Exp. 1, that S2Deval responding would be higher in second-order conditioning than sensory preconditioning.

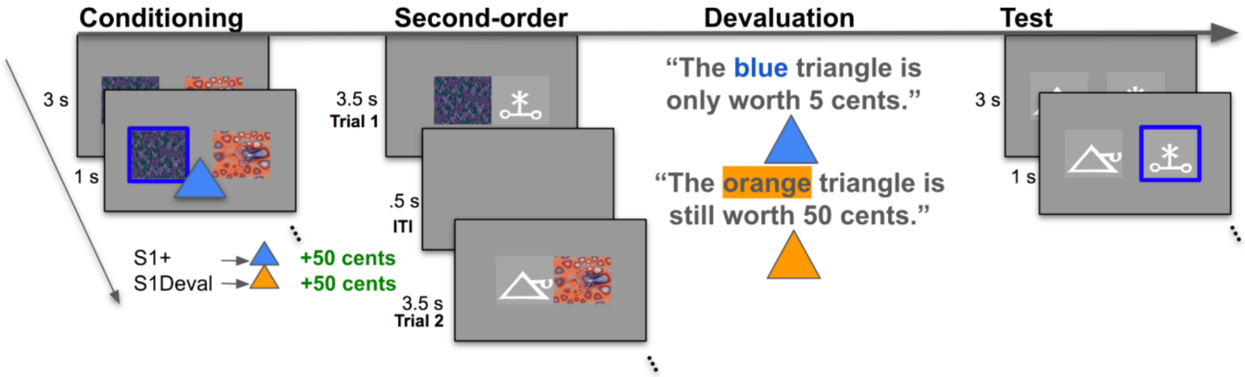

Figure 2.5 **Second-order conditioning devaluation task.** During conditioning, participants learned associations between each first-order cue (S1+, S1Deval), and a triangle, which was a monetary rewarded outcome ($). On each trial, participants had to choose between the presented cues by pressing the left or right key corresponding to the location of the cues presented on the screen. Then, the triangle associated with the chosen option (color counterbalanced) was presented on the screen. A blue square also surrounded the chosen option. In the second-order conditioning phase, participants learned associations between cue pairs. In sensory preconditioning, the order of conditioning and second-order conditioning was flipped such that conditioning happened after the preconditioning phase (i.e., equivalent of the second-order phase). During devaluation, participants were instructed that one of the triangles (counterbalanced) was devalued. During Test, participants were asked to choose the stimuli they thought would maximize their monetary outcome. No feedback was provided during Test. ITI, inter-trial interval. S1+, rewarded first-order cue. S1Deval, first-order cue associated with devalued outcome.

Results and Discussion

Second order conditioning: Devaluation of the first-order cue was robust and significant (Figure 2.6), participants picked S1+ cue 67.7% of the time compared to S1Deval (SD = 36.3) (t(54) = 3.621, p < .001, d = .488). Exp. 2 replicated the relatively strong second-order conditioning effect of Exp. 1, showing that S2+ was chosen more often than Neutral (t(54) = 4.677, p < .001, d= .631, M = 72.8, SD = 36.2). Furthermore, S2Deval was chosen more than Neutral (t(54) = 4.388, p < .001, d = .592, M = 70.7, SD = 35.0). Strikingly, there was no significant difference in the S2+ vs. S2Deval condition, with S2+ being chosen only 57% of the time compared to S2Deval (t(54) = 1.453, p = .152, d = .196, SD = 37.6). The main hypothesis was confirmed in that there was no difference between the S2+ vs. Neutral and the S2Deval vs. Neutral condition (t(54) = .498, p = .620, d = .067), even though the sample size was optimized to find such a difference. These results suggest that selectively devaluing the outcome did not affect its indirectly associated second-order cue, thereby replicating the results of Exp. 1 by finding support for the direct-link account for the associative structure of second-order conditioning, even when simultaneous presentations are used.

Sensory preconditioning: As in Exp. 1, it was hypothesized that devaluation of the outcome should affect responding to S2Deval. Results suggest that this prediction was confirmed. There was a significant difference between the S2+ vs. Neutral condition and the S2Deval vs. Neutral condition (t(56) = 2.451, p = .017, d = 0.325). S2+ was chosen 67.3% of the time compared to Neutral, showing robust sensory preconditioning (t(56) = 3.771, p < .001, d = .419, SD = 34.8). Furthermore, the S2Deval vs. Neutral condition was not significant (t(56) = .924, p = .360, d = .122, M = 54.4, SD = 35.9). The effect of selective outcome devaluation was further confirmed in the first-order cues, S1+ was chosen more than S1Deval (t(56) = 5.656, p < .001, d = .749, M = 75.4, SD = 34.02). There was also a significant difference in the S2+ vs. S2Deval condition (t(56) = 3.165, p = .003, d = .419, M = 64.3, SD = 34.1).

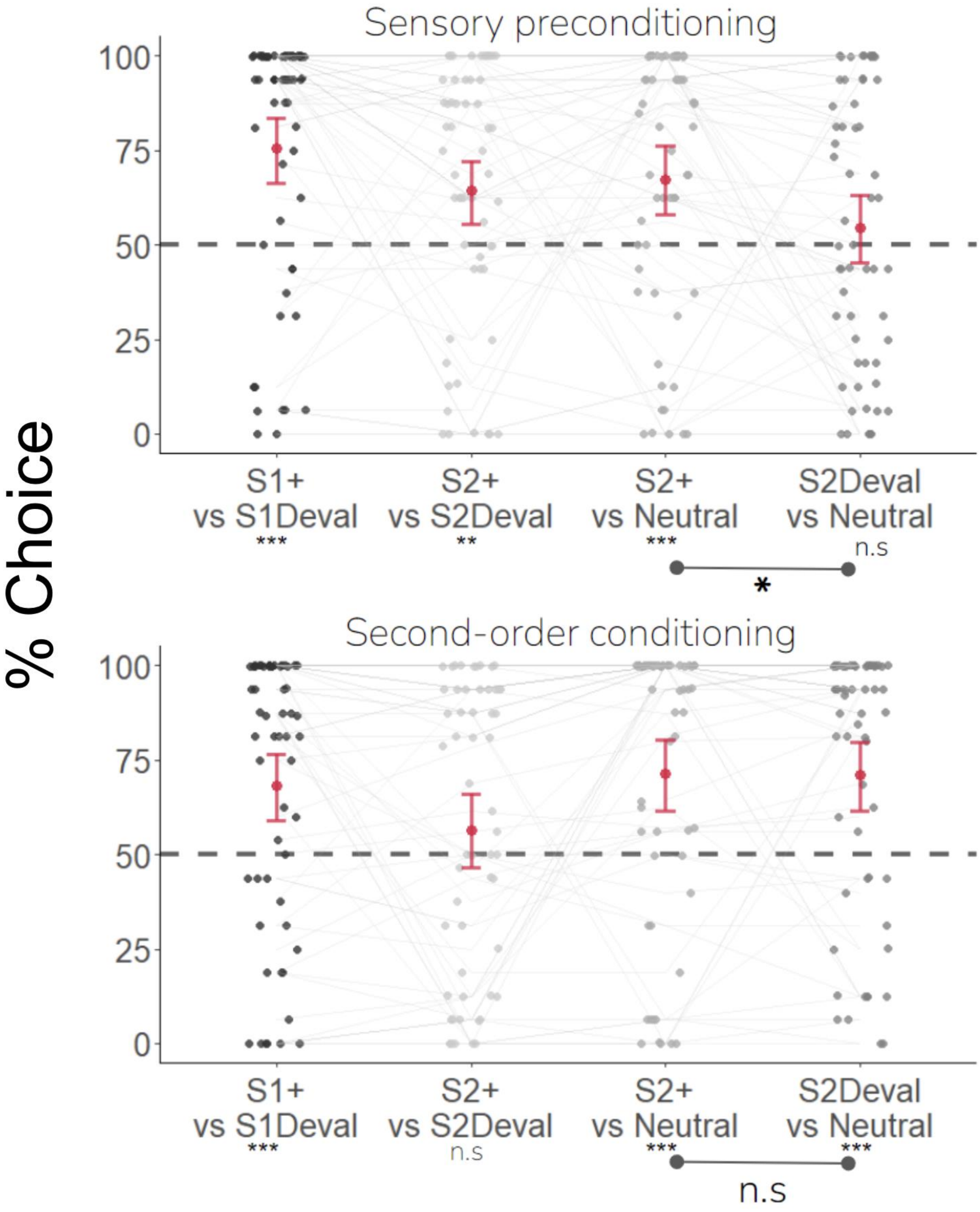

Figure 2.6 **Responding to second-order cues is devaluation-insensitive.** S1+, rewarded first-order cue. S1Deval, devaluation-associated first-order cue, S2+, second-order cue associated with S1+, S2Deval, second-order cue associated with S1Deval. Neutral, neutral cue. n.s = not significant. * p < .05, ** p < .01, *** p < .001

Exploratory analysis: The results of Exp. 2 replicated Exp. 1 using devaluation instead of extinction. Exp. 2 suggests the associative structure second-order conditioning differs from sensory preconditioning in a manner that aligns with most of the literature using sequential presentations and rodents as participants. However, one may point out that there is significant variability between participants in the S1+ vs. S1Deval condition.

To investigate the influence of this variability, participants that chose S1+ over S1Deval less than 60% of the time were excluded from the analysis. Excluding those subjects confirmed the robustness of the results from the full sample. All the one-sample t-tests in second-order conditioning were significant (ts(35) > 2.805, ps < .008), ds > .474). Again, there was no significant difference between the S2+ vs. Neutral condition and the S2Deval vs. Neutral condition (t(34) = .978, p = .335, d = .165). As in the full sample, this difference was found in sensory preconditioning even though the sample size was smaller than optimal for finding such a difference (t(41) = 2.892, p = .006, d = .446). Only the S2Deval vs. Neutral condition was not significant (t(41) = 0.915, p = .365, d = .141), all the other one-sample t-tests were significant (ts(41) > 4.118, ps < .001, ds > .635), suggesting that the preconditioned cue is affected by devaluation of the outcome.

Lastly, the S2Deval vs. Neutral condition was directly compared across second-order conditioning and sensory preconditioning using an independent sample t-test in the full sample of participants. Although this test lacked the optimal sample size for finding such a difference, the results were nevertheless significant (t(110) = 2.432, p = .017, d = .460). S2Deval in the second-order conditioning version was chosen 70.7% of the time (SD = 35.0) vs. Neutral, compared to only 54.4% of the time in the sensory preconditioning version (SD = 35.9).

Together, these results strongly support the view that the associative structures of second-order conditioning and sensory preconditioning differ: second-order conditioning is supported by a direct-link mechanism and sensory preconditioning by a chained-link (or mediated conditioning) mechanism. Of note is that while preconditioned responding has been shown to be sensitive to extinction (Rizley and Rescorla 1972), and devaluation in rats (Hart et al. 2020), the present study is the first to show that the preconditioned cue is sensitive to devaluation in humans.

Lastly, the exploratory finding that there was a significant difference between second-order conditioning and sensory preconditioning in the S2Deval vs. Neutral condition was unexpected given the sample size being too small for optimal comparison. It is therefore necessary to validate these results in an additional experiment with an appropriate sample size. Furthermore, while Exp 2. addresses questions pertaining to the behavioral modes that support higher-order conditioning as described by Balleine and Dickinson (Dickinson and Balleine 1994), it does not directly address whether the cues had access value in a model-free or model-based manner. This question was investigated in Experiment 3 using a conditioned reinforcement test in an appropriate sample size.

Experiment 3

One potential caveat to the results of Exp. 1 and 2 is that it is not clear whether the second-order cue in and of itself has intrinsic value as would be predicted by model-free learning, and whether the preconditioned cue does not. This is an important question because there could be several (not mutually exclusive) possibilities for why the associative structures of second-order conditioning and sensory preconditioning differ.

One possibility is that during sensory preconditioning, participants make an inference at Test which causes them to expect a reward whenever they see the preconditioned cue. Model-based learning envisions that this value is inferred through a mental simulation mechanism that retrieves successor stimuli. There is correlational evidence for this in humans (Wang, Schoenbaum, and Kahnt 2020; Barron et al. 2020). Another possibility is that there may be some mediated transfer of value from the first-order cue that has directly been paired with reward onto the associated preconditioned cue during the conditioning phase which may allow the preconditioned cue itself to accrue model-free value. In light of Rescorla's hypothesis about simultaneous presentations, this later possibility seems especially pertinent (Doll and Daw 2016; Wimmer and Shohamy 2012; Holland 1981; Kurth-Nelson et al. 2015). This is because simultaneous presentations may reactivate the preconditioned cue at the time of conditioning, allowing it to acquire value like a directly-conditioned cue, thereby making the acquisition of model-free value more likely.

Importantly, Exp. 2 does not answer this question directly. The preconditioned cue could maintain model-free value independent of devaluation of the outcome because the association between the first-order cue and the devalued outcome has not been directly experienced. First-order cues are normally sensitive to devaluation at Test in rats and humans (Costa et al. 2022; Howard et al. 2020; Pool et al. 2019). However, rats will continue to press a lever for presentations of first-order cues, even if its associated outcome has been devalued (Parkinson et al. 2005). This has been taken to mean that model-free value is cached in first-order cues which is sufficient to support lever pressing, but at the same time, the cue itself can access value in a model-based manner at Test which is why it is (normally) sensitive to devaluation. Indeed, Hart et al., (Hart et al. 2020) found that preconditioned cues and first-order cues (normally) are sensitive to devaluation at Test, and Sharpe et al. (Sharpe, Batchelor, and Schoenbaum 2017) found that rats are willing to press a lever for presentations of the first-order cue and the second-order cue, but they are not willing to do so for the preconditioned cue. So far, this hypothesis has not been confirmed in humans (Kahnt and Schoenbaum 2021). Therefore, Exp. 3 aims to address this gap.

To summarize, based on the results of Exp. 2 and Sharpe et al., (Sharpe, Batchelor, and Schoenbaum 2017), the preconditioned cue should not support conditioned reinforcement. In contrast, based on the results of Exp. 2 and on correlational evidence in humans, it is hypothesized that the second-order cue should support conditioned reinforcement (Luettgau et al. 2021). Exp. 3 used a conditioned reinforcement test to test these hypotheses.

Methods

Participants: 264 participants in total were recruited through the Prolific online platform. Of these, 133 were recruited for the second-order conditioning study, and 131 were recruited for the sensory preconditioning study. Power analysis from piloting determined that 116 participants were needed in each study for attaining power of 0.80. After applying the exclusion criteria (Supplementary Table 2.7), the final sample size in the second-order conditioning experiment was 120 (***Mean age*****:** 41.46; ***Range*****:** 19-75; ***Gender*****:** 36 Female, 1 Prefer not to say; ***Ethnicity***: 9 Asian, 11 Black, 91 White, 4 Other, 5 Mixed). The final sample size for sensory preconditioning was 120 after applying exclusion criteria (***Mean age*****:** 39.17; ***Range*****:** 19-74; ***Gender*****:** 38 Female, 1 Prefer not to say; ***Ethnicity:*** 9 Asian, 13 Black, 92 White, 1 Other, 5 Mixed). The experiment was approved by the University of California, Los Angeles Institutional Review Board and all participants provided informed consent.

Materials: The stimuli were the same as in Exp. 1 and Exp 2. The design of the experiment is summarized in Supplementary Tables 2.8 and 2.9.

Procedure: The procedure is presented in Figure 2.7. The trial numbers and counterbalancing procedures for conditioning, second-order conditioning, preconditioning, and Test were the same as in Exp. 1 and Exp. 2.

Conditioned reinforcement: After the Test phase, participants were instructed that the main part of the experiment was over and that they would now play a game in which they unboxed boxes in which “Gustav the Ghost” hid art pieces. They were told that they could help unbox the boxes to reveal an “art piece” inside. The stimuli represented the art pieces. Importantly, participants were told that they did not have to unbox the boxes and to only do so if they wanted to and that not unboxing the boxes did not affect their bonus in any way. This was to keep the motivation of the conditioned reinforcement test as close as possible to rat conditioned reinforcement tests which test the extent to which rats are willing to press a lever during a certain period for the presentation of a cue.

Two boxes appeared on top of each other with a gap in between at the start of each trial. Participants were told that they could, if they wanted to, “unbox” one of the boxes by pressing either the top or bottom arrow keys on their keyboard for a one second presentation of a cue. The location of each cue stayed the same across all trials. For example, if S1+ was “hidden” under the top box on trial one, it was also there on all subsequent trials. The location of each cue’s hidden location was randomized between participants and between the first and second conditioned reinforcement test.

Each trial lasted for four seconds and participants had three seconds to respond, with an ITI of 500 ms. The first conditioned reinforcement test was between the S2+ and Neutral (2) (i.e., associate cues). The second conditioned reinforcement test was between S1+ and Neutral (1) (i.e., first-order cues). This order was to ensure that the conditioned reinforcement for the associate cues was not influenced by the first-order cues. No reward was presented during the conditioned reinforcement test. There were 12 trials total, per session. The first two trials of each session were considered as “orientation” trials and were therefore discarded from analysis, leaving 10 valid trials for analysis per session.

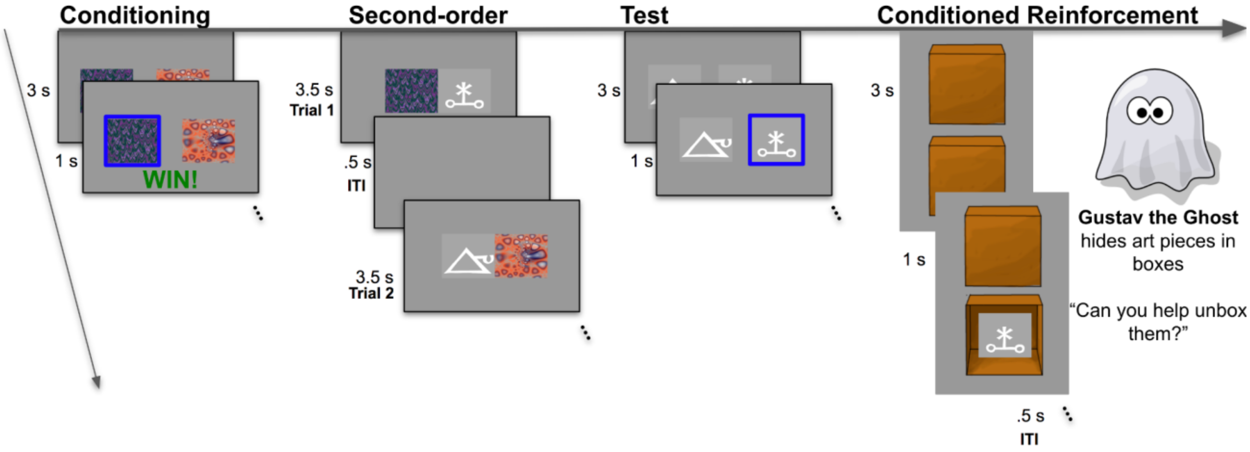

Figure 2.7 **Second-order conditioning task with conditioned reinforcement test.** During conditioning, participants learned associations between each first-order cue (S1+) and a monetary rewarded outcome ($). On each trial, participants had to choose between the presented cues by pressing the left or right key corresponding to the location of the cues presented on the screen. Then, feedback was presented on the screen. A blue square surrounded the chosen option. In the second-order conditioning phase, participants learned associations between cue pairs. In sensory preconditioning, the order of conditioning and second-order conditioning was flipped such that conditioning happened after the preconditioning phase (i.e., equivalent of the second-order phase). During Test, participants were asked to choose the stimuli they thought would maximize their monetary outcome. No feedback was provided during Test. Lastly, participants completed the conditioned reinforcement test in which they “unboxed” one of two boxes stacked on top of each other to reveal a stimulus. Each stimulus was in the same box each trial. In the first conditioned reinforcement test, opening one box led to presentation of cue S2+ and opening the other box led to presentation of the Neutral (2) cue. In the second conditioned reinforcement test, opening one box led to presentation of S1+ and opening the other led to presentation of Neutral (1). Each trial was followed by a 500 ms ITI. ITI, inter-trial interval. S1+, rewarded first-order cue.

Analysis: A mixed-design repeated measures analysis of variance (ANOVA) was conducted to compare the two dependent measures, Test and Conditioned Reinforcement behavior, across second-order conditioning and sensory preconditioning with [Experiment (second-order conditioning vs. sensory preconditioning)] as the between-subjects factor and [Condition (first-order cues vs. second-order cues)] as the within-subjects factor. All analysis for simple main effects were planned and orthogonal and therefore did not require controlling for multiple comparisons.

Results and Discussion

Test phase: Results from the mixed measures ANOVA [Experiment (Second-order conditioning vs. Sensory Preconditioning) X Condition (first-order cue vs. associate cue)] showed that there was a main effect of Condition ($F_{(1, 238)} = 125.20$, $p < .001$, $\eta^2 = .183$), demonstrating robust first-order conditioning as well as significant second-order conditioning and sensory preconditioning ($ts(120) > 3.434$, $ps < .001$, $ds > .314$) (Figure 2.8). There was also no difference between experiments in overall conditioning ($F_{(1, 238)} = 2.60$, $p = .109$, $\eta^2 = .005$). However there was a weak but significant Condition X Experiment interaction ($F_{(1, 238)} = 3.949$, $p = .048$, $\eta^2 = .006$). Planned post-hoc comparisons revealed that this difference emerged because sensory preconditioning was a relatively weaker effect ($M = 61.2$, $SD = 35.8$) compared to second-order conditioning ($M = 70.4$, $SD = 35.4$). It is not unusual that sensory preconditioning is relatively weak compared to second-order conditioning. Also, of note is that as in other studies, there was large variability between subjects in both sensory preconditioning and second-order conditioning (Luettgau et al. 2021; Wimmer and Shohamy 2012; Kurth-Nelson et al. 2015).

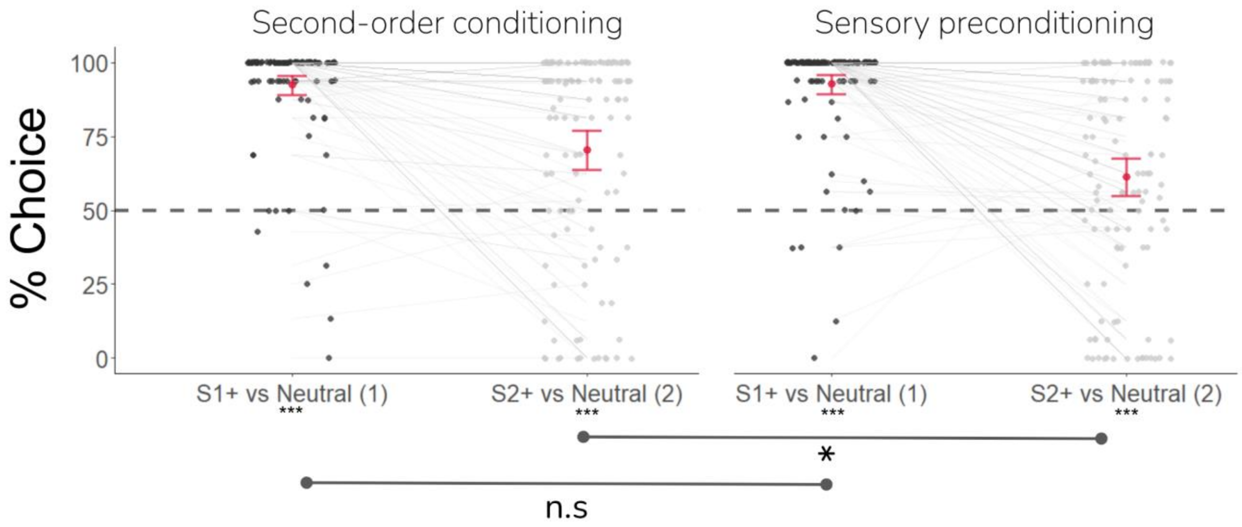

Figure 2.8 **Experiment 3 Test phase results.** S1+, first-order cue; Neutral (1), neutral first-order cue; S2+, second-order cue and preconditioned cue for second-order conditioning and sensory preconditioning respectively; Neutral (2), neutral second-order cue); n.s = not significant. * p < .05, *** p < .001

Conditioned reinforcement: Following the Test phase, participants were given two conditioned reinforcement tests. In the first, participants could open one of two boxes for a one second presentation of the S2+ or Neutral (2). In the second, participants could open one of two boxes for a one second presentation of S1+ or Neutral (1). That is, across both experiments, participants opened the boxes for the cue paired directly or indirectly with reward. A mixed-effects ANOVA [Experiment (Second-order conditioning vs. Sensory Preconditioning) X Condition (first-order cue vs. associate cue)] showed that there was a main effect of Condition ($F_{(1, 238)}$ = 113.64, p < .001, η^2 = .147); both first-order and associate cues supported conditioned reinforcement to a larger extent than neutral cues, but first-order cues did more so (ts > 2.831, ps < .005, ds > .258) (Figure 2.9).

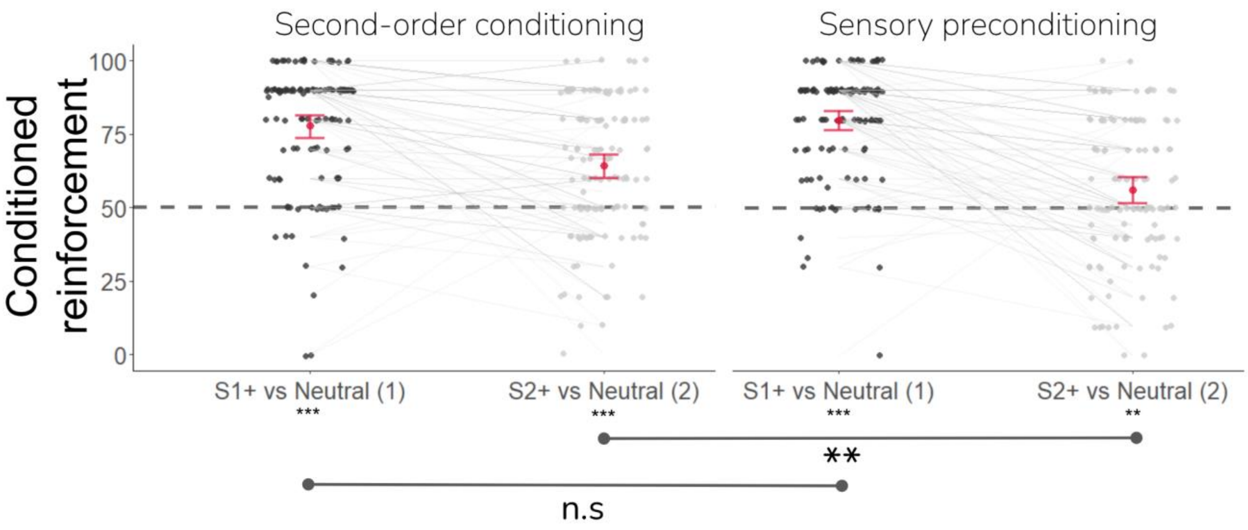

Figure 2.9 **Second-order cues support conditioned reinforcement more than preconditioned cues.** Experiment X Condition interaction ($F_{(1, 238)}$ = 8.28, p = .004, η^2 = .011). S1+, first-order cue; Neutral (1), neutral first-order cue; S2+, second-order cue and preconditioned cue for second-order conditioning and sensory preconditioning respectively; Neutral (2), neutral second-order cue); n.s, not significant. ** p < .01, *** p < .001

There was no difference between experiments in overall conditioned reinforcement effects as revealed by the lack of significant Experiment main effect ($F_{(1, 238)}$ = 1.61, p = .205, η^2 = .004). Critically, there was a significant Experiment X Condition interaction ($F_{(1, 238)}$ = 8.28, p = .004, η^2 = .011). Planned contrasts show that this effect emerged because the second-order cue could support conditioned reinforcement to a larger extent (M = 64.3, SD = 23.5) than the preconditioned cue (M = 56.3, SD = 24.6) (t(237) = 2.755, p = .012, d = .356).

Although second-order conditioning and sensory preconditioning were both demonstrated, sensory preconditioning was a slightly weaker, albeit significantly so, effect compared to second-order conditioning. Importantly, the main hypothesis of the present experiment is to test whether the preconditioned cue has value relative to the second-order cue (and not compared to the Neutral cue). The fact that that sensory preconditioning was weaker thus may lead one to rightly question whether the conditioned reinforcement effects are significant as a result of this difference, and not due to a difference between the two cues in terms of their cached value.

If preconditioned cues have no value but second-order cues do, then we should expect to see a significant correlation between second-order conditioning and conditioned reinforcement. Critically, there should not be such a correlation in the sensory preconditioning experiment. Namely, the extent of preconditioning should not correlate with the extent of conditioned reinforcement, despite strong preconditioning. Before running this analysis, it is important to note that as in other higher-order conditioning studies, there were large degrees of between subject variability. That is, since some participants did not demonstrate second-order conditioning and sensory preconditioning. Those participants may have also had weak conditioned reinforcement effects. Therefore, a correlation between higher-order conditioning and conditioned reinforcement may lead to the misguided conclusion that both second-order and preconditioned cues support conditioned reinforcement. Indeed, such an analysis on the full sample of participants showed that second-order conditioning was correlated with conditioned reinforcement ($r = .655$, $p < .001$), but so was sensory preconditioning ($r = .671$, $p < .001$). The correct way to address this is to only include participants that demonstrate strong second-order conditioning and sensory preconditioning.

After only including participants who demonstrated strong sensory preconditioning or second-order conditioning (>70% S2+ vs. Neutral), only second-order conditioning was correlated with conditioned reinforcement ($r = .286$, $p = .013$, $M = 95.0$, $SD = 6.1$). Importantly, sensory preconditioning was not correlated with conditioned reinforcement ($r = .104$, $p = .449$, $M = 93.8$, $SD = 7.9$) (Figure 2.10). Together, this suggests that the second-order cue can support conditioned reinforcement while the preconditioned cue does not. Therefore, these results replicate in humans a finding so far only demonstrated in rats, namely that preconditioned cues have no value (Sharpe, Batchelor, and Schoenbaum 2017).

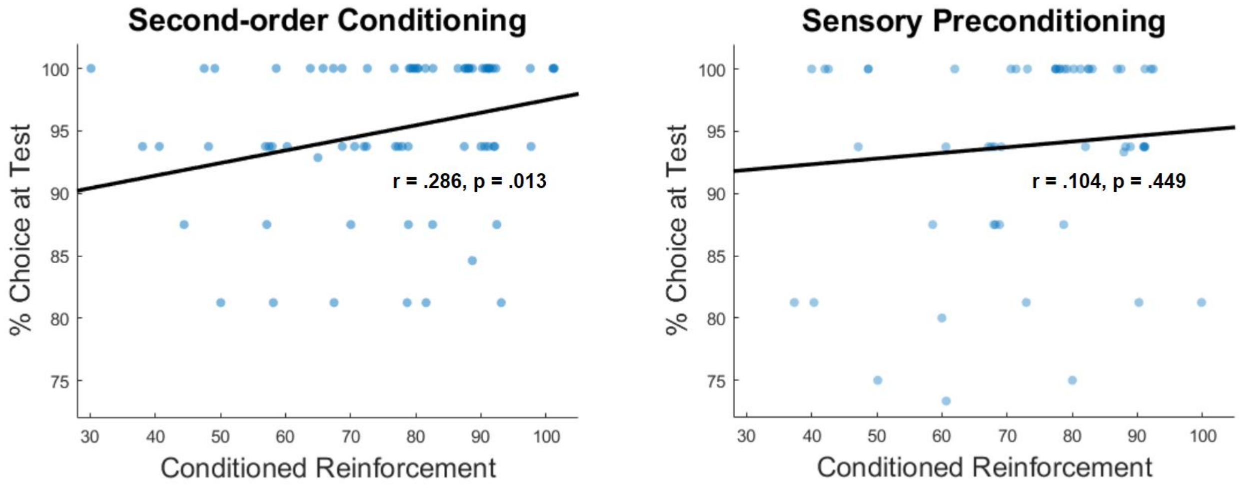

Figure 2.10 **Second-order cue, but not preconditioned cue responding correlates with conditioned reinforcement.** After only including preconditioned and second-order conditioned participants, only second-order conditioning correlated with conditioned reinforcement. Critically, preconditioned cues did not, suggesting preconditioned cues have no value.

Normally, conditioned reinforcement to first-order cues is not sensitive to devaluation of the outcome (Parkinson et al. 2005). That is, rats are willing to press a lever for presentations of the first-order cue, independent of what the cue itself predicts. The interpretation has therefore been that cached value in the cue (or model-free value) is enough to support conditioned reinforcement (Sharpe, Batchelor, and Schoenbaum 2017). Conversely, if a cue cannot support conditioned reinforcement, that may suggest that it does not have cached or model-free value. Given this, the present result suggests that at least in the procedure used in this experiment, the preconditioned cue does not have value.

An interesting question is why the (model-free) motivational properties of second-order cues allow them to support conditioned reinforcement. This is especially interesting given that previous results have only been demonstrated so far in rats in which a conditioning procedure is likely more of significance than it is to humans; intuitively, it should not have the same motivational properties in rats as in humans. The difference is that rat studies typically use primary reinforcers while the present experiment used secondary. Nonetheless, participants were still willing to work to obtain presentations of the second-order cue, but not for the preconditioned cue. Indeed, previous studies have shown that humans can flexibly assign value to abstract stimuli or events when they are used as an outcome, and standard reinforcement learning can explain such learning (McDougle et al. 2021). These findings were extended in the present experiment, suggesting second-order cue becomes “good”, such that participants are willing to work to obtain them (Seitz, Blaisdell, and Sharpe 2021). The boundary of such an ability is unclear and should be investigated in future studies. The present study suggests at a minimum, that primary and secondary reinforcers share some motivational properties.

One potential caveat with the present results is that the results from the conditioned reinforcement test may simply reflect a desire to “help” the experimenter, thus potentially leading to suboptimal participant behavior and misleading the experimenter to draw erroneous conclusions about the data (Doll et al. 2009). Indeed, participants were asked to “help” unbox the art. Along those lines, notably, all the participants completed the conditioned reinforcement tests with very few missed trials in total, suggesting participants were motivated throughout the task. This may have been the case even though participants were explicitly instructed that they did not have to respond during the conditioned reinforcement test, and that the outcome of the conditioned reinforcement test would not affect their bonus in any way.

While it is clear that participants were motivated, the results suggest that the motivation was not simply due to a desire to “help”, since it cannot explain why second-order cues support conditioned reinforcement while preconditioned cues do not. Furthermore, first-order cues in both sensory preconditioning and second-order conditioning supported conditioned reinforcement, and to a similar extent. Again, it is not clear why responding would not differ between first-order cues but would for second-order and preconditioned cues if the sole motivation was to help the experimenter. Instead, what seems clear is that second-order and preconditioned cues do not share the same motivational properties as it pertains to performance on the conditioned reinforcement test.

To summarize, second-order cues can support conditioned reinforcement, while preconditioned cues do not. This suggests that second-order cues have model-free value. This is compatible with the direct-link account for second-order conditioning, and the chained-link account for sensory preconditioning.

General Discussion

These three experiments together strongly suggest that the associative structure of higher order conditioning in humans differs between second-order and sensory preconditioning. The first two experiments found that even when simultaneous presentations are used, the direct-link account more readily explains the structure of second-order conditioning while the chained-link account more readily explains the structure of sensory preconditioning. The direct-link account postulates that the second-order cue becomes directly associated with the value of the outcome. This is analogous to model-free value which was confirmed in Exp. 3.

Most of the evidence in non-humans suggest that the most likely structure of second-order conditioning is a direct-link, and a chained (or mediated) structure in sensory preconditioning; especially when sequential presentations are used. The present study is the first to suggest that simultaneous second-order conditioning is underpinned by a direct-link structure. This finding was surprising and contrary to the initial hypothesis.

It is possible that despite simultaneous presentations, participants still represented each pair as two elements of a compound instead of one configural stimulus. If that were the case, as

Rescorla argued, then S2 could enter a “between-event” association with S1 (and likewise for the Neutral controls), instead of a “within-event” association (Rescorla 1981; Rescorla 1982). Following Rescorla’s logic, when cues are presented first as a compound, as in the sensory preconditioning experiments here, there is less opportunity for them to be represented as independent elements (at least until conditioning). Indeed, second-order conditioning first exposes participants to one of the cues as an element before as part of a compound (the instructions emphasized that the two cues should be considered part of a whole stimulus, and the two cues were presented with no gap in between). If this were the case, then we should perhaps have expected relatively low conditioned reinforcement to the first-order cue in sensory preconditioning compared to second-order conditioning. This is because elements of a compound elicit relatively low levels of conditioned reinforcement (González et al. 2022). Instead, the first-order cues elicited very similar levels of responding across experiments. Of course, subsequent phases may have broken the potential compound representation in sensory preconditioning. There may have also been a ceiling effect on responding to the first-order cues. The procedure used here did not differ significantly from other simultaneous procedures (e.g., (Wimmer and Shohamy 2012).

It is difficult to say whether in the present experiment participants truly represented the pairs as a compound, or whether they were represented as two elements of a pair. Even so, this was not the main goal of the experiments, which aimed to test the associative structure of higher-order conditioning, *even when* simultaneous presentations are *presented,* not *represented*. Thus, in a sense, the procedure was even more strict for the emergence of evidence for the direct-link structure than it would have been with sequential presentations. Indeed, what the right conditions are for compound versus elemental representations is still debated (Alvarado and Rudy 1992; Williams 1994). Future studies could more closely test what those conditions are and how that affects the associative structure of higher-order conditioning.

It is worth emphasizing that these three experiments do not necessarily isolate model-free from model-based behaviors. That is, the definition of model-free behavior has rested on the assumption that model-free behaviors are not sensitive to extinction and devaluation, as was shown in Exp. 1 & 2. While this may be true, it is also possible that the extinction and devaluation treatments simply failed to affect the model being used in second-order conditioning. There are a number of conditions in which this may happen (e.g., if the original model is too strong). Another factor could perhaps be because the choice task does not lend itself so easily to the expression of a relatively weak model compared to strong model-free action. For example, one study measuring behavior in two ways simultaneously found that there are two Pavlovian systems operating in parallel, one outcome sensitive and another outcome insensitive, each expressed differently (Pool et al. 2019). Compared to model-free behavior, Isolating model-based behavior is easier as devaluation sensitivity can only be based on inferred expectations and because model-free behavior should lead to a qualitatively different response.

As discussed in the introduction, mediated learning is an alternative explanation for responding to the preconditioned cue that does not require inference at the time of test (Holland 1981). This explanation has correlational evidence from simultaneous sensory preconditioning in humans (Wimmer and Shohamy 2012), but one study failed to find evidence for it in second-order conditioning (Craddock et al. 2018). Mediated conditioning posits that the preconditioned cue is reactivated during conditioning. Simultaneous presentations should make acquisition of model-free value more likely, fitting mediated conditioning predictions. This did not appear to occur in Exp. 3, which showed that preconditioned cues do not support conditioned reinforcement, the gold-standard for assessing model-free value. However, it is also possible that behavior is driven by several independent mechanisms; model-based inference may occur in parallel with mediated learning. Hence, the present results do not exclude the possibility of mediated learning as a mechanism. Instead, the results suggest that if responding to the preconditioned cue is supported by mediated learning, then that does not cause the preconditioned cue to accrue model-free value.

By looking at the results from all three experiments, one may note that there were some levels of individual differences in choice behavior. Indeed, even Pavlov noted that there was significant variability between subjects in terms of their conditioned response to the second-order stimulus (Honey and Dwyer 2021). Individual differences are common in human associative learning studies but have been relatively neglected until recently. Computational models are now being developed to elucidate them as a difference between learning and performance (Byrom and Murphy 2016; Lee 2021; Honey and Dwyer 2021; Honey, Dwyer, and Iliescu 2020). Some studies have taken advantage of this variability to attempt to elucidate the neural mechanisms underlying higher-order conditioning (Kurth-Nelson et al. 2015; Wimmer and Shohamy 2012). These two approaches could be combined in future experiments (Wuensch, Pool, and Sander 2021).

Overall, the data support the idea that during second-order conditioning, events that have never been explicitly paired can be linked by taking advantage of the relational structure between the first-order and second-order cues. Importantly, unlike first-order cues and preconditioned cues, second-order cues are not affected by treatments aimed at manipulating its relationship with the outcome through extinction and devaluation. Furthermore, second-order cues can support conditioned reinforcement, suggesting that they accrue cached value themselves in a model-free manner. The equivalent is not true for preconditioned cues, which do not support conditioned reinforcement, suggesting that preconditioned cues do not have value.

These experiments are important because the mechanisms they study have been used to explain disorders such as obsessive-compulsive disorder, substance abuse disorder, anorexia nervosa, and anxiety, as a failure of adaptive decision-making. These have all been proposed as failure to use inference to form outcome expectations (Kahnt and Schoenbaum 2021; Kahnt 2022). By using computationally defined frameworks, we can start to understand why some individuals are more predisposed than others and provide new avenues to individualized treatment. Therefore, a detailed understanding of the neurocomputational mechanisms that support them is necessary. One way to study this is by using decoded neuroreinforcement (Kahnt 2022).

Chapter 4

Establishing whether unconscious learning can support model-based inference

Introduction

Everyday decisions are shaped by past experiences. Choices that previously led to good outcomes are likely to be repeated and choices that led to bad outcomes are likely to be avoided (Thorndike 1898). Model-free decisions rely on precomputed values stored or ‘cached’ in the relevant cues during previous experiences. Responding reflects associations between the cues and responses that have been directly reinforced in their presence (Doya 2008). While efficient and robust, model-free decisions do not take into account changes in the value of the expected reward.

Often however, we are faced with situations that involve decisions between options we have never considered before. Those situations may call for making inferences based on indirectly associated information. In contrast to model-free, model-based decisions solve such a problem by integrating information we have learned in the past, and use that information to guide current decisions, allowing adaptive and flexible responding (Daw 2018). By utilizing the associative structure of the environment, value can be inferred when needed.

Evidence suggests that when the associative structure is established consciously, the information transfer process supporting model-based decisions can occur outside conscious awareness (Wimmer and Shohamy 2012; Garvert, Dolan, and Behrens 2017). That is, when the relationship between consciously experienced events are represented as a structure, participants seem to be able to make model-based inferences using that information, despite not being able to report how the structure is organized (i.e., the events are consciously experienced but the structure is *implicitly* represented).

It is unknown whether this model-based inference can also be accomplished when the events themselves are represented *unconsciously*. That is, without conscious experience of the event in the first place. Indeed, it has long been debated whether unconscious processes share the same abilities as conscious ones (Hassin 2013; Newell and Shanks 2014). Some have argued that model-based decisions are a sign of consciousness (Frith 2019).[1] Dual-systems theories argue that model-free and model-based learning rely differentially on explicit processes, with model-based learning requiring utilization of explicit processes. The present experiment therefore asks whether model-based decisions can be supported even in the complete absence of conscious experience of the associative structure - i.e., when the associative structure is unconsciously learned, thereby precluding the possibility of explicit processes.

Model-based decisions are flexible but require high computational cost. Using their representations can be too difficult, even for realistically large problems (Momennejad et al. 2017). Instead, a feature map known as the *successor representation* (SR) can simplify the computation of future rewards by combining the efficiency of model-free learning with some of the flexibility of model-based learning. An increasing amount of evidence has found support for the proposal that the brain uses SR as part of its repertoire to enable inferences of indirect relationships among events (Stachenfeld, Botvinick, and Gershman 2017; Garvert, Dolan, and Behrens 2017; Momennejad et al. 2017; Geerts et al. 2020).

The present experiment used real-time fMRI neuroreinforcement combined with a version of sensory preconditioning to identify model-based behavior when the events that establish their associative structure is learned unconsciously (Brogden 1945; Wimmer and Shohamy 2012;

---

[1] Frith did not argue that model-based inferences require consciousness - just that its content should be active in working memory, consciously or unconsciously.

Sharpe et al. 2017; Sharpe, Batchelor, and Schoenbaum 2017). Classic sensory preconditioning involves three phases. First, an association is established between the preconditioned cue and first-order cue (S2-S1, respectively); followed by a phase in which only S1 is associated with reward (i.e., Conditioning phase). This causes subjects to respond to cue S2 (i.e., the preconditioned cue) over control stimuli in the last and critical test phase. Importantly, participants respond to cue S2 as if it is predictive of reinforcement despite the fact that S2 had never been directly reinforced.

The key distinction between the procedure used here from the classic sensory preconditioning procedure is that reward is introduced to teach the S2-S1 association during the first phase. This is a necessary feature of the unconscious learning technique since reinforcement is one of its driving forces. Inferences that solve sensory preconditioning in the classic task are based on model-based representations (Sharpe et al. 2017; Gardner, Schoenbaum, and Gershman 2018).

This distinction is important because once a stimulus is associated with reward, SR can update its value and is therefore sensitive to devaluation (Gardner, Schoenbaum, and Gershman 2018). Therefore, in the second phase of the novel sensory preconditioning procedure when the first-order cue is punished (the equivalent of devaluation), the preconditioned cue could also become associated with the same outcome through an outcome-sensitive SR mechanism.

In the SR model, only stimuli that are already predictive of value can change their value after devaluation. This contrasts with model-free learning which cannot handle devaluations. Importantly, SR computations are model-based in a sense that they utilize a model of the reward structure of the environment. They are different from purely model-based models because they cannot update the transition structure, which is not necessary in this version of sensory preconditioning. Thus, these two computations make different predictions about the behavior that they support.

The task used here pits the two mechanisms against each other. During preconditioning, over the course of three days of neuroreinforcement training, all participants should learn that the consciously presented first-order cue (i.e., dots stimulus) on the screen is associated with value in a model-free manner (Figure 3.1). Likewise, if participants learn through decoded neuroreinforcement to evoke S2, it is expected that S2 should also acquire model-free value. This is because evoked cues are sensitive to reinforcement (Holland 1990; Holland 1981). Then, the conditioning phase is aimed at separating a model-free learner from a learner using SR by heavily punishing the first-order cue. Therefore, a model-free learner should be unaffected by this punishment, and pick it more often than a novel Neutral Control at Test because they lack a vector of all the state rewards (Momennejad et al. 2017). Punishing the dot motion also avoids a potential mere exposure confound. This is because since the dot motion is presented during induction, it may acquire preference simply by being more familiar than the other stimuli. Hence, it would be difficult to conclude whether any preference for the dot motion is due to it being associated with reward, or mere exposure.

In contrast, a learner using SR should have learned the association between the first-order cue and the preconditioned cue, and should therefore avoid the preconditioned cue by picking the novel Neutral Control more often, because the first-order cue it is associated with was punished during Conditioning.

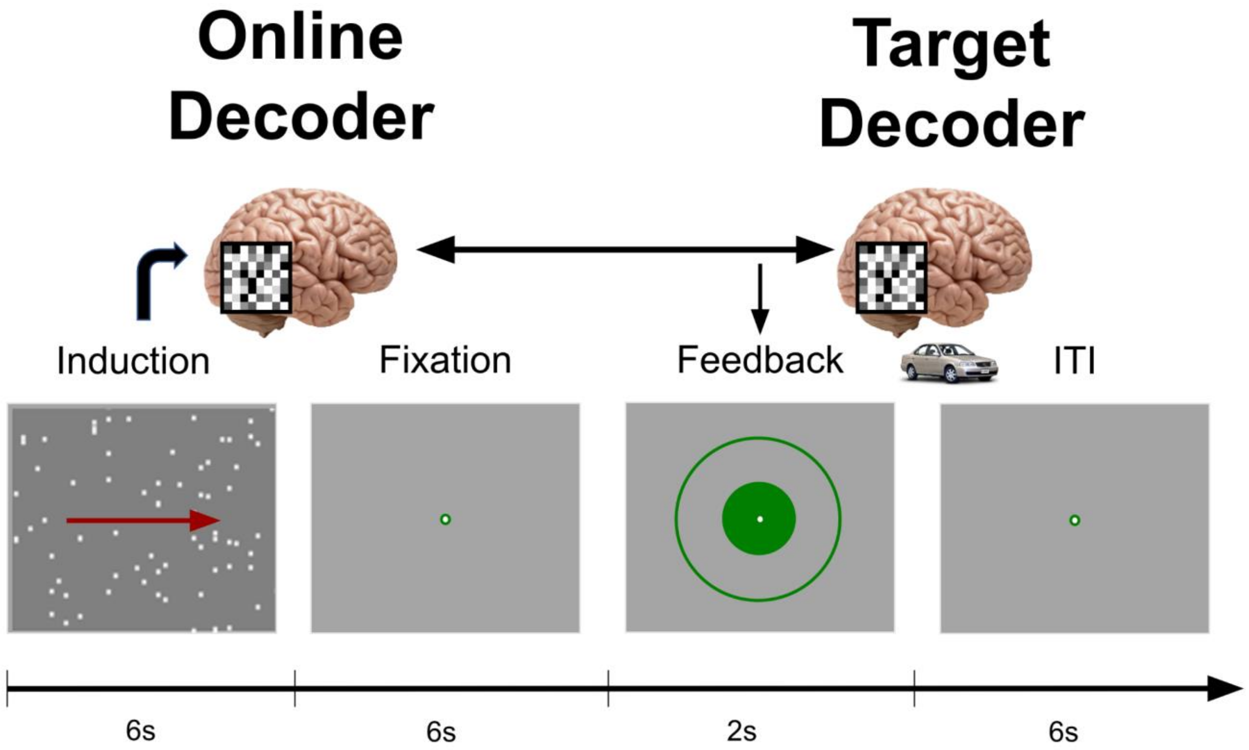

Figure 3.1 **Associative neuroreinforcement procedure**. Online decoding is used to reinforce the target category (cars in this example) while dot motion is presented on the screen. The size of the disc is proportional to the likelihood of the target being represented, and to monetary gain. Brain activity pattern representing cars is endowed with value.

Methods

Participants: 16 participants were recruited for the study (***Mean age*** = 22.2, ***Female*** = 7). The expected effect size based on a review of the decoded neurofeedback literature indicated that the average effect was 1.02. The experiment that most matches the present design had a behavioral effect of 1.07 (Amano et al. 2016). Therefore, an expected effect size was around 1.00 as the best approximation based on previous decoded neurofeedback studies. 16 participants are needed to obtain such an effect with a two-tailed paired samples t-test. A chi-square power analysis was also conducted to determine the number of trials required to find such an effect, which revealed that 16 trials per condition is required. To optimize fMRI signal-to-noise, the trial number was doubled. There were therefore 32 trials per condition, and 96 trials in total. All participants provided informed consent, and the study was approved by the Institutional Review Board of University of Florida.

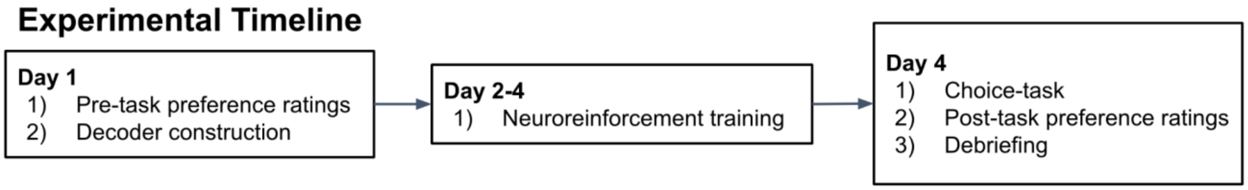

Figure 3.2 **Experimental timeline** for *Establishing whether unconscious learning can support model-based inference.*

Procedure: The experimental timeline is presented in Figure 3.2. The experiment involved three phases (Figure 3.3). 1) Association phase, 2) Conditioning phase, and 3) Test phase. Pre and post-task preference ratings were collected before and after the experiment. Participants were presented with a sheet to rate their preference for each of the categories (from “Extremely/strongly dislike” to “Extremely strongly like”, responses were scaled from 1-7). Categories rated a ‘4’ (i.e., Neutral’) were chosen to be used as stimuli for the Choice task based on criteria specified under “*Decoder Construction”*.

Association phase (Neural-reinforcement): As in Chapter 2, following Decoder Construction, participants came back on a later day for neural-reinforcement sessions. There were 2-3 days between decoder construction and neuroreinforcement, determined by scanner and participant availability. Using multi-voxel neuro-reinforcement, activation of the Target category was paired with reward. While in the scanner, participants were told “while noticing the direction of the dots on the screen, do whatever you can", to maximize the size of a subsequently presented feedback disc. A dot motion stimulus was presented on the screen during the six second induction phase. This stimulus was 100% coherent, moving in either the up, down, left, or right direction, counterbalanced between participants. All other procedures, including the induction calculation method were the same as in Chapter 2. Importantly, the participants did not know what the purpose of the intervention was. Blinding was conducted in the same way as in Chapter 2. That is, neither the participant nor the experimenters knew what category of the Target. Thus, any learning that occurred toward the Target category was unconscious.

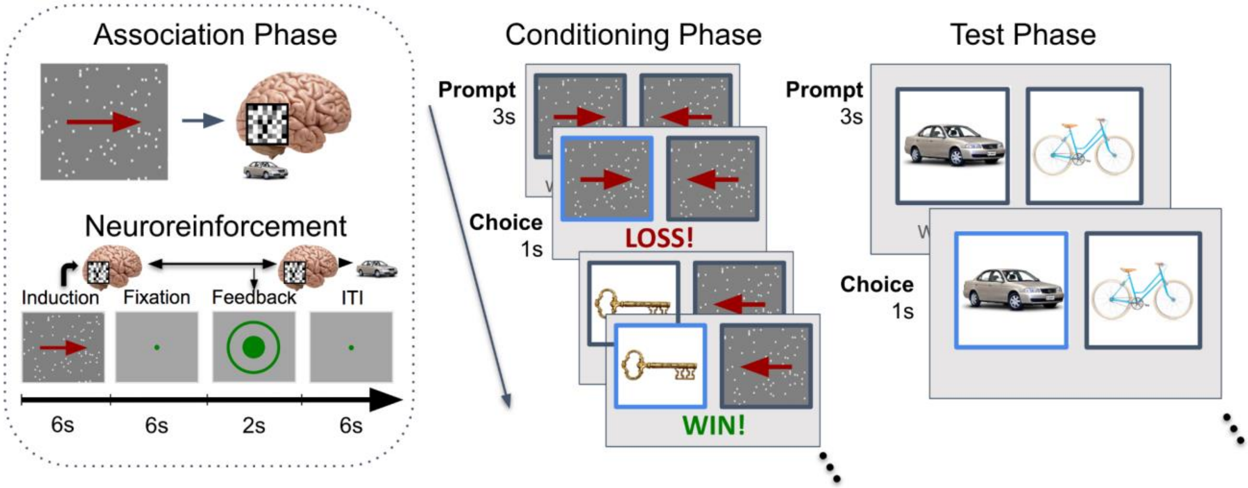

Figure 3.3 **Procedure** for *Establishing unconscious learning can support model-based inference*. Participants underwent three days of neuroreinforcement which aimed to pair a Target category (cars in this example) with reward while a 100% coherent dot motion stimulus was presented on the screen during the Association phase. On the last day, participants completed the Conditioning phase in which they learned through trial and error that picking the dot motion presented during induction results in a heavy monetary loss. Hence, participants should avoid it in the subsequent Test phase. During the Test phase, there was no feedback and participants were told to do their best to pick the stimulus (or stimuli) they thought would lead to a reward. Here, the preconditioned cue (i.e., the stimuli corresponding to the Target category) was presented for the first time along with two neutral control cues. If participants learned to associate the punished dot motion with the preconditioned cue, then they should avoid the Target category during this phase. If the Target category accrued model-free value, then they should pick it more often than the controls. Conversely, if participants learned nothing, then the preconditioned cue should be picked equally as often as the two control category stimuli. Participants used the keys associated with the right and left stimulus on the screen. A blue square surrounded the chosen option in each phase of the choice task.

Conditioning phase: After the Association phase, participants were told that they were going to play a choice task in which they had to choose between different stimuli on the screen inside the scanner. They were told that some stimuli will lead to a reward, others to no outcome, and others to a large loss. The loss was always the dot direction presented to the participants during the Association phase (hereby referred to as S1+) however participants were not told this before the task.

Conditioning phase conditions: The conditioning phase had four conditions and there were four stimuli, hence the stimuli were paired against each other. Choosing S1+ led to a $5 loss from the earnings gained during the Association phase. Choosing a novel dot motion stimulus (orthogonal direction to S1+) did not lead to a loss, or reward (hereby referred to as S1-). The direction of the two dot motion stimuli was always orthogonal to each other and counterbalanced (i.e., up, down, left, right). There were also two category stimuli, Mildly Rewarded Control, and Neutral. Choosing Mildly Rewarded Control led to a 50 cent reward, while choosing the Neutral led to no outcome. The two category stimuli were always within the same animate/inanimate category as the Target and were chosen based on Euclidean distance from the Target, as in Chapter 2. The inclusion of the two category stimuli was to balance reward and punishment throughout the experiment.

Participants chose using the left and right buttons of a button box. The two stimuli were separated by a gap in the middle. Each trial lasted for four seconds and participants had three seconds to make their choice. There was no ITI. After their choice, a blue square surrounded the chosen stimulus to indicate the selected option followed by feedback on whether the choice was a win or not. Participants therefore learned by trial and error which choice was winning and which was not. There was no penalty for missed trials. The total accumulated reward was constantly presented on the top right of the screen. There was also feedback on the screen after each choice that indicated the amount of money they won or lost. Each condition had 20 trials, thus the total number of trials was 120.

Test phase: As in the conditioning phase, participants completed the test phase in the scanner. The test phase was exactly the same as the test phase in Chapter 2. Participants were told that they again would play the choice task but that this time there was not going to be any feedback. Therefore, they needed to do their best to use what they learned throughout the experiment to pick the stimuli they think are rewarded, and avoid picking the stimuli they think are punished. There were 32 trials per condition and four conditions, thus there were 128 trials total. Each trial lasted for four seconds, and participants had three seconds to respond.

Test phase conditions: The choices in the test phase were between A) Target Category (i.e., category stimuli that represents the target brain activity pattern previously trained during induction) vs. Control Category 1, B) Target Category vs. Control Category 2, and C) Control Category 1 vs. Control Category 2, and D) S1+ (punished dot motion) vs. S1- (neutral dot motion). (Supplementary Table 3.1) The Target, as well as the two Controls had not been seen until this point of the experiment. Thus, all of the categories presented during this phase were novel. This way, the mere exposure confound was also avoided. As in Chapter 2, the two control categories were chosen in order to reduce the possibility of false positive results in the comparisons between the target and each of the controls, and to test whether the two control conditions would be picked equally when pitted against each other in order to establish an active baseline; i.e., each control should be chosen around 50% of the time on average across participants. Also, to balance conditions, the dot motion stimuli were never involved in a condition against any of the categories since dots have been punished in the conditioning phase.

MRI Parameters and Behavioral analysis: Same as in Chapter 2.

Results and Discussion

The design of the experiment was aimed at dissociating between two possible outcomes. Namely, if participants learned to associate the Target with value during the Association phase in a model-free manner, then they should be more likely to pick the Target in the choice task over each of the two Controls. On the other hand, if participants learned through a more model-based manner to associate the Target with the simultaneously presented dot motion during the Association phase, then they should be more likely to pick each of the Controls over the Target. This is because in the Conditioning phase, participants learned that the dot motion presented during the Association phase is now associated with loss, hence participants should generalize this to the Target. Alternatively, if there was no learning related to neuroreinforcement during the experiment then the Target should be picked equal to the Controls. The Controls condition thus served as an active baseline for this comparison.

The results are presented in Figure 3.4. Participants learned to avoid the punished S1+ dots over S1- (i.e., non-punished dots) by only picking it 24.6% (SD = 29.9) of the time (t(15) = 3.394, p = .004, d = .848, M = 24.), suggesting the conditioning phase was successful for the first-order cues. As expected, there was no preference between the Controls (t(15) = .516, p = .613, d = .129, M = 48.9, SD = 8.8; i.e., Control 1 was picked 48.9% of the time), suggesting both control conditions were appropriately balanced.

On average, the pooled result from the two critical conditions found that participants chose the Target 51.6% (SD = 28.9) of the time versus the pooled average of the two control stimuli (t(15) = .367, p = .719, d = .092). Overall, the target did not accrue model-free value, or access value in a model-based manner. Indeed, both the Target vs. Control 1 condition (M = 51.5, SD = 30.5), and Target vs, Control 2 condition (M = 51.7, SD = 36.6) were not significant (ts(15) < .197, ps > .846, d < .049). These results suggest that neuroreinforcement did not lead to a specific preference for the stimuli that represent the target pattern induced during neuroreinforcement at the group level. Neither does the result suggest that participants overall learned to associate the dot motion presented simultaneously on the screen (i.e., S1+) with the Target. This is because the participants overall learned to avoid the punished dots in the Test phase, suggesting conditioning was efficient. Instead, the results agree with those in Chapter 2 in that the presence of the Target induced higher levels of between-subject variability in choice; an effect that did not emerge in the Controls condition. Indeed, there was a marginally significant correlation between the Target vs. Controls 1 and Target vs. Controls 2 condition (r = .472, p = .065).

Perhaps one important caveat to the above results is that there were a few participants that picked the punished dots *more often* (n = 5, M = 63.2). That is, these participants all picked the punished dots more than 50% of the time. Three of these participants also did not avoid the punished dots in the conditioning phase. To investigate the influence of these participants they were excluded in the exploratory analysis.

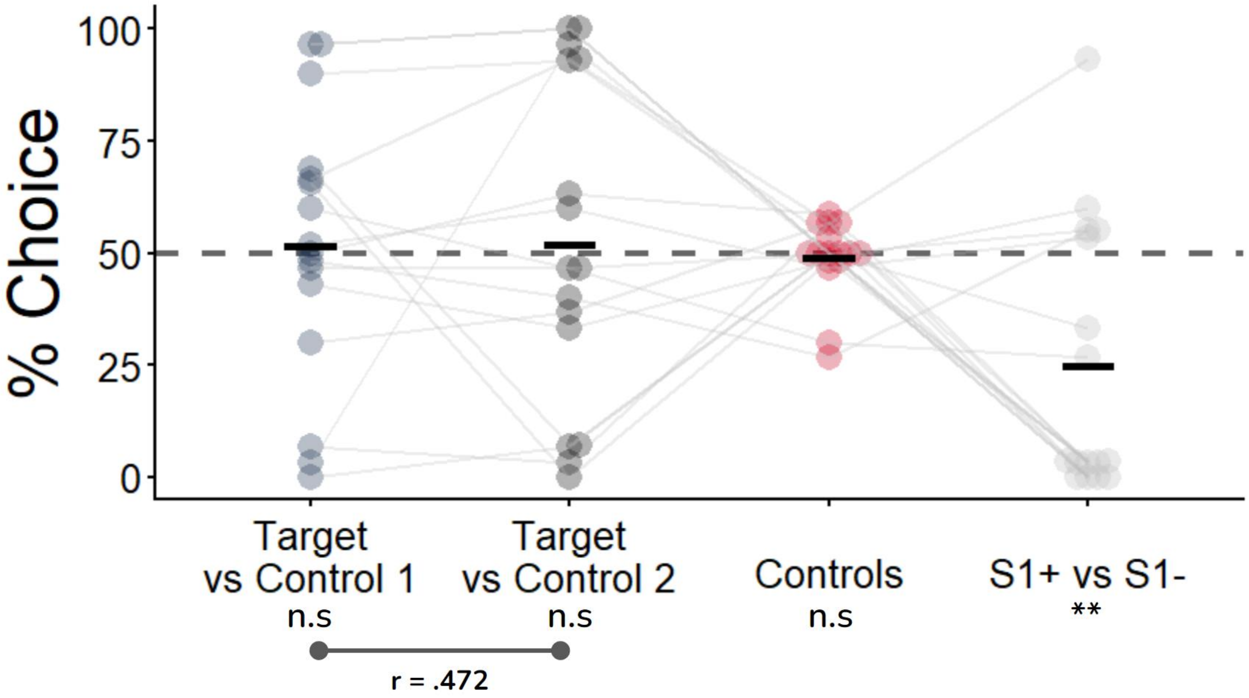

Figure 3.4 **Choice task results.** Dots indicate individual participants. Center lines represent means for each condition. Connecting lines represent individual data across conditions. n.s, not-significant; ** p > .01.

The pooled condition (i.e., Pooled Target vs. Pooled Neutrals) remained non-significant after exclusions (t(10) = 1.517, p = .160, d = .457). Numerically, the target was chosen more often than the controls (M = 61.7, SD = 25.6). The two critical conditions (i.e., Target vs. Control 1, and Target vs. Control 2) also remained not significant (t(10) < 1.318, p > .217, d < .397). The Target was chosen 61.3% of the time (SD = .28.5) compared to Control 1, and 62.1% of the time (SD = 37.4) compared to Control 2. The Controls were not significantly different even after exclusions (M = 50.3, SD = 7.590, t(10) = .142, p > .890, d = .043).

The variation in choice remained in the critical conditions even after exclusions were applied. However, the marginally significant correlation in the full sample between the Target vs. Control 1 condition and the Target vs. Control 2 condition was not significant after exclusions (r = .195, p = .565), suggesting the variation was not consistent between participants.

These results validate the results from Chapter 2 in that although the neuroreinforcement procedure did not instill preference bias (or avoidance) of the Target, it had an effect of changing the relative status of the Target category compared to the control categories. This variability does not however seem to fit along a model-free to model-based dimension. Note that the variability between participants stayed intact even after exclusions (but the consistency did not), hence it is unlikely that the variability was due to lack of conditioning.

Even so, after exclusions, participants did seem to pick the Target more often, at least numerically. If that were the case, then we should also expect that the correlation between the two critical trials should have been significant, which it was not. Of course, one reason for this is because the sample size is small. It may have required 42 participants to find the effect at 80% power, after excluding non-conditioned participants.

Next, the induction likelihood was investigated (Figure 3.5). The average induction likelihood did not increase significantly across days (**Means:** *Day 2* = 57.9; *Day 3* = 53.8; *Day 4* = 59.2), ($F_{(2,45)}$ = .548, p = .582, η^2 = .024). This was also confirmed in the planned contrasts between days (ts(45) < 1.011, p > .317). The induction likelihood was significantly above 50% on Day 4 (t(15) = 3.166, p = .006, d = .792), marginally significant on Day 2 (t(15) = 1.803, p = .092, d = .451), and not significant on Day 3 (t(15) = .947, p = .359, d = .237), suggesting that participants learned to activate the target pattern, just that they did not consistently improve across days. These results were similar in the sample that excluded the participants who did not demonstrate proper conditioning (**Means:** *Day 2* = 55.5; *Day 3* = 52.0; *Day 4* = 59.9). None of the participants reported that they had the Target in mind during induction (Supplementary Table 3.2). Participants were also not above chance when guessing which of the three stimuli was the Target (i.e., Target, Control 1, Control 2).

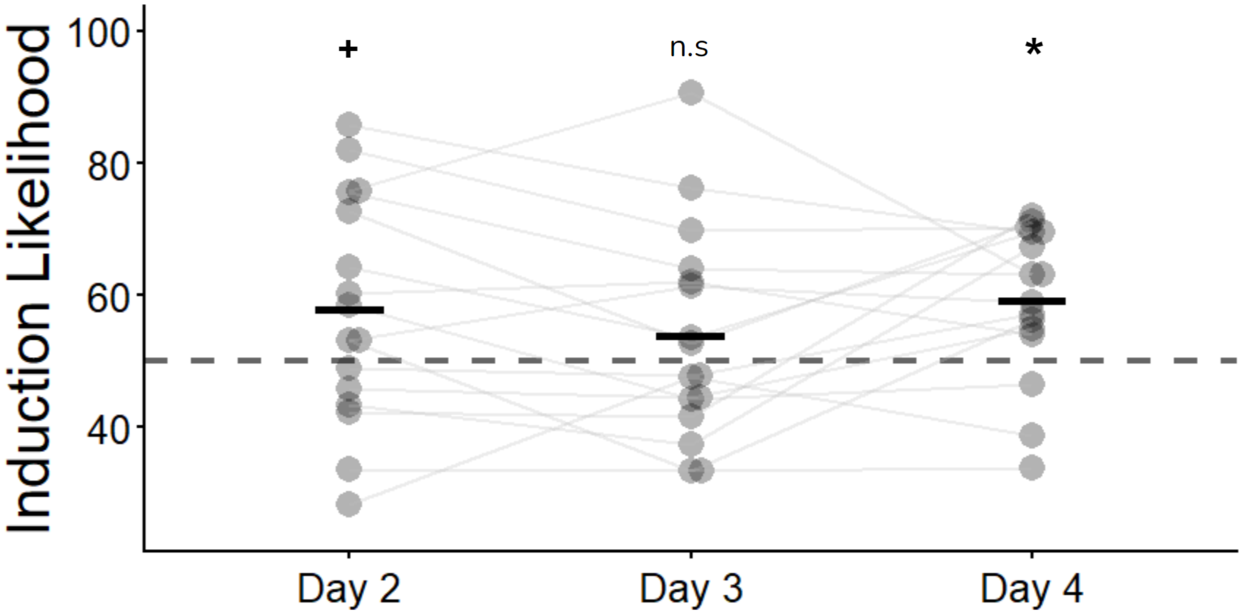

Figure 3.5 **Induction likelihood results.** Each dot represents one participant. The lines represent mean induction likelihood per day. * p < .05. **+** marginally significant (p < .10). n.s, not significant.

Next, the relationship between the induction likelihood and choice was investigated (Figure 3.6). Previous decoded neurofeedback studies have found a weak but positive relationship between absolute induction likelihood (calculated as the total score across all days of training) and behavior change (Cortese et al. 2021). This correlation did not materialize in the present experiment for any of the days individually (r < .188, p > .486). Neither did the average induction score (r = .086, p = .750), nor the difference between Day 4 and Day 2 predict choice (r = -.287, p = .281). Instead, a number of participants seem to have been non-responders which is common in neurofeedback studies (Sitaram et al. 2017; Oblak, Sulzer, and Lewis-Peacock 2019). Future studies should investigate how to identify likely non-responders prior to neurofeedback training.

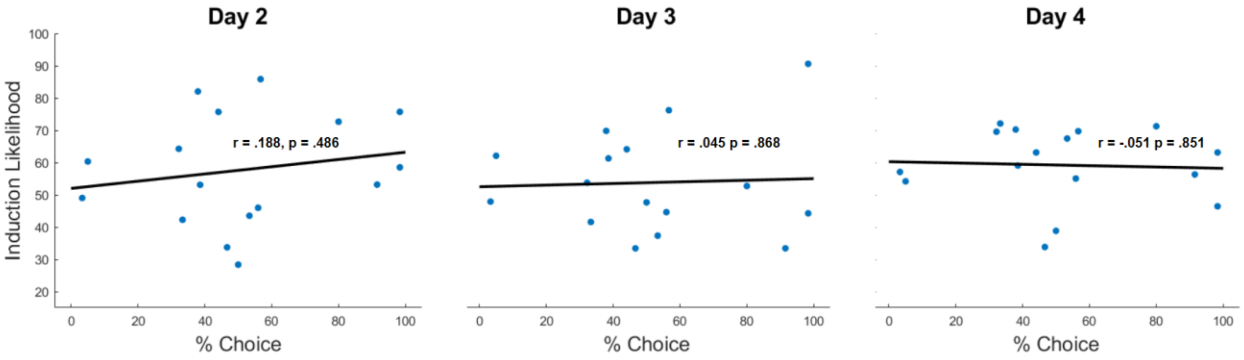

Figure 3.6 **Induction likelihood and choice correlation results.** Each dot represents one participant. The lines represent the fitted correlation line between induction likelihood and choice.

It can therefore be concluded that the neuroreinforcement procedure did not affect choice in a manner that is consistent with model-free or model-based learning in the choice task used here. Instead, neuroreinforcement seems to have affected the relative status of the Target as seen in the larger degree of variability in choice in the condition with the target than in the Controls condition.

Taken together, these results cannot conclude whether unconscious learning can support model-based inference. Specifically, the premise was that activating a multivoxel pattern in conjunction with another event, either reward or a consciously presented dot motion stimuli (which itself is later paired with loss), changes the evaluation of the pattern in a way that depends on those events. Instead, it is unclear what was learnt, if anything. The variability in choice is perhaps the most noteworthy result but one can only speculate about the cause.

Chapter 5

Conclusion

This dissertation thesis investigated conscious and unconscious learning using higher-order conditioning procedures. In Chapter 2, real-time fMRI neuroreinforcement was used to attempt to causally endow multivoxel patterns with value in a double-blind study in order to establish unconscious learning. Then, Chapter 3 showed in three experiments that the associative structures of the two higher-order conditioning procedures - sensory preconditioning and second-order conditioning - differ, and that they access different computations for value generalization. Lastly, Chapter 4 attempted to further establish whether unconscious learning can support these computations in a sensory preconditioning-inspired task using real-time fMRI neuroreinforcement.

Associative learning is a fundamental component of human decision making, but the relationship between learning and consciousness remains controversial. Chapter 2 and 4 of the dissertation proposed using a new neurofeedback technique to evaluate whether associative learning can be unconsciously incepted in humans. This neurofeedback technique, referred to as neuroreinforcement, is the closest approximation to non-invasive causal brain manipulation in humans, with high degrees of specificity (Shibata et al. 2019), and can be administered without participants' awareness, and without having to stimulate the brain.

The aim of Chapter 2 was to reinforce a visual category representation (i.e., the Target) with the expectations that the Target should be chosen more often over a neutral control category. This work would uncover causal links between brain representations and associated behavioral outcomes and thereby provide insight regarding the mechanisms of learning (Martin and Morris

2002). In brief, Chapter 2 did not reveal whether the target category was more valuable than the neutral controls. Instead, the results seem to suggest that the neuroreinforcement procedure caused participants to become more variable in their choice.

While speculative, it seems that unconscious learning took place, but that learning did not express itself as an increased tendency to choose the Target category. Strikingly, the same choice pattern observed in Chapter 2 emerged in Chapter 4. Chapter 4 aimed to explore two competing outcomes, namely whether unconscious learning can be model-free or more model-based. Following the discussion in Chapter 2, it therefore may have been expected that model-based inferences should emerge in Chapter 4. This is because previous neuroreinforcement procedures have shown that injecting value can increase the tendency to utilize generalization processes akin to model-based inference (Cortese et al. 2021). The results from this experiment were inconclusive and instead replicated the choice variability effect seen in Chapter 2, importantly, this variability in choice was not consistent across participants. Hence, neither study can definitively conclude whether unconscious learning is possible, or which individual differences drive the computations that solve the event structure of learning tasks – at least as measured in the choice task used here.

Another difference between the Chapter 2 and 4 was the use of punishment in the conditioning phase of sensory preconditioning. This manipulation is done not only to arbitrate between model-free and model-based processing but also to eliminate contributions due to the mere exposure effect. This is particularly pertinent in Chapter 2. That is, preference for the Target may have emerged due to evoking it, and not necessarily because of the reward associated with it. Hence, if this were the case then that could be a fatal flaw of Chapter 2, and to some extent also for Chapter 4.

While this may be a valid concern, it is only so if one believes that the mere exposure effect by itself is responsible for neuroreinforcement. Indeed, the biggest contribution to neuroreinforcement is thought to be through an ‘unconscious exposure’ effect (Chiba et al. 2019). Chiba et al. (2019) separated the contributions of reward and mere exposure by assuming that exposure is linearly proportional to the numbers of trials in which the target activity pattern was successfully induced above chance, and reward learning was assumed to be proportional to the induction likelihood of the target pattern multiplied by the amount of reward on each trial. This study found that while exposure seemed to be the main driver of the effect, the contribution of reward was still significant. Therefore, it can be assumed that activating the target pattern should also lead to it acquiring value, even if reward is not necessarily the main driving force behind the effect. There is also reason to believe that reward should be more effective than exposure, since that seems to be the case outside of neurofeedback contexts (Keller, Hennings, and Dunsmoor 2020). As Chiba et al., (2019) point out, future studies should apply these models in a larger sample size to better understand the contribution of each mechanism.

An interesting question is what role value plays in learning. This is relevant to the neuroreinforcement studies because they suggested that *some* learning took place, just not in the direction that was expected. It seems that despite the aim of the study to instill value in the Target, this was not what was learnt. Perhaps not even weakly so. Because if value was learnt, but just not strongly enough, then both critical conditions overall should have looked the same. That is, participants’ overall tendency to choose one stimulus over another should have been consistent between the two critical conditions. This is especially evident since the control conditions were very consistent between participants. Instead, it seems like value was not learned in the sense that was hypothesized.

Of course, this is all speculative with the caveat that the experiments had relatively small sample sizes and because the induction likelihood did not relate to choice. Future studies could employ a value localizer session to investigate this (Knutson et al. 2005). For example, if value played a role, then perhaps that signal should have been reactivated in the Test phase. Similarly, if value helped model-based inference for sensory preconditioning then this signal (i.e., a negative value signal since the dots were punished) should have reactivated in the Test phase when the preconditioned cue was present (supporting model-based inference). As is, the design of the present experiments was unable to address these questions.

Following the theme of whether value played a role, many popular computational models of learning appeal to Thorndike's "Law of Effect ", which holds that action that is followed by a reward is more likely to be repeated in the future (Thorndike 1898). These ideas were formalized within modern computational reinforcement frameworks in which actions are selected based on cached values from previous experience (Sutton and Barto 2018; Daw, Niv, and Dayan 2005). One recent proposal suggested that these ideas do not align with the view that behavior can be due to "value-less" stimulus-response learning, operating on a repetition-based mechanism that is strengthened by reward (Miller, Shenhav, and Ludvig 2019).

According to the value-less framework, reward simply influences the strength of the stimulus-response association and not whether the learning is established in the first place. Hence, if value actually played no role in neuroreinforcement, then if the predictions of this framework is correct, value correlates may not necessarily be expected in the brain. And if they are found, those value-correlates would not necessarily be the mechanisms by which learning takes place, but simply serve to strengthen the learning. Indeed, pure model-free signals can be difficult to identify and manipulate in the brain (Hayden and Niv 2021). Of course, the assumption of the neuroreinforcement experiment was that injecting a representation with value (broadly defined) in the brain will influence choice. It seems based on the results of the two studies that just because the brain can compute value (as suggested by an increase in induction likelihood), that does not *necessarily* mean that it will use it for the process of guiding choice.

The neuroreinforcement experiments used animate and inanimate categories as stimuli. These stimuli were chosen because they were used in previous neuroreinforcement experiments and showed good decoding results. As discussed in Chapter 2, one concern with using naturalistic stimuli is that pre-existing preferences may have influenced the results, which is why only the categories rated “Neutral” were chosen. Although the results do not suggest such an influence, the benefit of using categories did not extend to Chapter 3, hence they were abandoned for abstract symbols and fractals in those experiments.

Three experiments were conducted in Chapter 3 to investigate the associative structure of higher-order conditioning and the role of value. These experiments used extinction, devaluation and conditioned reinforcement, respectively. Specifically, Experiment 1 and 2 asked if second-order conditioning and sensory preconditioning shared their associative structures.

Two accounts were compared. The direct-link account argues that the second-order cue becomes directly associated with the outcome during second-order conditioning in a manner that is not affected by extinction and devaluation. According to the chained-link account the second-order cue should be affected by extinction and devaluation because on this view, the second-order cue evokes a representation of the first-order cue which in turn evokes a representation of the outcome. Sensory preconditioning was used as a control because the structure is thought to be a chained-link and not a direct-link.

Experiment 1 and 2 found the associative structure of second-order conditioning is a direct-link, while sensory preconditioning (as expected) is a chained-link. Lastly, Experiment 3 tested an analogous and critical question related to Experiment 1 and 2. Namely, if the second-order cue is directly linked to the outcome, then does the second-order cue become valuable? And, if the preconditioned cue is not directly linked to the outcome, then does the preconditioned cue not become valuable? To answer these questions, Experiment 3 used conditioned reinforcement – a gold-standard procedure for accessing model-free value. The results showed that second-order cues can support conditioned reinforcement, while preconditioned cues do not.

The results of Chapter 3 were surprising because it was initially hypothesized that using simultaneous presentations should facilitate chained-link representations (or mediated learning). This is because if two stimuli are presented simultaneously, those stimuli may be represented as a compound, hence any subsequent manipulation should affect both. The results of Chapter 3 helped inform the hypothesis about the associative structure in the sensory preconditioning-like study in Chapter 4. That is, in Chapter 4, a dot motion was presented on the screen at the same time as participants were inducing the target pattern in order to create an association between them.

However, one may wonder whether the order in the association phase of the experiment in Chapter 4 was nevertheless reversed from classic sensory preconditioning tasks. That is, the order of pairing for classic preconditioning (i.e., S2 -> S1, S1 -> -$; i.e., dots -> Target, Target -> loss) implies that S2 ultimately predicts the outcome in a forward temporal direction (i.e., S2 -> S1 -> -$). It is possible that the order in Chapter 4 instead has the temporal order of S1 -> S2, S1 -> -$ (i.e., dots -> Target, dots -> loss). This order may fail to produce the model-based transitive implication even if both conditioning stages succeeded since the merged timelines of the two phases may imply temporally backward ordering from the Target to the outcome (i.e., -

$). This learning can be weak or even inhibitory (Prével and Krebs 2021; Seitz et al. 2022). Hence, any learning in Chapter 4 would be qualitatively different from that studied in Chapter 3 which would have made it difficult to answer how conscious learning differs from unconscious learning.

There are reasons to believe this may not be the case. This is because the dots co-terminated with the termination of the Induction phase. Furthermore, participants were told to rest during the ITI period (i.e., “Rest” appeared on the screen), and the induction feedback was calculated against this period as the baseline for every trial. Therefore, if participants were inducing the Target before or after actual induction, then feedback would have been small. Hence, if participants avoided the Target at test, it is likely that they did so due to a more model-based inference.

Although unconscious learning was not demonstrated in the neuroreinforcement experiments, there are examples of unaware learning using subliminal masking technique. Of these, two stand out. The first one used instrumental conditioning, however as discussed in the Background, this study has failed to replicate. The other used a perceptual learning task and seems more robust. Seitz et al. (2009) found that reward-based feedback on correct trials drove participants’ ability to learn the orientation of a grating stimulus that was masked by continuous flash suppression. Importantly, a recognition test indicated that participants were not better than chance at guessing the presence of the stimulus. This effect is dependent on sleep (Tamaki et al. 2020).

This, together with neuropsychological evidence such as unconscious conditioning demonstrated in patients and animal models of blindsight, suggests that strong unconscious learning should be possible with the right methodology.

This dissertation thesis used an especially promising technology, neuroreinforcement, coupled with online behavioral experiments to study how learning is established, the computations that support it, and the role of consciousness. This approach attempted to put the experimenter "in the driver seat" of the brain to gain a better understanding of learning and memory and ultimately, of mental health conditions (Martin and Morris 2002; Kahnt 2022).

Appendices

Chapter 2 Appendix

Supplementary Table 1.1

Participant induction strategies

Participant	**Strategies**
1	Thinking about Euphoria on HBO. Thinking about a guy with a gun. Thinking about random big words and spelling them.
2	Saying the alphabet. Singing in her head. Counting backwards.
3	Thinking about things that brought strong emotion such as skydiving.
4	Thinking about scenarios that might kill her. Thinking about singing songs from SpongeBob and Ted.
5	Singing songs from Encanto, Trolls of Prince of Egypt. Imagined modern music cover groups.

Chapter 3 Appendix

Supplementary Table 2.1

Exclusions for Experiment 1

	5+ trials missed	*Button press repetitions*	*Early dropout*	*Failed extinction*	*Technical issue*
SOC	2	0	2	64 (*39%*)	3
SPC	8	3	6	51 (*34%*)	0

Supplementary Table 2.2

Design for second-order conditioning in Experiment 1

	Conditioning	Second-order	Extinction	Test
Second-order conditioning	S1+	S1+ – S2+		S1+ vs. S1Ext S2+ vs. S2Ext S2+ vs. Neutral (2) S2Ext vs. Neutral (2)
Second-order extinction	S1Ext	S1Ext – S2Ext	S1Ext	
Neutral Controls	Neutral (1)	Neutral (1) – Neutral (2)		
Fillers	Neutral Filler		R+ (rewarded) Neutral Filler	

Supplementary Table 2.3

Design for sensory preconditioning in Experiment 1

	Preconditioning	**Conditioning**	**Extinction**	**Test**
Preconditioning	S1+ – S2+	S1+		S1+ vs. S1Ext S2+ vs. S2Ext S2+ vs. Neutral (2) S2Ext vs. Neutral (2)
Preconditioning Extinction	S1Ext – S2Ext	S1Ext	S1Ext	
Neutral Controls	Neutral (1) – Neutral (2)	Neutral (1)		
Fillers		Neutral Filler	R+ (rewarded) Neutral Filler	

Supplementary Table 2.4

Exclusion for Experiment 2

	5+ trials missed	*Button press repetitions*	*Failed devaluation check*	*Technical issue*
SOC	1	0	20 (*29%*)	1
SPC	0	0	9 (*14%*)	0

Supplementary Table 2.5

Design for second-order conditioning in Experiment 2

	Conditioning	**Second-order**	**Devaluation**	**Test**
Second-order Conditioning	S1+	S1+ – S2+	S1Deval associated outcome devalued	S1+ vs. S1Deval S2+ vs. S2Deval S2+ vs. Neutral (2) S2Deval vs. Neutral (2)
Second-order Devalued	S1Deval	S1Deval–S2Deval		
Neutral Controls	Neutral (1)	Neutral (1) – Neutral (2)		
Fillers	Neutral Filler			

Supplementary Table 2.6

Design for sensory preconditioning in Experiment 2

	Preconditioning	Conditioning	Devaluation	Test
Preconditioning	S1+ – S2+	S1+	S1Deval associated outcome devalued	S1+ vs. S1Deval S2+ vs. S2Deval S2+ vs. Neutral (2) S2Deval vs. Neutral (2)
Preconditioning Devalued	S1Deval– S2Deval	S1Deval		
Neutral Controls	Neutral (1) – Neutral (2)	Neutral (1)		
Fillers		Neutral Filler		

Supplementary Table 2.7

Exclusions for Experiment 3

	5+ trials missed	*Button press repetitions*	*Early dropout*	*Technical issue*
SOC	7	3	2	1
SPC	3	4	4	0

Supplementary Table 2.8

Design for second-order conditioning in Experiment 3

	Conditioning	Second-order Conditioning	Test	Conditioned Reinforcement Indirect Cues	Conditioned Reinforcement Direct Cues
Second-order Conditioning	S1+	S1+ – S2+	S1+ vs. Neutral (1) S2+ vs. Neutral (2)	S2+ Neutral (2)	S1+ Neutral (1)
Neutral Controls	Neutral (1)	Neutral (1) – Neutral (2)			
Fillers	Neutral Filler Reward Filler				

Supplementary Table 2.9

Design for sensory preconditioning in Experiment 3

	Preconditioning	Conditioning	Test	Conditioned Reinforcement Indirect Cues	Conditioned Reinforcement Direct Cues
Preconditioning	S1+ – S2+	S1+	S1+ vs. Neutral (1) S2+ vs. Neutral (2)	S2+ Neutral (2)	S1+ Neutral (1)
Neutral Controls	Neutral (1) – Neutral (2)	Neutral (1)			
Fillers		Neutral Filler Reward Filler			

Chapter 4 Appendix

Supplementary Table 3.1

Design for *Establishing whether unconscious learning can support model-based inference*

Conditioning	Test
S1+, S1-, Mildly Rewarded Category Control, Neutral Category Control	Target vs. Control 1 Target vs. Control 2 Control 1 vs. Control 2 2 S1+ vs. S1-

Supplementary Table 3.2

Participant induction strategies	
Participant	**Strategies**
1	Pretended to be hiding in a pile of bodies, attempting not to move.
2	Multiplication of numbers.
3	Tried to visualize times that made him feel good, such as playing with cute animals or being with friends and family.
4	Thought about emotions, imagined Harry Potter.
5	Thought about random foods.
6	Mentally painting the green circle.
7	Thinking about the circle in motion increasing and shrinking in size.
8	Tried to experience different emotions, tried math problems, activating different muscles.
9	Thinking about memories and complex topics.
10	Sang Rick Roll to himself in his head, and Beatles.
11	[Not recorded]
12	Thought about emotional memories.
13	Relax during the rest period then “intensely” stimulate his visual cortex.
14	Think about memories or imagine navigating through buildings.
15	Focused on the direction of the dots.
16	Count as many dots as possible, then try memorizing the dot pattern. [the dot pattern was random].

References

Adams, Rick, Quintin Huys, and Jon Roiser. 2016. “Computational Psychiatry: Towards a Mathematically Informed Understanding of Mental Illness.” *Journal of Neurology, Neurosurgery, and Psychiatry*, no. 87: 53–63.

Allman, Melissa J., Iser G. DeLeon, Michael F. Cataldo, Peter C. Holland, and Alexander W. Johnson. 2010. “Learning Processes Affecting Human Decision Making: An Assessment of Reinforcer-Selective Pavlovian-to-Instrumental Transfer Following Reinforcer Devaluation.” *Journal of Experimental Psychology. Animal Behavior Processes* 36 (3): 402–8.

Alvarado, M. C., and J. W. Rudy. 1992. “Some Properties of Configural Learning: An Investigation of the Transverse-Patterning Problem.” *Journal of Experimental Psychology. Animal Behavior Processes* 18 (2): 145–53.

Amano, Kaoru, Kazuhisa Shibata, Mitsuo Kawato, Yuka Sasaki, and Takeo Watanabe. 2016. “Learning to Associate Orientation with Color in Early Visual Areas by Associative Decoded fMRI Neurofeedback.” *Current Biology: CB* 26 (14): 1861–66.

Arzi, Anat, Limor Shedlesky, Mor Ben-Shaul, Khitam Nasser, Arie Oksenberg, Ilana S. Hairston, and Noam Sobel. 2012. “Humans Can Learn New Information during Sleep.” *Nature Neuroscience* 15 (10): 1460–65.

Balleine, B. W., and A. Dickinson. 1998. “Goal-Directed Instrumental Action: Contingency and Incentive Learning and Their Cortical Substrates.” *Neuropharmacology* 37 (4-5): 407–19.

Barron, Helen C., Hayley M. Reeve, Renée S. Koolschijn, Pavel V. Perestenko, Anna Shpektor, Hamed Nili, Roman Rothaermel, et al. 2020. “Neuronal Computation Underlying Inferential Reasoning in Humans and Mice.” *Cell* 183 (1): 228–43.e21.

Block, Ned. 2007. “Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience.” *The Behavioral and Brain Sciences* 30 (5-6): 481–99; discussion 499–548.

Block, Ned, David Carmel, Stephen M. Fleming, Robert W. Kentridge, Christof Koch, Victor A.

F. Lamme, Hakwan Lau, and David Rosenthal. 2014. “Consciousness Science: Real Progress and Lingering Misconceptions.” *Trends in Cognitive Sciences*.

Bouton, Mark E. 2021. “Context, Attention, and the Switch between Habit and Goal-Direction in Behavior.” *Learning & Behavior* 49 (4): 349–62.

Breitmeyer, Bruno G. 2015. “Psychophysical ‘Blinding’ Methods Reveal a Functional Hierarchy of Unconscious Visual Processing.” *Consciousness and Cognition* 35 (September): 234–50.

Brogden, B. Y. W. J. 1945. “Sensory Pre-Conditioning of Human Subjects*,” 527–39.

Brown, Richard, Hakwan Lau, and Joseph E. LeDoux. 2019. “Understanding the Higher-Order Approach to Consciousness.” *Trends in Cognitive Sciences* 23 (9): 754–68.

Byrom, Nicola C., and Robin A. Murphy. 2016. “Individual Difference in Configural Associative Learning.” *Journal of Experimental Psychology: Animal Learning and Cognition* 42 (4): 325–35.

Chiba, Toshinori, Tetsufumi Kanazawa, Ai Koizumi, Kentarou Ide, Vincent Taschereau-Dumouchel, Shuken Boku, Akitoyo Hishimoto, et al. 2019. “Current Status of Neurofeedback for Post-Traumatic Stress Disorder: A Systematic Review and the Possibility of Decoded Neurofeedback.” *Frontiers in Human Neuroscience* 13 (July): 1–13.

Cohen, Jonathan D., Nathaniel Daw, Barbara Engelhardt, Uri Hasson, Kai Li, Yael Niv, Kenneth A. Norman, et al. 2017. “Computational Approaches to fMRI Analysis.” *Nature Neuroscience* 20 (3): 304–13.

Collins, Anne G. E., and Jeffrey Cockburn. 2020. “Beyond Dichotomies in Reinforcement Learning.” *Nature Reviews. Neuroscience* 21 (10): 576–86.

Cortese, Aurelio, Asuka Yamamoto, Maryam Hashemzadeh, Pradyumna Sepulveda, Mitsuo Kawato, and Benedetto De Martino. 2021. “Value Signals Guide Abstraction during Learning.” *eLife* 10 (July). https://doi.org/10.7554/eLife.68943.

Costa, Kauê Machado, Robert Scholz, Kevin Lloyd, Perla Moreno-Castilla, Matthew P. H.

Gardner, Peter Dayan, and Geoffrey Schoenbaum. 2022. “The Role of the Orbitofrontal Cortex in Creating Cognitive Maps.” *bioRxiv*. https://doi.org/10.1101/2022.01.25.477716.

Craddock, Paul, Jessica S. Wasserman, Cody W. Polack, Thierry Kosinski, Charlotte Renaux, and Ralph R. Miller. 2018. “Associative Structure of Second-Order Conditioning in Humans.” *Learning & Behavior* 46 (2): 171–81.

Davey, G. C., and T. Arulampalam. 1982. “Second-Order ‘Fear’ Conditioning in Humans: Persistence of CR2 Following Extinction of CR1.” *Behaviour Research and Therapy* 20 (4): 391–96.

Davey, G. C., and I. McKenna. 1983. “The Effects of Postconditioning Revaluation of CS and UCS Following Pavlovian Second-Order Electrodermal Conditioning in Humans.” *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology* 35 (Pt 2) (May): 125–33.

Daw, Nathaniel D. 2018. “Are We of Two Minds?” *Nature Neuroscience*.

Daw, Nathaniel D., Yael Niv, and Peter Dayan. 2005. “Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control.” *Nature Neuroscience* 8 (12): 1704–11.

Daw, Nathaniel D., John P. O’Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. 2006. “Cortical Substrates for Exploratory Decisions in Humans.” *Nature* 441 (7095): 876–79.

Dayan, Peter, and Kent C. Berridge. 2014. “Model-Based and Model-Free Pavlovian Reward Learning: Revaluation, Revision, and Revelation.” *Cognitive, Affective & Behavioral Neuroscience* 14 (2): 473–92.

Dehaene, Stanislas, Hakwan Lau, and Sid Kouider. 2017. “What Is Consciousness, and Could Machines Have It?” *Science* 358 (6362): 486–92.

Dickinson, Anthony, and Bernard Balleine. 1994. “Motivational Control of Goal-Directed Action.” *Animal Learning & Behavior* 22 (1): 1–18.

Dolan, Ray J., and Peter Dayan. 2013. “Goals and Habits in the Brain.” *Neuron* 80 (2): 312–25.

Doll, Bradley B., and Nathaniel D. Daw. 2016. “The Expanding Role of Dopamine.” *eLife* 5. 3–5.

Doll, Bradley B., W. Jake Jacobs, Alan G. Sanfey, and Michael J. Frank. 2009. “Instructional Control of Reinforcement Learning: A Behavioral and Neurocomputational Investigation.” *Brain Research* 1299 (November): 74–94.

Doll, Bradley B., Dylan A. Simon, and Nathaniel D. Daw. 2012. “The Ubiquity of Model-Based Reinforcement Learning.” *Current Opinion in Neurobiology* 22 (6): 1075–81.

Doya, Kenji. 2008. “Modulators of Decision Making.” *Nature Neuroscience* 11 (4): 410–16.

Evans, Jonathan St B. T. 2008. “Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition.” *Annual Review of Psychology* 59: 255–78.

Fanselow, Michael S., and Zachary T. Pennington. 2018. “A Return to the Psychiatric Dark Ages with a Two-System Framework for Fear.” *Behaviour Research and Therapy* 100: 24–29.

Fischl, Bruce, Andre van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, et al. 2004. “Automatically Parcellating the Human Cerebral Cortex.” *Cerebral Cortex* 14 (1): 11–22.

Frith, Chris D. 2019. “The Neural Basis of Consciousness.” *Psychological Medicine* 1: 1–13.

Gardner, Matthew P. H., Geoffrey Schoenbaum, and Samuel J. Gershman. 2018. “Rethinking Dopamine as Generalized Prediction Error.” *Proceedings of the Royal Society B: Biological Sciences* 285 (1891). https://doi.org/10.1098/rspb.2018.1645.

Garvert, Mona M., Raymond J. Dolan, and Timothy E. J. Behrens. 2017. “A Map of Abstract Relational Knowledge in the Human Hippocampal–entorhinal Cortex.” *eLife* 6 (April): e17086.

Geerts, Jesse P., Fabian Chersi, Kimberly L. Stachenfeld, and Neil Burgess. 2020. “A General Model of Hippocampal and Dorsal Striatal Learning and Decision Making.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (49): 31427–37.

Gershman, Samuel J. 2018. “The Successor Representation: Its Computational Logic and Neural Substrates.” *Journal of Neuroscience* 38 (33): 7193–7200.

González, Valeria V., Benjamin M. Seitz, Rachel Formaker, and Aaron P. Blaisdell. 2022. “Elements of a Compound Elicit Little Conditioned Reinforcement.” *Journal of Experimental Psychology. Animal Learning and Cognition* 48 (1): 67–72.

Greenwald, A. G., M. R. Klinger, and E. S. Schuh. 1995. “Activation by Marginally Perceptible (‘subliminal’) Stimuli: Dissociation of Unconscious from Conscious Cognition.” *Journal of Experimental Psychology. General* 124 (1): 22–42.

Greenwald, Anthony G., and Jan De Houwer. 2017. “Unconscious Conditioning: Demonstration of Existence and Difference from Conscious Conditioning.” *Journal of Experimental Psychology. General* 146 (12): 1705–21.

Hamm, Alfons O., Almut I. Weike, Harald T. Schupp, Thomas Treig, Alexander Dressel, and Christof Kessler. 2003. “Affective Blindsight: Intact Fear Conditioning to a Visual Cue in a Cortically Blind Patient.” *Brain: A Journal of Neurology* 126 (Pt 2): 267–75.

Hanke, Michael, Yaroslav O. Halchenko, Per B. Sederberg, Emanuele Olivetti, Ingo Fründ, Jochem W. Rieger, Christoph S. Herrmann, James V. Haxby, Stephen José Hanson, and Stefan Pollmann. 2009. “PyMVPA: A Unifying Approach to the Analysis of Neuroscientific Data.” *Frontiers in Neuroinformatics* 3 (February): 3.

Hart, Evan E., Melissa J. Sharpe, Matthew Ph Gardner, and Geoffrey Schoenbaum. 2020. “Responding to Preconditioned Cues Is Devaluation Sensitive and Requires Orbitofrontal Cortex during Cue-Cue Learning.” *eLife* 9 (August). https://doi.org/10.7554/eLife.59998.

Hassin, Ran R. 2013. “Yes It Can: On the Functional Abilities of the Human Unconscious.” *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 8 (2): 195–207.

Haugg, Amelie, Ronald Sladky, Stavros Skouras, Amalia McDonald, Cameron Craddock, Matthias Kirschner, Marcus Herdener, et al. 2020. “Can We Predict Real-Time fMRI

Neurofeedback Learning Success from Pre-Training Brain Activity?" *bioRxiv*, 2020.01.15.906388.

Haxby, James V., M. Ida Gobbini, Maura L. Furey, Alumit Ishai, Jennifer L. Schouten, and Pietro Pietrini. 2001. "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Corten." *Science* 293 (September): 87–96.

Hayden, Benjamin Y., and Yael Niv. 2021. "The Case against Economic Values in the Orbitofrontal Cortex (or Anywhere Else in the Brain)." *Behavioral Neuroscience* 135 (2): 192–201.

Henke, Katharina. 2010. "A Model for Memory Systems Based on Processing Modes rather than Consciousness." *Nature Reviews. Neuroscience* 11 (7): 523–32.

Holland, Peter. C. 1977. "Conditioned Stimulus as a Determinant of the Form of the Pavlovian Conditioned Response." *Journal of Experimental Psychology. Animal Behavior Processes* 3 (1): 77–104.

Holland, Peter. C 1990. "Event Representation in Pavlovian Conditioning: Image and Action." *Cognition* 37 (1-2): 105–31.

Holland, Peter C. 1981. "Acquisition of Representation-Mediated Conditioned Food Aversions." *Learning and Motivation* 12 (1): 1–18.

Holland, Peter C., and Deswell T. Forbes. 1982. "Representation-Mediated Extinction of Conditioned Flavor Aversions." *Learning and Motivation*. https://doi.org/10.1016/0023-9690(82)90004-2.

Holland, Peter C., and Robert A. Rescorla. 1975. "Second-Order Conditioning with Food Unconditioned Stimulus." *Journal of Comparative and Physiological Psychology* 88 (1): 459–67.

Holmes, Nathan M., Francesca S. Wong, Youcef Bouchekioua, and R. Fred Westbrook. 2021. "Not 'Either-or' but 'Which-When': A Review of the Evidence for Integration in Sensory Preconditioning." *Neuroscience and Biobehavioral Reviews*, October.

https://doi.org/10.1016/j.neubiorev.2021.10.032.

Honey, Robert C., and Dominic M. Dwyer. 2021. “Higher-Order Conditioning: What Is Learnt and How It Is Expressed.” *Frontiers in Behavioral Neuroscience* 15 (September): 726218.

Honey, Robert C., Dominic M. Dwyer, and Adela F. Iliescu. 2020. “Individual Variation in the Vigor and Form of Pavlovian Conditioned Responses: Analysis of a Model System.” *Learning and Motivation* 72 (November): 101658.

Howard, James D., Rachel Reynolds, Devyn E. Smith, Joel L. Voss, Geoffrey Schoenbaum, and Thorsten Kahnt. 2020. “Targeted Stimulation of Human Orbitofrontal Networks Disrupts Outcome-Guided Behavior.” *Current Biology: CB* 30 (3): 490–98.e4.

Kahnt, Thorsten. 2022. “Computationally Informed Interventions for Targeting Compulsive Behaviors.” *Biological Psychiatry*, September. https://doi.org/10.1016/j.biopsych.2022.08.028.

Kahnt, Thorsten, and Geoffrey Schoenbaum. 2021. “Cross-Species Studies on Orbitofrontal Control of Inference-Based Behavior.” *Behavioral Neuroscience* 135 (2): 109–19.

Kato, Rikako, Abdelhafid Zeghbib, Peter Redgrave, and Tadashi Isa. 2021. “Visual Instrumental Learning in Blindsight Monkeys.” *Scientific Reports* 11 (1): 14819.

Keller, Nicole E., Augustin C. Hennings, and Joseph E. Dunsmoor. 2020. “Behavioral and Neural Processes in Counterconditioning: Past and Future Directions.” *Behaviour Research and Therapy* 125: 103532.

Klein, Tilmann A., Markus Ullsperger, and Gerhard Jocham. 2017. “Learning Relative Values in the Striatum Induces Violations of Normative Decision Making.” *Nature Communications* 8 (June): 16033.

Knotts, Jeffrey. D., Hakwan Lau, and Megan A. K. Peters. 2018. “Continuous Flash Suppression and Monocular Pattern Masking Impact Subjective Awareness Similarly.” *Attention, Perception & Psychophysics* 80 (8): 1974–87.

Knutson, Brian, Jonathan Taylor, Matthew Kaufman, Richard Peterson, and Gary Glover. 2005.

“Distributed Neural Representation of Expected Value.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25 (19): 4806–12.

Kouider, Sid, and Stanislas Dehaene. 2007. “Levels of Processing during Non-Conscious Perception: A Critical Review of Visual Masking.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 362 (1481): 857–75.

Krajbich, Ian, Björn Bartling, Todd Hare, and Ernst Fehr. 2015. “Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference.” *Nature Communications* 6 (July): 7455.

Krishnapuram, Balaji, Lawrence Carin, Mário A. T. Figueiredo, and Alexander J. Hartemink. 2005. “Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (6): 957–68.

Kurth-Nelson, Zeb, Gareth Barnes, Dino Sejdinovic, Ray Dolan, and Peter Dayan. 2015. “Temporal Structure in Associative Retrieval.” *eLife* 4: 1–18.

Lashley, Karl. 1923. “The Behavioristic Interpretation of Consciousness II.” *Psychological Review*, no. 30: 329–53.

Lau, Hakwan C. 2008. “Are We Studying Consciousness Yet?” Edited by L. Weiskrantz and M. Davies. https://doi.org/10.1093/ acprof:oso/9780199233151.001.0001.

Lau, Hakwan, and David Rosenthal. 2011. “Empirical Support for Higher-Order Theories of Conscious Awareness.” *Trends in Cognitive Sciences* 15 (8): 365–73.

LeDoux, Joseph, and Nathaniel D. Daw. 2018. “Surviving Threats: Neural Circuit and Computational Implications of a New Taxonomy of Defensive Behaviour.” *Nature Reviews. Neuroscience* 19 (5): 269–82.

LeDoux, Joseph E., and Hakwan Lau. 2020. “Seeing Consciousness through the Lens of Memory.” *Current Biology: CB* 30 (18): R1018–22.

LeDoux, Joseph E., Matthias Michel, and Hakwan Lau. 2020. “A Little History Goes a Long Way toward Understanding Why We Study Consciousness the Way We Do Today.”

Proceedings of the National Academy of Sciences of the United States of America 117 (13): 6976–84.

LeDoux, Joseph E., and Daniel S. Pine. 2016. “Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework.” *The American Journal of Psychiatry* 173 (11): 1083–93.

Lee, Jessica C. 2021. “Second-Order Conditioning in Humans.” *Frontiers in Behavioral Neuroscience* 15 (July): 672628.

Leeuw, Joshua R. de. 2015. “jsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser.” *Behavior Research Methods* 47 (1): 1–12.

Lovibond, Peter F., and David R. Shanks. 2002. “The Role of Awareness in Pavlovian Conditioning: Empirical Evidence and Theoretical Implications.” *Journal of Experimental Psychology. Animal Behavior Processes* 28 (1): 3–26.

Luettgau, Lennart, Emanuele Porcu, Claus Tempelmann, and Gerhard Jocham. 2021. “Reinstatement of Cortical Outcome Representations during Higher-Order Learning.” *Cerebral Cortex* 32 (1): 93–109.

Martin, S. J., and R. G. M. Morris. 2002. “New Life in an Old Idea: The Synaptic Plasticity and Memory Hypothesis Revisited.” *Hippocampus* 12 (5): 609–36.

Mattar, Marcelo G., and Nathaniel D. Daw. 2018. “Prioritized Memory Access Explains Planning and Hippocampal Replay.” *Nature Neuroscience* 21 (11): 1609–17.

McDougle, Samuel D., Ian C. Ballard, Beth Baribault, Sonia J. Bishop, and Anne G. E. Collins. 2021. “Executive Function Assigns Value to Novel Goal-Congruent Outcomes.” *Cerebral Cortex* 32 (1): 231–47.

Mertens, Gaëtan, and Iris M. Engelhard. 2019. “A Systematic Review and Meta-Analysis of the Evidence for Unaware Fear Conditioning.” *Neuroscience and Biobehavioral Reviews* 108 (April 2019): 254–68.

Michel, Matthias, Diane Beck, Ned Block, Hal Blumenfeld, Richard Brown, David Carmel, Marisa Carrasco, et al. 2019. “Opportunities and Challenges for a Maturing Science of Consciousness.” *Nature Human Behaviour* 3 (2): 104–7.

Miller, Kevin J., Amitai Shenhav, and Elliot A. Ludvig. 2019. “Habits without Values.” *Psychological Review* 126 (2): 292–311.

Mitchell, Chris, Jan De Houwer, and Lovibond. 2009. “The Propositional Nature of Human Associative Learning.” *The Behavioral and Brain Sciences* 32 (2): 183–246.

Mitchell, Chris J., Evan Livesey, and Peter F. Lovibond. 2007. “A Dissociation between Causal Judgement and the Ease with Which a Cause Is Categorized with Its Effect.” *Quarterly Journal of Experimental Psychology* 60 (3): 400–417.

Momennejad, Ida. 2020. “Learning Structures: Predictive Representations, Replay, and Generalization.” *Current Opinion in Behavioral Sciences* 32: 155–66.

Momennejad, I., E. M. Russek, J. H. Cheong, M. M. Botvinick, N. D. Daw, and S. J. Gershman. 2017. “The Successor Representation in Human Reinforcement Learning.” *Nature Human Behaviour* 1 (9): 680–92.

Muñiz-Diez, Clara, Judit Muñiz-Moreno, and Ignacio Loy. 2021. “Second-Order Conditioning and Conditioned Inhibition in Different Moments of the Same Training: The Effect of A+ and AX- Trial Number.” *Frontiers in Behavioral Neuroscience* 15 (April): 632548.

Muñoz-Moldes, Santiago, and Axel Cleeremans. 2020. “Delineating Implicit and Explicit Processes in Neurofeedback Learning.” *Neuroscience and Biobehavioral Reviews* 118 (November): 681–88.

Newell, Ben R., and David R. Shanks. 2014. “Unconscious Influences on Decision Making: A Critical Review.” *The Behavioral and Brain Sciences* 37 (1): 1–19.

Niv, Yael. 2019. “Learning Task-State Representations.” *Nature Neuroscience* 22 (October). https://doi.org/10.1038/s41593-019-0470-8.

Oblak, Ethan F., Jarrod A. Lewis-Peacock, and James S. Sulzer. 2017. “Self-Regulation Strategy, Feedback Timing and Hemodynamic Properties Modulate Learning in a

Simulated fMRI Neurofeedback Environment." *PLoS Computational Biology* 13 (7): 1–25.

Oblak, Ethan F., James S. Sulzer, and Jarrod A. Lewis-Peacock. 2019. "A Simulation-Based Approach to Improve Decoded Neurofeedback Performance." *NeuroImage* 195 (March): 300–310.

O'Doherty, John, Peter Dayan, Johannes Schultz, 1. Ralf Deichmann, Karl Friston, and Raymond J. Dolan. 2004. "Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning." *Science* 304 (5669). https://doi.org/10.1126/science.1094285.

O'Doherty, John P., Sang Wan Lee, Reza Tadayonnejad, Jeff Cockburn, Kyo Iigaya, and Caroline J. Charpentier. 2021. "Why and How the Brain Weights Contributions from a Mixture of Experts." *Neuroscience and Biobehavioral Reviews* 123 (April): 14–23.

Parkinson, J. A., A. C. Roberts, B. J. Everitt, and P. Di Ciano. 2005. "Acquisition of Instrumental Conditioned Reinforcement Is Resistant to the Devaluation of the Unconditioned Stimulus." *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology* 58 (1): 19–30.

Pedregosa, Fabian. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research: JMLR*.

Persaud, Navindra, Matthew Davidson, Brian Maniscalco, Dean Mobbs, Richard E. Passingham, Alan Cowey, and Hakwan Lau. 2011. "Awareness-Related Activity in Prefrontal and Parietal Cortices in Blindsight Reflects More than Superior Visual Performance." *NeuroImage* 58 (2): 605–11.

Pessiglione, Mathias, Predrag Petrovic, Jean Daunizeau, Stefano Palminteri, Raymond J. Dolan, and Chris D. Frith. 2008. "Subliminal Instrumental Conditioning Demonstrated in the Human Brain." *Neuron* 59 (4): 561–67.

Peters, Megan A. K., and Hakwan Lau. 2015. "Human Observers Have Optimal Introspective Access to Perceptual Processes Even for Visually Masked Stimuli." *eLife* 4 (October): e09651.

Pool, Eva R., Wolfgang M. Pauli, Carolina S. Kress, and John P. O'Doherty. 2019. "Behavioural Evidence for Parallel Outcome-Sensitive and Outcome-Insensitive Pavlovian Learning Systems in Humans." *Nature Human Behaviour* 3 (3): 284–96.

Prével, Arthur, and Ruth M. Krebs. 2021. "Higher-Order Conditioning With Simultaneous and Backward Conditioned Stimulus: Implications for Models of Pavlovian Conditioning." *Frontiers in Behavioral Neuroscience* 15 (November): 749517.

Raio, Candace M., David Carmel, Marisa Carrasco, and Elizabeth A. Phelps. 2012. "Nonconscious Fear Is Quickly Acquired but Swiftly Forgotten." *Current Biology: CB* 22 (12): R477–79.

Rescorla, R. A. 1980. "Simultaneous and Successive Associations in Sensory Preconditioning." *Journal of Experimental Psychology: Animal Behavior Processes* 6 (3): 207–16.

Rescorla, R. A. 1982. "Simultaneous Second-Order Conditioning Produces S-S Learning in Conditioned Suppression." *Journal of Experimental Psychology. Animal Behavior Processes* 8 (1): 23–32.

Rescorla, Robert. 1981. "Simultaneous Associations." In *Predictability, Correlation, and Contiguity*, edited by P. Harzem and M. D. Zeiler, 47–80. John Wiley & Sons Ltd.

Rizley, Ross C., and Robert A. Rescorla. 1972. "Associations in Second-Order Conditioning and Sensory Preconditioning." *Journal of Comparative and Physiological Psychology* 81 (1): 1–11.

Scott, Ryan B., Jason Samaha, Ron Chrisley, and Zoltan Dienes. 2018. "Prevailing Theories of Consciousness Are Challenged by Novel Cross-Modal Associations Acquired between Subliminal Stimuli." *Cognition* 175 (June): 169–85.

Seitz, Aaron R., Dongho Kim, and Takeo Watanabe. 2009. "Rewards Evoke Learning of Unconsciously Processed Visual Stimuli in Adult Humans." *Neuron* 61 (5): 700–707.

Seitz, Benjamin M., Aaron P. Blaisdell, and Melissa J. Sharpe. 2021. "Higher-Order Conditioning and Dopamine: Charting a Path Forward." *Frontiers in Behavioral*

Neuroscience 15: 228.

Seitz, Benjamin M., Ivy B. Hoang, Lauren E. DiFazio, Aaron P. Blaisdell, and Melissa J. Sharpe. 2022. “Dopamine Errors Drive Excitatory and Inhibitory Components of Backward Conditioning in an Outcome-Specific Manner.” *Current Biology: CB* 32 (14): 3210–18.e3.

Shanks, David R. 2017. “Regressive Research: The Pitfalls of Post Hoc Data Selection in the Study of Unconscious Mental Processes.” *Psychonomic Bulletin & Review* 24 (3): 752–75.

Sharpe, Melissa J., Hannah M. Batchelor, and Geoffrey Schoenbaum. 2017. “Preconditioned Cues Have No Value.” *eLife* 6: 1–9.

Sharpe, Melissa J., Chun Yun Chang, Melissa A. Liu, Hannah M. Batchelor, Lauren E. Mueller, Joshua L. Jones, Yael Niv, and Geoffrey Schoenbaum. 2017. “Dopamine Transients Are Sufficient and Necessary for Acquisition of Model-Based Associations.” *Nature Neuroscience* 20 (5): 735–42.

Shibata, Kazuhisa, Giuseppe Lisi, Aurelio Cortese, Takeo Watanabe, Yuka Sasaki, and Mitsuo Kawato. 2019. “Toward a Comprehensive Understanding of the Neural Mechanisms of Decoded Neurofeedback.” *NeuroImage* 188 (July 2018): 539–56.

Shibata, Kazuhisa, Takeo Watanabe, Yuka Sasaki, and Mitsuo Kawato. 2011. “Perceptual Learning Incepted by Decoded fMRI Neurofeedback without Stimulus Presentation.” *Science* 334 (6061): 1413–15.

Sitaram, Ranganatha, Tomas Ros, Luke Stoeckel, Sven Haller, Frank Scharnowski, Jarrod Lewis-Peacock, Nikolaus Weiskopf, et al. 2017. “Closed-Loop Brain Training: The Science of Neurofeedback.” *Nature Reviews. Neuroscience* 18 (2): 86–100.

Skora, L. I., M. R. Yeomans, H. S. Crombag, and R. B. Scott. 2021. “Evidence That Instrumental Conditioning Requires Conscious Awareness in Humans.” *Cognition* 208 (March): 104546.

Squire, L. R. 1992. “Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory.” *Journal of Cognitive Neuroscience* 4 (3): 232–43.

Stachenfeld, Kimberly L., Matthew M. Botvinick, and Samuel J. Gershman. 2017. “The Hippocampus as a Predictive Map.” *Nature Neuroscience* 20 (11): 1643–53.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. 2nd ed. Vol. 66. Cambridge, MA: MIT Press.

Takakuwa, Norihiro, Rikako Kato, Peter Redgrave, and Tadashi Isa. 2017. “Emergence of Visually-Evoked Reward Expectation Signals in Dopamine Neurons via the Superior Colliculus in V1 Lesioned Monkeys.” *eLife* 6 (June). https://doi.org/10.7554/eLife.24459.

Tamaki, Masako, Aaron V. Berard, Tyler Barnes-Diana, Jesse Siegel, Takeo Watanabe, and Yuka Sasaki. 2020. “Reward Does Not Facilitate Visual Perceptual Learning until Sleep Occurs.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (2): 959–68.

Taschereau-Dumouchel, Vincent, Aurelio Cortese, Toshinori Chiba, J. D. Knotts, Mitsuo Kawato, and Hakwan Lau. 2018. “Towards an Unconscious Neural Reinforcement Intervention for Common Fears.” *Proceedings of the National Academy of Sciences* 115 (13): 3470–75.

Taschereau-Dumouchel, Vincent, Matthias Michel, Hakwan Lau, Stefan G. Hofmann, and Joseph E. LeDoux. 2022. “Putting the ‘Mental’ Back in ‘Mental Disorders’: A Perspective from Research on Fear and Anxiety.” *Molecular Psychiatry* 27 (3): 1322–30.

Thorndike, Edward. 1898. “Animal Intelligence: An Experimental Study of the Associative Processes in Animals.” *The Psychological Review: Monograph Supplements*, no. 2: i – 109.

Turner, Benjamin O., Jeanette A. Mumford, Russell A. Poldrack, and F. Gregory Ashby. 2012. “Spatiotemporal Activity Estimation for Multivoxel Pattern Analysis with Rapid Event-Related Designs.” *NeuroImage* 62 (3): 1429–38.

Vadillo, Miguel A., Emmanouil Konstantinidis, and David R. Shanks. 2016. “Underpowered Samples, False Negatives, and Unconscious Learning.” *Psychonomic Bulletin & Review* 23

(1): 87–102.

Vadillo, Miguel A., Simone Malejka, Daryl Y. H. Lee, Zoltan Dienes, and David R. Shanks. 2021. “Raising Awareness about Measurement Error in Research on Unconscious Mental Processes.” *Psychonomic Bulletin & Review*, June. https://doi.org/10.3758/s13423-021-01923-y.

Wang, Fang, Geoffrey Schoenbaum, and Thorsten Kahnt. 2020. “Interactions between Human Orbitofrontal Cortex and Hippocampus Support Model-Based Inference.” *PLoS Biology* 18 (1): e3000578.

Watanabe, Takeo, Yuka Sasaki, Kazuhisa Shibata, and Mitsuo Kawato. 2017. “Advances in fMRI Real-Time Neurofeedback.” *Trends in Cognitive Sciences* 21 (12): 997–1010.

Watson, Poppy, Claire O’Callaghan, Iain Perkes, Laura Bradfield, and Karly Turner. 2022. “Making Habits Measurable beyond What They Are Not: A Focus on Associative Dual-Process Models.” *Neuroscience and Biobehavioral Reviews* 142 (September): 104869.

Williams, B. A. 1994. “Conditioned Reinforcement: Experimental and Theoretical Issues.” *The Behavior Analyst / MABA* 17 (2): 261–85.

Wimmer, G. Elliott, and Daphna Shohamy. 2012. “Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions.” *Science* 338 (6104): 270–73.

Wuensch, Lavinia, Eva R. Pool, and David Sander. 2021. “Individual Differences in Learning Positive Affective Value.” *Current Opinion in Behavioral Sciences* 39 (June): 19–26.