

MASTER THESIS

BY

SOFIA LAZAROVA

MATRICULATION NUMBER: 108018265082

COGNITIVE SCIENCE M.Sc.

GENERATIVE EFFECTS OF EXTINCTION LEARNING EMERGING
FROM AN EMBODIED ASSOCIATIVE NETWORK.

Supervisor I:

Prof. Dr. Sen Cheng
sen.cheng@rub.de

Supervisor II:

Dr. José R. Donoso
jose.donoso@rub.de

Ruhr-Universität Bochum

January 13, 2021

DECLARATION

I HEREWITH CERTIFY THAT I HAVE WRITTEN THIS WORK INDEPENDENTLY AND THAT I HAVE NOT USED ANY SOURCES OR AIDS OTHER THAN THOSE INDICATED, AND HAVE CITED ALL THE QUOTATIONS.



/Sofia Lazarova/

Contents

1	Abstract	3
2	Introduction	4
2.1	Operant conditioning	4
2.2	Extinction learning and the renewal effect	4
2.3	The generative effects of extinction	6
3	Methods	8
3.1	T-maze experiment	8
3.2	Computational model	9
3.3	Simulation	15
3.3.1	Virtual environment	15
3.3.2	Single trial dynamics	16
3.3.3	Experimental design	18
3.4	Analysis	21
4	Results	22
4.1	Variability of behavior in the experimental and simulated data . . .	22
4.2	Extinction-induced variability between subjects	25
4.2.1	Smooth vs. abrupt transitions	26
4.2.2	Omission dominant vs. alternative dominant	28
4.2.3	Resistance to extinction	30
5	Discussion	31
5.1	Associative model accounts for the extinction of a previously rein- forced response	33
5.2	Choice variability on a subject level	33
5.3	Abrupt transitions in an associative model?	34
5.4	On the emergence of alternative dominant behavior	34
5.5	Resistance to extinction	35
5.6	Limitations	36
5.7	Future Directions	37
6	Conclusion	38

1 Abstract

The adaptation to changes in the environment has a great importance for survival of the individual. A process which actively drives behavioral adaptation is extinction learning - the process of ceasing a previously learned behavior in response to altered reinforcement contingencies. An often neglected aspect of extinction learning is the generation of new, previously non-reinforced behaviors in parallel to the gradual elimination of the learned response. However, the mechanism that generates these behaviors is loosely understood. I analyzed the behavior of 16 rats executing a T-maze context-dependent extinction task. The analysis confirmed the presence of extinction learning on a sample level and uncovered a variety of individual behaviors observed during extinction. Then by adapting the model proposed in *Donoso et al., In review*, I demonstrated that an associative model can produce extinction learning as well as behavioral variability similar to the one observed in rats.

2 Introduction

2.1 Operant conditioning

Animals, including humans, are capable of establishing associations between stimuli and behaviour as they operate in their environment. Such associations usually arise as a function of the consequences of undertaken actions in a given context. Thus, the positive outcome (reinforcement) of an undertaken action tends to increase the likelihood of this behavior being observed again in a similar situation and vice versa - a negative outcome (punishment) would decrease the chances of this behavior being observed again. Such a process of learning through creating associations based on punishment and rewards is referred to as operant (or instrumental) conditioning (*Byrne, 2013; Skinner, 1938*).

Despite its simplicity (*Allen and Bekoff, 1999; Osvath and Osvath, 2008; Suddendorf, 2006*), operant conditioning plays a key role in shaping behaviour as it aids the identification of purposeful and gain-oriented actions. Learning through associations allows individuals to anticipate recurrences of meaningful events on the basis of already familiar cues. The ability to identify familiar circumstances aids the generation of appropriate behavior as it allows the reuse of actions which previously lead to success in similar situations. The importance of studying operant conditioning is further emphasized by the fact that it has been implicated in the development of addictive behaviors such as drug abuse (*Everitt and Robbins, 2005*) and overeating (*Bouton, 2011*).

2.2 Extinction learning and the renewal effect

While finding an optimal behavior in response to the environment is undoubtedly crucial for survival, in some cases, the ability to quickly change course of action due to changes in the environment is just as important.

Extinction refers to the gradual decrease in a response to a conditioned stimulus that arises when the stimulus is presented without a reinforcer (*Hartley and Phelps, 2012*). For example, take a rat that has acquired the knowledge (has been condi-

tioned) that the right arm of a T-maze contains a small amount of food. Such a rat will persistently choose to go to the right arm of the maze since this action is known to bring an added value (food reward in this case). If the food is removed, however, the existing association between going to the right arm and finding food will become meaningless. Thus, the rat would eventually cease to go for the right arm in favor of exploring other behavioral options. In such cases, the previously established association is said to undergo extinction. The importance of this phenomenon is emphasized by the fact that the ability to extinguish already established behaviors has been observed in multiple species ranging from invertebrates to Homo sapiens (*Eisenberg and Dudai, 2004; Gottfried and Dolan, 2004; K. A. Lattal et al., 2013; Lengersdorf et al., 2014; Stollhoffa et al., 2005*).

Extinction was predominantly seen in the past as a process that weakens the underlying original learning (e.g., *Pearce and Hall, 1980; Rescorla and Wagner, 1972*). However, nowadays it is widely accepted that extinction involves new learning that seems to be context-dependent (*Bouton, 2019; Bouton et al., 2006; Todd, 2013*). The latter view of extinction suggests that the memory trace formed during extinction is actively inhibiting the trace acquired prior to extinction. Whether this new learning is going to be expressed depends on the context which is also consistent with the idea that contextual cues can drive memory retrieval (*Tulving and Thomson, 1973*). A commonly observed phenomenon that provides support for this view is the renewal effect.

In the renewal effect, a previously extinguished behavior reemerges once the animal is removed from the context where extinction learning took place (*Bouton, 2019; Nieto et al., 2017; Todd, 2013; Todd et al., 2014*). With respect to the context change at least three forms of renewal have been described - AAB, ABC and ABA renewal (*Bouton, 2002*). However, here we will focus on the ABA renewal where acquisition is taking place in context A and extinction occurs in context B. Renewal of the extinguished behavior is observed once context A is presented again (*Bouton et al., 2011*). Note that while acquisition and renewal occur in the same context A, during renewal no rewards are provided to the subjects. Thus, the

renewal effect is emerging purely due to the existing contextual association. The fact that the previous conditioning reemerges again in the absence of the extinction context shows two things: 1) Extinction cannot be simply an erasing process; it is rather actively inhibiting the previous conditioning. 2) The context in which a learning process occurs is becoming directly associated with the contingencies of the environment and the presentation of a new context is likely to drive the formation of a new trace of learning. Thus, extinction can be explained in terms of a context-dependent learning process during which the subject is learning to inhibit the previous conditioning without erasing it. This mechanism offers the benefit of preserving already learned behaviors which is key for survival and behavioral flexibility. A behavior that was once useful might become useful again and in some cases a quick recovery of such behavior might be life saving. However, this feature of extinction can also contribute to relapses in scenarios like exposure therapy, where pathological behaviors that were extinguished during therapy emerge again once the patient goes back to their everyday environment (*Conklin and Tiffany, 2002; Lleving et al., 2004*). Therefore, understanding what drives and influences extinction and the renewal effect might be helpful in addressing such relapse episodes and preventing the reemergence of pathological behaviors (*Conklin and Tiffany, 2002*).

2.3 The generative effects of extinction

Apart from the elimination of a conditioned response, the procedures of extinction have been shown to also drive the emergence of new, previously non-reinforced behaviors (*Antonitis, 1951; Grow et al., 2008; K. M. Lattal and Lattal, 2012*).

Even though during extinction the previously reinforced response seems to be diminished, and even completely extinguished, it is certainly not the case that no behavioral patterns are observed at this time. Rather, while the rate of the conditioned response gradually decreases, other behaviors seem to emerge. These newly emerging behaviors are commonly referred to as the generative effects of extinction. A commonly observed generative effect is the extinction-induced variability where subjects are exhibiting variations of response topography after the onset of

extinction (*K. A. Lattal et al., 2013*). One of the earliest studies demonstrating extinction-induced variability was conducted by Joseph Antonitis (*Antonitis, 1951*). In his study rats were reinforced for poking their noses anywhere along a 50-cm opening. Even though an exact location was not reinforced, the location of the pokes became increasingly predictable during acquisition. Once reinforcement was withheld, behavioral variability increased and responses started occurring all along the strip (*Antonitis, 1951*). Extinction-induced variability has been observed across other response dimension as well as multiple species ranging from bacteria to humans (*Neuringer and Kornell, 2001*).

Even though there is a substantial body of evidence showing that extinction induces behavioral variability (*Neuringer and Kornell, 2001*), it is still a common practice for studies of operant conditioning to account only for the eliminative quality of extinction. As a result, the effect that extinction might have on other available actions is often neglected. For example, in the case of a rat put in a T-maze and conditioned to choose the right arm, traditionally, the focus during extinction will be on whether the rat ceases to choose the right arm even though there are more behavioral options (e.g., going left, returning, not leaving the starting point) available to the rat (e.g., *Méndez-Couz et al., 2019*). Thus, this gap in the literature leaves the impact that extinction has on decision-making partly unexplored. Shedding light on this matter might be crucial in cases where multiple behavioral options are available at the time of extinction. This is especially true in cases where some of these options might be undesired or even disruptive.

The variability caused by extinction can potentially induce the expression of every behavior available to the subject during extinction. Thus, there is no guarantee that a newly generated behavior is more socially appropriate or even beneficial than the extinguished one. This implication may play a critical role in the treatment of pathological behaviors since it may be the case that by applying extinction to one form of disruptive behaviour we might unintentionally facilitate the emergence of even more severe forms of disruption (*Lleving et al., 2004*). In fact, if there is no specific alternative response to be reinforced, there is no guarantee what newly

generated behaviours may occur (*K. A. Lattal et al., 2013*). However, the same qualities of extinction may also ensure the generation of new responses that can be targeted for reinforcement. This way the generative aspect of extinction can be used to facilitate the exchange of an unfavorable behavior for more appropriate forms of response (e.g., *Grow et al., 2008*).

The purpose of this thesis is to explore the putative mechanisms underlying extinction learning and to assess the extent to which a minimalistic associative model can account for the dynamics observed in the behavioral data. To address the gap in regard to the generative aspects of extinction, instead of focusing only on the eliminative quality of the extinction procedure, the scope of this work includes investigation of the extinction-induced behavioral variability as well. Within this thesis I will present the results from analyzing behavioral data from rats collected during the execution of a T-maze context-dependent extinction task and further show that the computational model, proposed by *Donoso et al., In review* can, at least partially, account for the rich behavioral phenomena observed in rats.

3 Methods

3.1 T-maze experiment

For the purpose of my investigation I used data from 16 rats obtained in an experiment, designed and performed by *Méndez-Couz et al., 2019*. The experiment was designed to investigate the context-dependent extinction in rats. Both neural and behavioral data was collected during the experimental sessions. The animals were trained to choose a predefined goal arm (always right arm) in a T-maze. The maze contained both visuospatial and olfactospatial cues. The floor of the maze had a specific, context-dependent pattern and identical odor cues (Vanilla food aroma, Dr. Oetker, Bielefeld, Germany) were placed at the end of each sidearm. Both arms of the maze contained identical food ponds but food could be found only in the right arm.

The T-maze was composed of a starting box (25×20 cm) with a sliding door

that separated the starting box from the main corridor (100×20 cm) (*Méndez-Couz et al., 2019*). Two side arms (10×40 cm) were positioned at the end of the corridor (*Méndez-Couz et al., 2019*). Two different contexts were used for the different phases of the experiment: context A describes the conditions used for the acquisition and renewal phases, whereas context B describes the conditions during extinction (*Méndez-Couz et al., 2019*).

The experiment had an acquisition phase spanning over 3 days as well as extinction and renewal phases that took place on the 4th and final day of the experiment. Each day consisted of 20 trials. On the first day the probability of obtaining a reward was 100% and then gradually decreased to 80% on the second day and to 30% on the last sessions of the third day. The duration of one trial was maximum 2 minutes or until the animal got a reward from the pond. The inter-trial time was 15 seconds. During the runs rats were let to explore the maze and enter the correct or incorrect arm. Both maze arms were provided with identical ponds. However, only the correct choice was rewarded with a chocolate sprinkle inside the pond. In case of an incorrect choice the chosen arm was blocked and the animal would stay in the non-reinforced arm for a period of 15 seconds before being returned to the starting position. The phases of extinction and renewal were both carried out on the fourth day of the experiment. Just like the first three days, the final day of the experiment had a total of 20 trials. However, in this case, the first 15 trials made up the extinction phase and the last 5 trials counted for renewal.

3.2 Computational model

This thesis proposes the use of a model based on a simple associative network and a winner-takes-all decision process, originally proposed by *Donoso et al., In review* to explain pigeon behavior in an analogous experiment. In the current work I am adapting this model to the T-maze task described above.

The model is proposing a simple mechanism associating sensory input with motor output (*Donoso et al., In review*). Sensors and executors are connected by inhibitory and excitatory synapses which values are altered whenever the activation of a sensor

leads to the activation of an executor (*Donoso et al., In review*). This way the synaptic values are representing the association between sensory input and motor output. For instance, if the agent makes a right turn as a result of a sensory input pattern, the values of all the synapses connected to the respective executor will be altered (*Figure 1*).

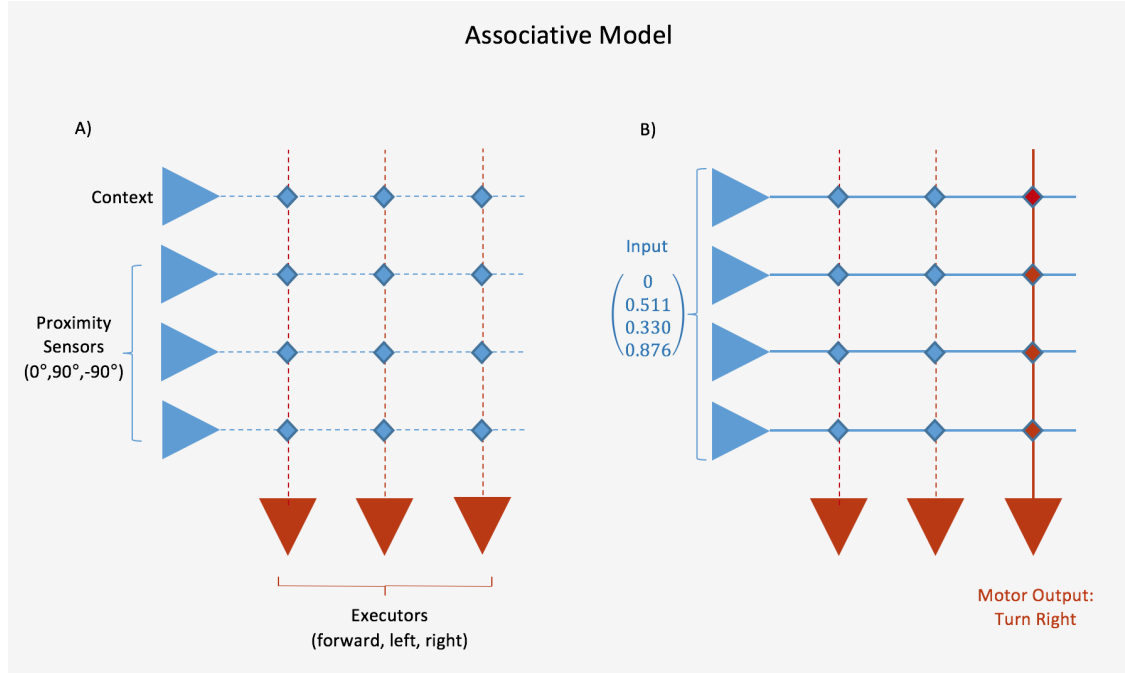


Figure 1: A schematic representation of the associating sensory input with motor output. A) The model is directly associating sensory input captured by proximity sensors with motor output and is able to generate three possible actions (moving forward, left or right) each action corresponding to a single executor. Sensors and executors are connected with synapses (blue diamonds). Note that the context is treated as another stimulus and is creating associative connections with the motor output. B) A sample input is shown to illustrate the mechanism of altering synaptic values. The sensory input is given as a vector which generates a particular action in the case "turn right". The synapses connecting the respective executor with the input (marked in red) are the ones which values are being altered in order to reflect the association between the input and the resulting choice of action.

The model is guided by raw sensory input obtained by three proximity sensors each providing information about the proximal distance to any objects located in front, to the left or to the right of the agent. The sensors are located at 0°, 90° and -90° providing a greater ability to explore the environment. Sensory input values are normalized between 0.0 and 1.0. Accompanying the proximal input, an additional input value is given in order to signify the current context. The context value is binary and can take the values of 0 and 1 where 0 represents the learning

context(context A) and 1 represents context B. Thus, the context is treated the same way as a stimulus - it can establish direct inhibitory and excitatory associations with specific motor actions (*Donoso et al., In review*). This representation of the context is in accordance with recent studies (*Bernal-Gamboa et al., 2018; Bouton, 2019; Nieto et al., 2017; Todd, 2013; Todd et al., 2014*).

There are three possible courses of action available to the virtual agent at any point of time - go forward, turn left and turn right. Each action is represented by a single executor and only one executor can be active at a time. Meaning the virtual agent can execute only one action at a time. Turning away (left and right) actions are defined as turn 1° in the designated direction and make a 2-point move forward. Respectively forward movement is defined as 2-point move forward (*Figure 2*). This small-step definition of the available actions, supports the smooth and continuous movement of the agent.

Next, I will provide a formal description of the associative model and its operational flow. *Figure 3* provides a visual overview of the dynamics within the model and its interaction with the environment.

The implementation of this model allows the stimuli (and the context) to create both excitatory and inhibitory associations. Thus, the model has two pathways that are influencing the choice of behavior. Respectively one pathway keeps track of the inhibitory associations and the other one of the excitatory associations. Therefore, the synaptic input is given by:

$$u = W_{exc}s - W_{inh}s$$

where u is a vector containing the input to the motor units, W_{exc} and W_{inh} are matrices containing the excitatory and inhibitory synaptic weights and s is the sensory input available at a given moment. All inhibitory and excitatory synaptic weights are initialized with values 0.0. The sensory input is given as a vector containing the proximity sensory data plus a context value. The executors are implemented as rectifying linear units driven by the net synaptic input and excitatory noise:

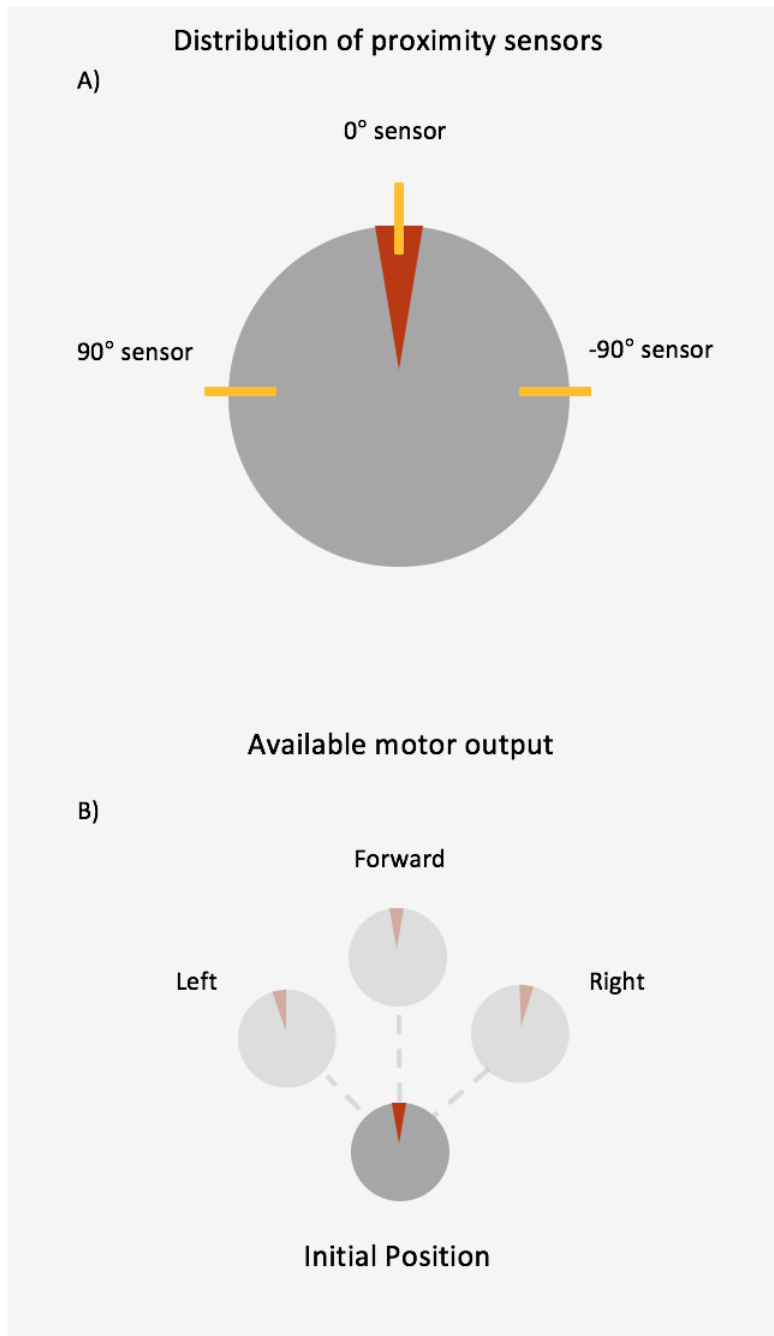


Figure 2: Virtual agent with its capacities of sensory input and motor output. A) Locations of the proximity sensors. B) Available motor output. At any point of time the agent has three motor actions available - forward, left and right. Moving forward is defined as moving forward with two points. Turning away (left and right) is defined as turning 1° in the respective direction and moving two points forward.

$$m = \text{ReLU}(u + \epsilon)$$

where m is a three element vector describing the possible actions executed by the motor units and ϵ is a three element vector containing the accompanying noise

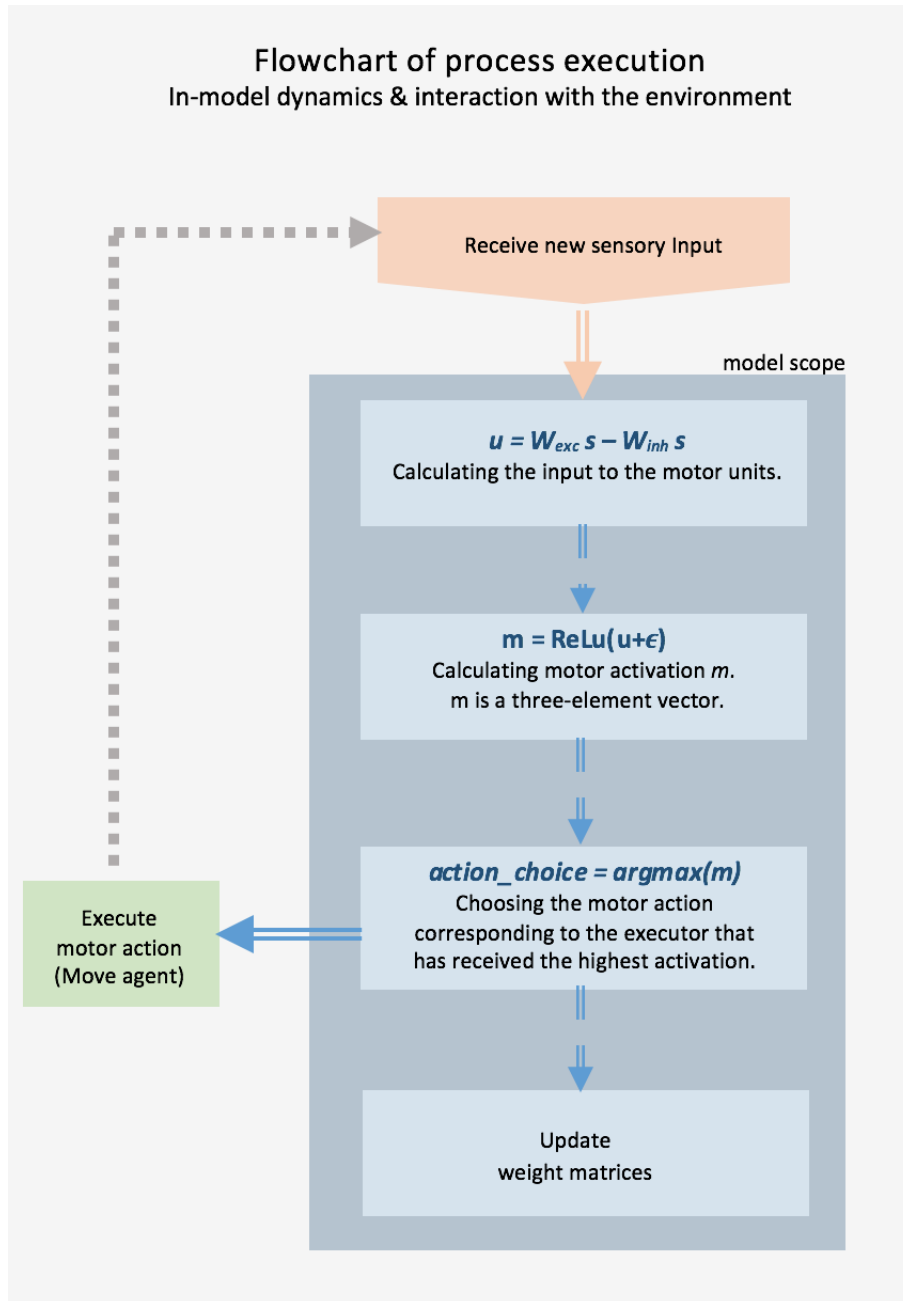


Figure 3: The operational flow of the model and its most basic interactions with the environment. The model is kept as a cohesive unit that receives sensory input on which it operates and as a result outputs a single motor action to be executed. This dynamic creates a circular flow between the agent and the environment. As the agent executes a motor action the sensory flow changes which as a result produces a new motor action and so on.

inputs. The values for ϵ are drawn from independent uniform distributions on the interval (0, 1). The rectifying activation function returns 0 for all values that are less than 0 and keeps the original values in case they are greater than 0. This way m contains values greater or equal to 0 that correspond to the activation passed to the respective executor. The choice of motor action is based on the 'winner-takes-it-all'

principle. In other words, the executor receiving the greatest activation value will be one that will get activated. Thus, choice of action is defined as follows:

$$choice = \operatorname{argmax}(m)$$

In case of insufficient activation values, where m has value of 0, a random command is executed.

On every iteration or every time a move is executed both excitatory and inhibitory connections between the sensory units and the respective motor unit are being reinforced. This way the weight matrices are being updated to reflect every choice of action that has been undertaken while traversing the maze. Associative strengths are updated on every iteration as follows:

$$\Delta w_{ij} = \Delta_{exc,inh}(w_{exc,inh}^{\infty} - w_{ij})$$

where Δw_{ij} is the change of the synaptic connection between input i and a motor unit j , $\Delta_{exc,inh}$ is the learning rate of excitation or inhibition and $w_{exc,inh}^{\infty}$ is the maximum possible value that excitatory or inhibitory weights can reach. The maximum synaptic value is set to 20.0. The learning rates of inhibition and excitation are not identical - inhibition has a slower learning rate set to 4×10^{-7} while excitation has a learning rate of 9×10^{-6} .

Once a trial has ended the synaptic weights are either consolidated or let to decay naturally to the values they had at the beginning of the trial. The synapses of both the excitatory and the inhibitory pathways are implemented with a synaptic decay which is given by:

$$w(i) = w_0 e^{(-\lambda i)}$$

Where i is the number of the current iteration and w_0 is the initial synaptic value. Every time a synaptic value gets altered the synaptic decay is being activated. The exponential decay of the synaptic values ensures that in case of consolidation the

most recent actions that were undertaken will have a higher significance than the ones prior to them. λ is the exponential constant and is set to 9×10^{-7} .

If the agent performs a successful trial and obtains a reward the weights of the excitatory matrix are being consolidated while the inhibitory weights are left to naturally decay to the values they had at the beginning of the trial. Conversely, if the agent reaches the target arm but no reward is provided or it fails to reach the target arm altogether then the inhibitory weights are being consolidated while the excitatory connections are left to decay.

3.3 Simulation

The model was tested with the means of a virtual simulation, a method which ensures a reliable and consistent environment with minimum external interference. In order to recreate the conditions under which the original experiment was carried out, the environment of the maze and the supporting conditions were modeled using Blender (<https://www.blender.org/>) and the Python programming language (Python Software Foundation, <https://www.python.org/>).

A single run of the simulation produces the data equivalent to the data obtained from a single animal. In other words, the data produced by a single run of the simulation can be treated as the run of one simulated animal.

3.3.1 Virtual environment

The environment was modeled as a 3D T-maze spacious enough to host the free movement of the virtual agent. The arena and the agent are illustrated in *Figure 4*. The proportions of the maze were kept similar to the proportion of the T-maze used in the animal experiment. The simulated maze does not depict reward trays and context cues because their visual presence is irrelevant for testing the model and would have a purely decorative nature. The agent has a very simplified shape and design, the only feature of concern is an indicator showing the facing (*Figure 4, the red marker*).

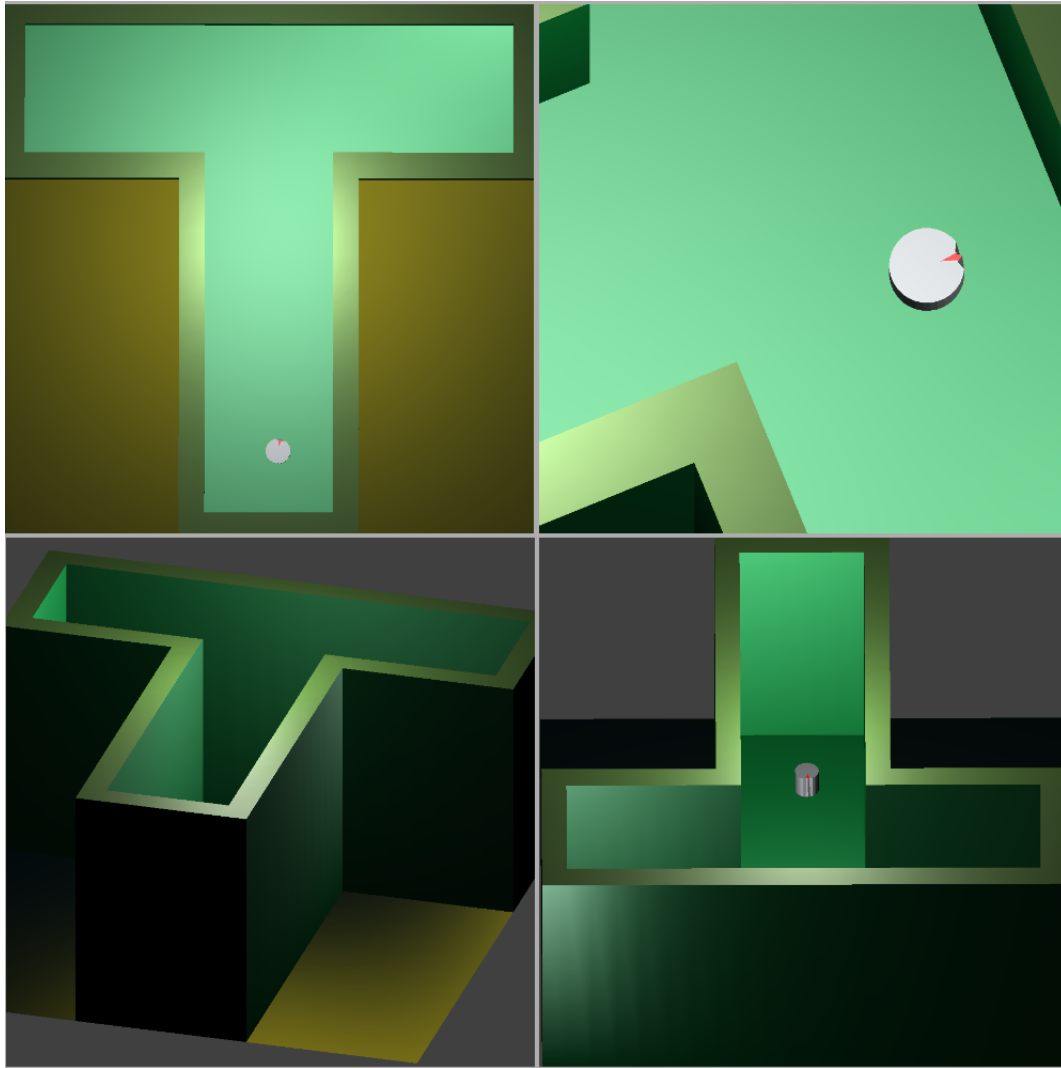


Figure 4: Screenshots of the Blender environment while running the simulation. T-maze arena and virtual agent displayed from different angles.

3.3.2 Single trial dynamics

Each trial has a maximum duration of 90s. At the beginning of each trial the agent is placed at the starting position and let to navigate the maze freely (*Figure 6*). The starting position remained unchanged throughout the trials. A trial ends in one of four conditions - target arm is reached, alternative arm is reached, the agent returns to the starting zone or time out. Thus, there are three relevant zones in the layout of the maze that have a landmark significance and the event of entering them is terminating the current trial (*Figure 5*).

Once a trial begins, the virtual agent is left to explore the maze freely. The path traversed by the agent is forming a directed trajectory with beginning the 'start zone'. This trajectory is recorded and can be later used to visualize the motion

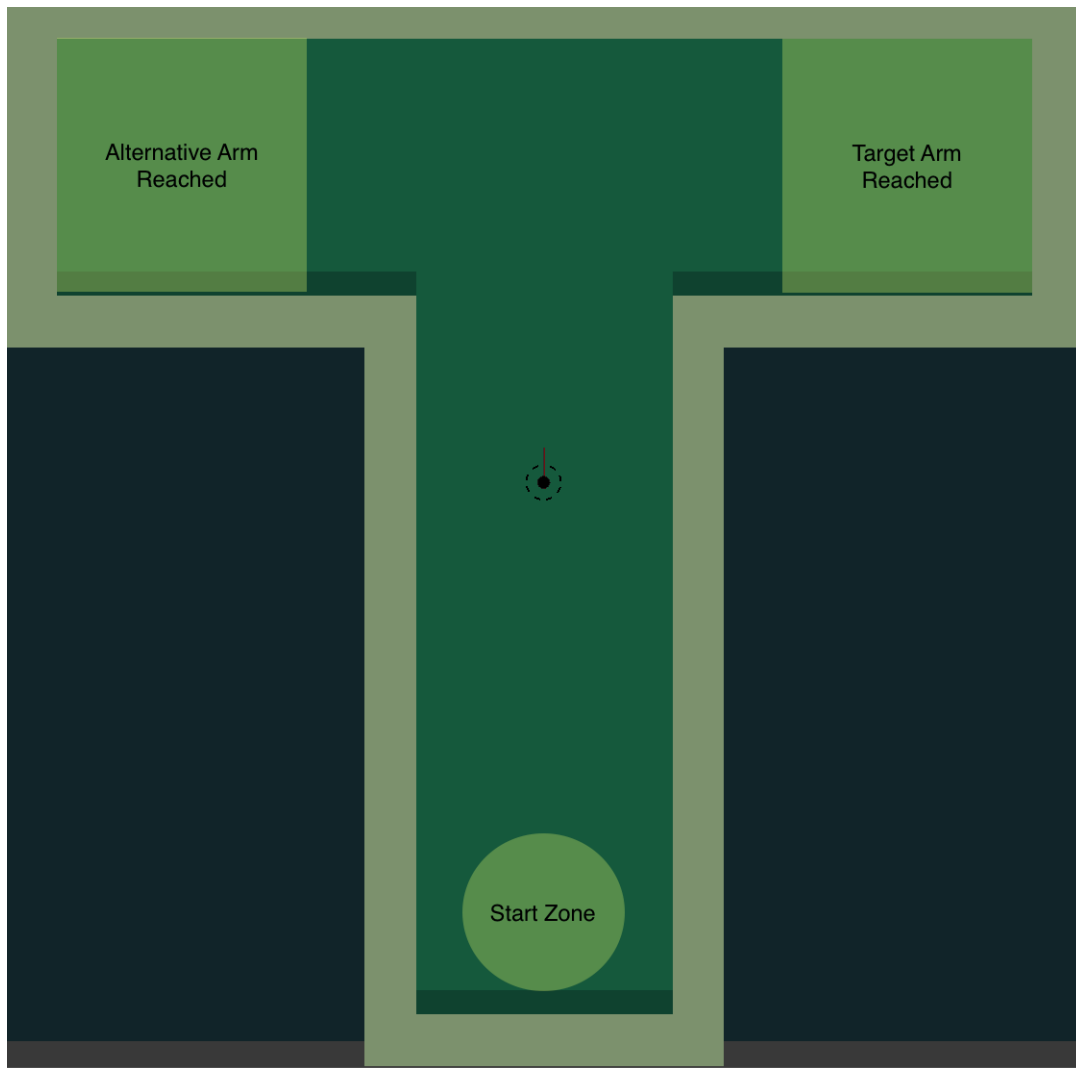


Figure 5: Landmark zones in the layout of the maze. Entering any of these zones after a trial has begun is terminating the execution of the current trial.

of the agent during a particular trial. The ending point of a trajectory shows the outcome of the respective trial. *Figure 6* follows the formation of a trajectory from the beginning to the end of a trial. In the depicted trial the agent chooses the alternative arm.

Such a trajectory is produced by every trial and is represented as a collection of points in the two-dimensional space, recorded as x and y coordinates. The coordinates are recorded on every iteration from the beginning of the trial to the end of the trial. The number of the $\langle x, y \rangle$ tuples forming the trajectory depends on the duration of the trial - the longer the trial is the more data points will be recorded. The notation used to save and plot path coordinates is the same as the notation used in the behavioral data obtained from the rat experiment. *Figure 7* illustrates

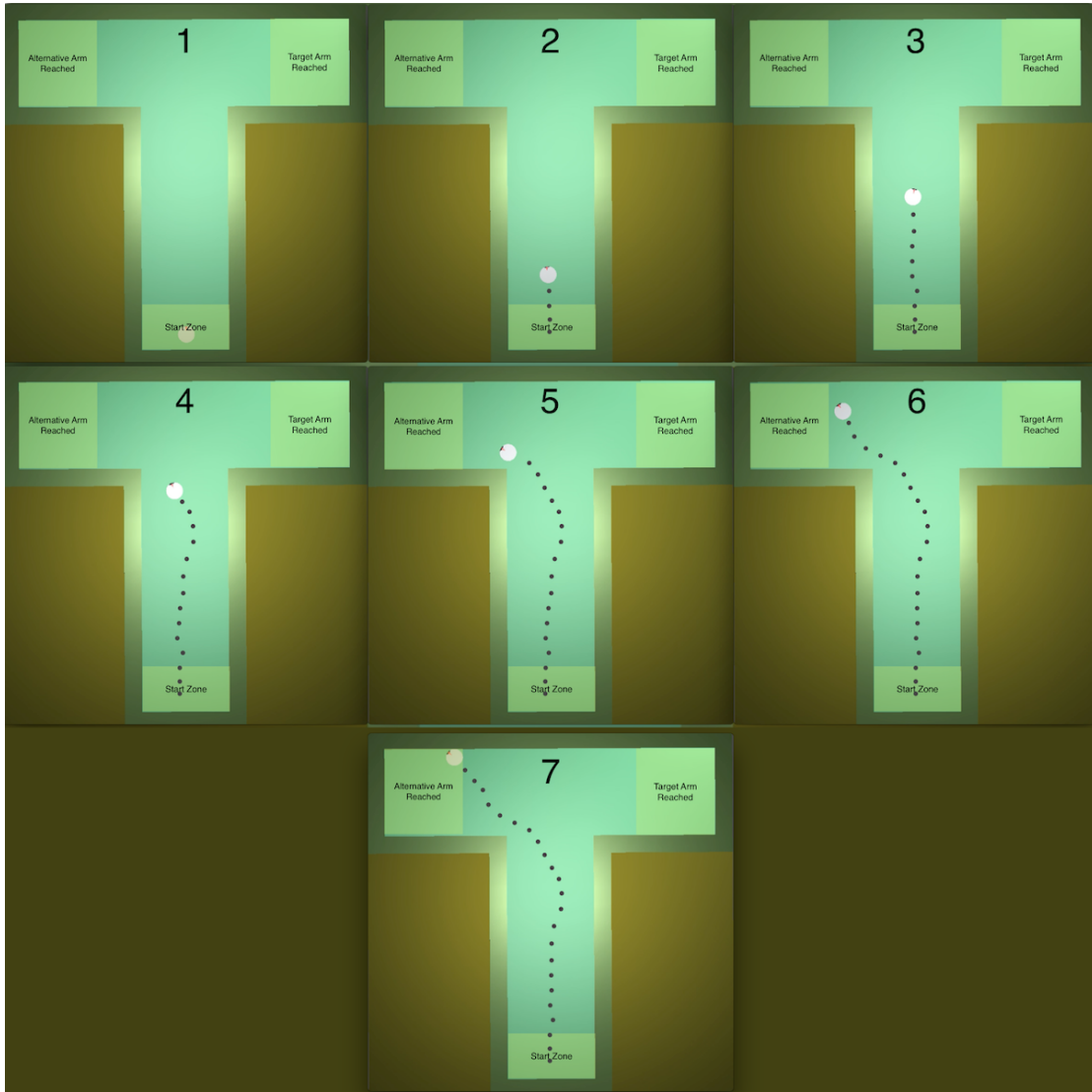


Figure 6: Trajectory formation during an incorrect trial. Panels 1-7 are depicting consequent screenshots following the progress of a trial. At the beginning of the trial the agent is located at the starting position within the 'start zone' (Panel 1). During the trial the agent is freely navigating from the 'start zone' to the 'alternative arm' (Panels 2-6). This motion is depicted as a dotted path.

various examples of trial trajectories obtained from behavioral and simulation data.

3.3.3 Experimental design

The experimental paradigm used by the simulation is a simplified version of the design used to conduct the rat experiment (*Méndez-Couz et al., 2019*). However, the simulation contains more trials compared to the rat experiment. While rats exhibit the ability to learn fast the rules of such a restricted environment, a strictly associative model is likely to need more repetitions to establish associations and start learning. A single simulation contains a total of 250 trials separated into three

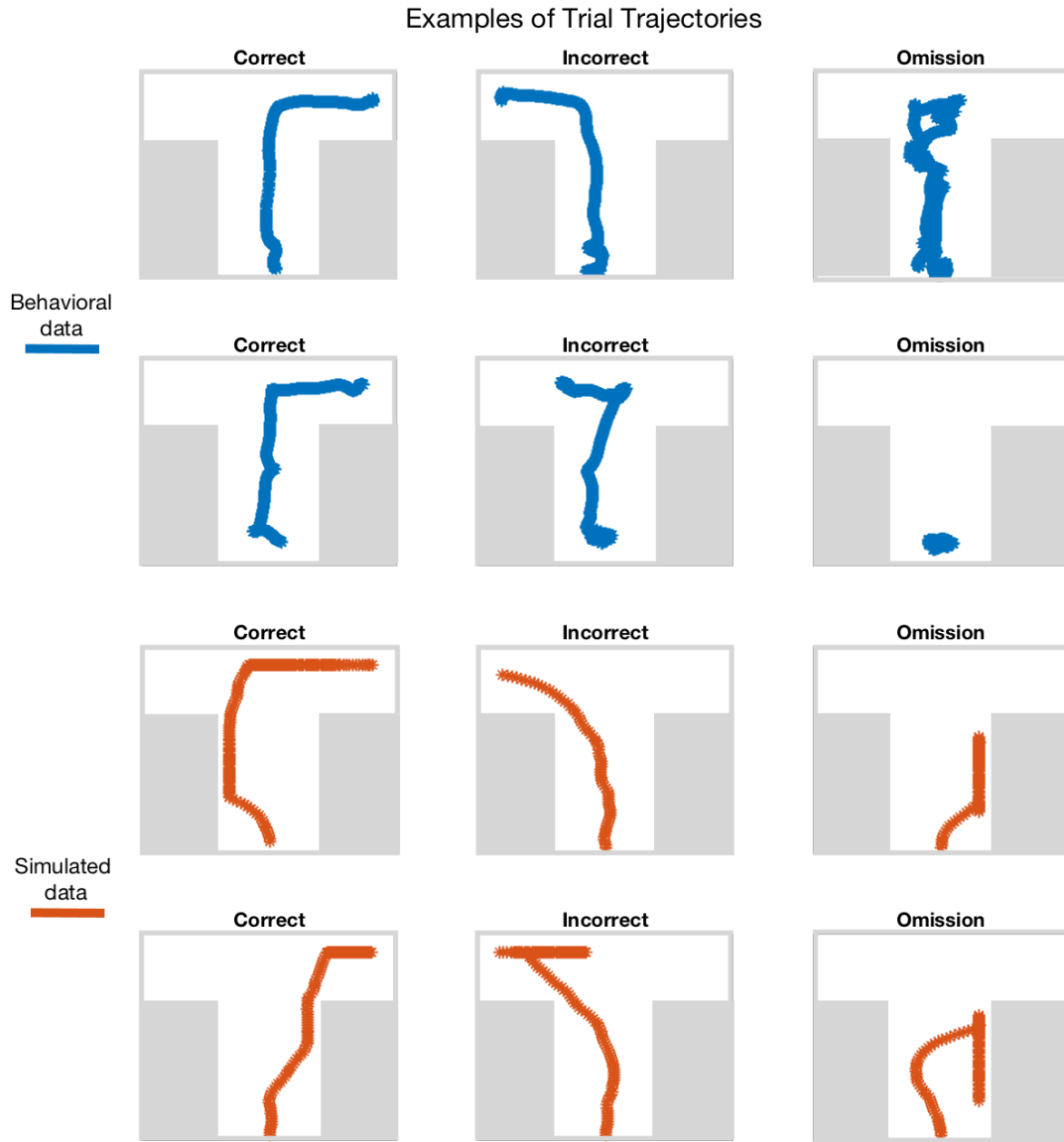


Figure 7: Sample trial trajectories obtained from the behavioral data (in blue) and from the simulation data (in red). Note that all behaviors that are not reaching an arm are considered as omissions for the purpose of this thesis.

phases - acquisition, extinction and renewal (*Figure 8*). There are two possible contexts that are presented during the different phases of the experiment. Context A is presented during acquisition and renewal and context B during extinction. Contexts are passed to the model as binary values - 0 for acquisition and renewal and 1 for extinction.

Acquisition phase: The acquisition phase expands over the first 100 trials. During acquisition, the agent is learning to choose the right arm (target arm) so this behavior is being rewarded every time it is executed. Unlike the rat experiment where the

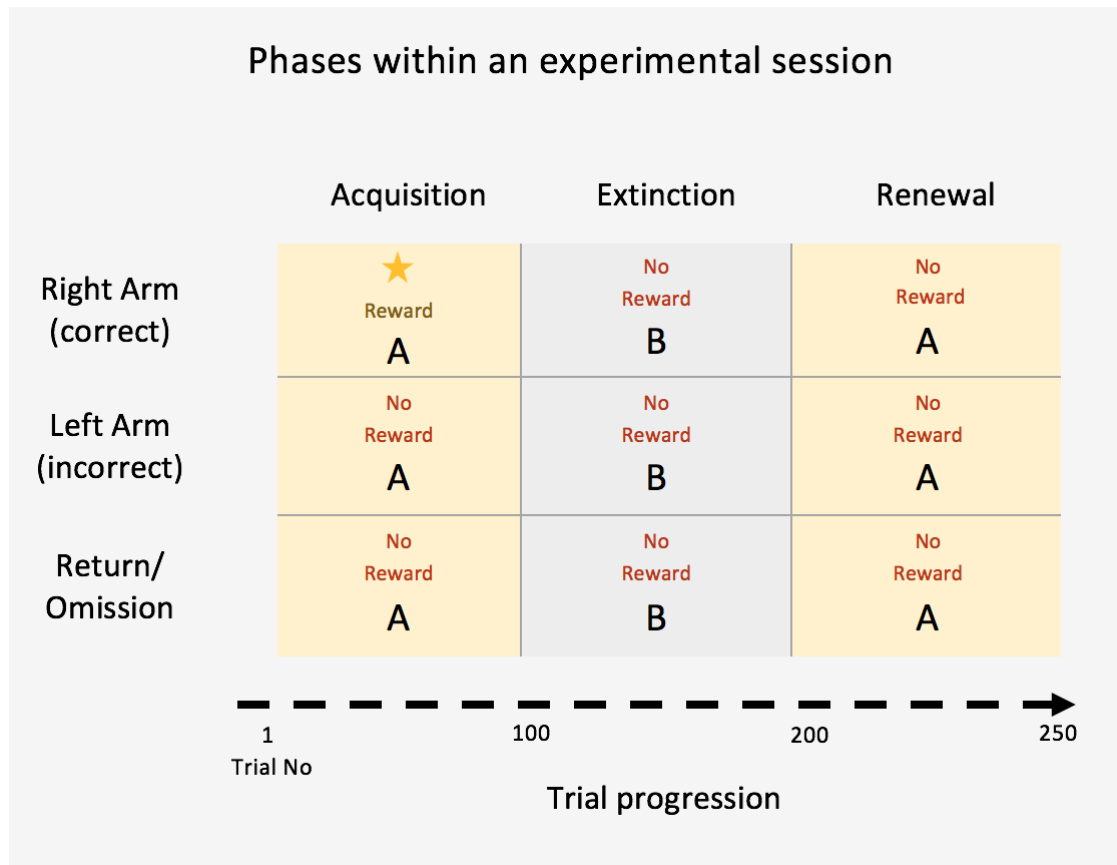


Figure 8: Experimental design. The experiment consist of three phases - acquisition, extinction and renewal with respective duration 100, 100 and 50 trials. Rewards are provided only during acquisition in case the reinforced behavior(choosing the right arm) is performed by the agent. Acquisition and renewal were performed in context A whereas, during extinction, context B was presented. Note that the context was presented as a binary value so context A corresponds to 0 and context B to 1.

rate of providing a reward is declining, here a reward is given 100% of the time upon reaching the target arm.

Extinction phase: The extinction phase has a duration of 100 trials and is taking place from trial 101 to trial 200 including. During extinction no rewards are provided and there is no rewarded course of action. It is important to note that during extinction the context is changed.

Renewal test: The last 50 trials are making up for the renewal test. During the renewal phase the original context, presented during acquisition, is returned while the delivery of rewards is discontinued.

The results of each experimental session were saved and exported in JSON format, a lightweight and language-independent data format allowing further flexibility in the choice of tools for data analysis (<https://www.json.org/>). The saved data con-

tains metadata related to the simulation and trial-specific data - path trajectories recorded in X and Y coordinates, trial outcome, duration and reward availability.

3.4 Analysis

The software used for conversion and analysis was Matlab R2019a (Mathworks Inc., <https://www.mathworks.com/>). All statistical analyses were conducted with RStudio version 1.1.383 (<http://www.rstudio.com/>). The full implementation of the hereby described work, including the statistical analyses, is open-sourced and available on GitHub (https://github.com/sophialazarova/GEffects_Ext).

All analyses described here were applied both to the behavioral data obtained from the rat experiment and to the simulated data. The choice of analyses was guided by the main purpose of this thesis - to assess the extent to which an associative model can account for the dynamics observed in the behavioral data.

The preference for choice response is playing a key role in the analysis applied here. As already mentioned, there are three types of possible behaviors - reinforced(learned) choice, alternative choice and omission. Estimating the preference for each of these behaviors can uncover the presence of behavioral changes during a phase or between phases. Furthermore, by exploring the choice preference during extinction one can also make suggestions about the generative aspects of extinction learning. The preference for a choice response was quantified by expressing the target responses as a proportion of the total number of trials considered. This way, the choice rate is expressed as k/N where k is the counted responses and N is the total number of trials. The difference between proportions was measured with one-tailed χ^2 statistical tests. A significant difference between proportions was considered to be a difference that yields a p-value below .05. Additionally, one-tailed two-sample t-tests were used to establish the difference in the means of the response rates calculated for learned choice, alternative choice and omissions. Again, results were considered for significant if the p-values were below .05.

To visualize the behavioral variability on a subject level, I plotted the cumulative record of responses as a function of trial number (*Gallistel et al., 2004; Leslie et*

al., 2005). Traditionally, the cumulative record is created in a unipolar fashion which signals only the presence or the absence of the learned choice with 1 and 0 respectively. This kind of encoding is useful when the focus is only on the learned response which is often the case. However, in the present work the data is encoded in a bipolar fashion as suggested by *Donoso et al.*, *In review*. Under this notation not only the learned choices are encoded but also the alternative choices and omissions respectively with 1, -1 and 0 (*Donoso et al.*, *In review*). The bipolar encoding reveals biases towards any of the possible behavioral choices. Thus, a learning curve with a positive or a negative slope reveals a tendency to prefer the learned or the alternative choice, respectively. A slope of 0 indicates either a chain of omissions or a random behavior expressed as a mixture of different responses (*Donoso et al.*, *In review*).

The analysis of the behavioral data included a total number of 16 rats while the total number of simulated subjects used for analysis was 5.

4 Results

4.1 Variability of behavior in the experimental and simulated data

In the following section, the focus will be on the choice preference expressed during the three experimental stages. For the purpose of the analysis, the choice preference is expressed as the response rate proportion k/N where k is the number of target outcomes and N is the number of trials in the analyzed batch.

The behavioral data obtained from the rat experiment showed a high level of preference for the learned choice during the last trials of acquisition, suggesting that the animals learned successfully (Figure 9, A left).

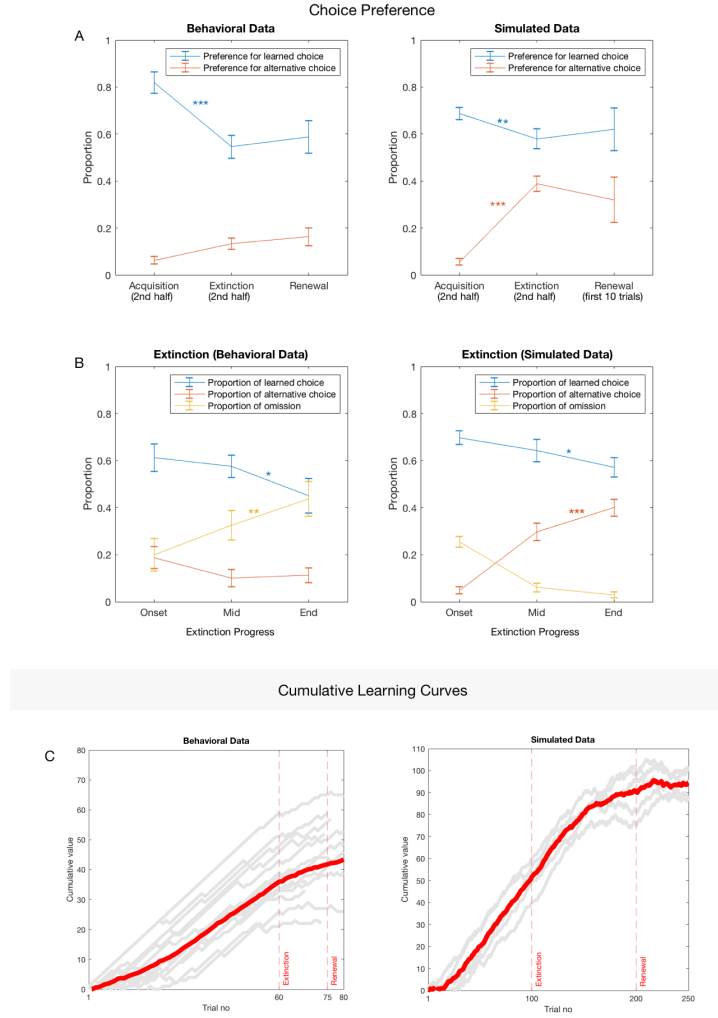


Figure 9: Dynamics of choice preference. A) The average preference of rats (left) and simulated subjects (right) plotted against phase progression. There is a significant decay of learned choices during extinction in both rats ($\chi^2 = 47.514$, $p < .001$) and simulated subjects ($\chi^2 = 5.5591$, $p < .01$). Simulated subjects are expressing an increased preference for the alternative choice during extinction ($\chi^2 = 66.208$, $p < .001$) compared to acquisition. B) Average preference of rats (left) and simulated subjects (right) during extinction. Both behavioral ($\chi^2 = 5.6285$, $p = .02865$) and simulated data ($\chi^2 = 5.2247$, $p = .01113$) is showing a significantly decreasing preference for the learned choice during extinction. While the behavioral data is showing an increase in the proportion of omission ($\chi^2 = 9.3254$, $p = .00113$) the simulated subjects shows an increased preference for the alternative choice ($\chi^2 = 56.999$, $p < .001$). C) Cumulative learning curves for rats (left) and simulated subjects (right). Marked in red are the average learning curves.

During the second half of the extinction phase, the proportion of trials having the learned choice as an outcome decreases significantly ($\chi^2 = 47.514$, $p < .001$) compared to the second half of the acquisition (Figure 9, A left). This observation suggests that the learned behavior was successfully extinguished. Curiously, there seem to be no statistically meaningful increase in the learned responses during renewal suggesting that, at least on a sample level, the renewal effect did not occur.

Thus, according to the presented data we can conclude two things: first, the observed sample of rats is gradually extinguishing the learned response while exhibiting an increasing preference for omissions. Second, upon return of the original context A, the rats are not exhibiting a significant recovery of their preference for the learned response.

Identical analysis showed that the simulated data exhibited similar properties to the behavioral data (Figure 9, A, B, C right). The preference for the learned choice during acquisition is again well-defined, even though it is not as strong as the preference observed in the behavioral data. This suggests that a level of learning is present but it is not as pronounced as in real rats. However, this behavior might be at least partially accounted for by the choice of a learning rate. Similarly to the behavioral data, a significant decrease in the preference for the learned choice was observed between acquisition and extinction ($\chi^2 = 5.5591$, $p < .01$). Interestingly, there is a significant spike of the preference for the alternative choice during extinction ($\chi^2 = 66.208$, $p < .001$) which was not present in the behavioral data. This finding suggests that while extinction of the reinforced behavior is observed in the experimental and the simulated data, it may be the case that the generative effects driven by the extinction procedure are somewhat different.

To get a fine-grained view of extinction-induced behaviors I analyzed the dynamics of choice preference during extinction. Both in the behavioral and simulated data, the choice proportion of the learned choices decreased significantly by the end of the extinction phase (Figure 9, B). The decrease begins from the onset of the extinction and gradually progresses. Note that in the second half of the extinction phase, the decrease is more rapid than during the first half. Also, as the preference

for the reinforced choice gradually declines, the preference for omissions seems to grow linearly in rats (*Figure 9, B left*). This is in line with the widespread view of extinction learning according to which when animals experience a withdraw of reinforcers upon the onset of context B (onset of extinction) they initially persist on the reinforced choice before gradually change their behavior towards omissions (*Nevin, 2012; Podlesnik and Shahan, 2009, 2010*). In the simulated data, on the other hand, the preference for the alternative choice seems to be gradually and persistently increasing (*Figure 9, B right*). Thus, the simulated rats seem to gradually change their behavior towards the alternative choice rather than omissions.

4.2 Extinction-induced variability between subjects

So far we saw that the associative model has the capacity to partially reproduce the average behavior of rats during extinction. However, while the simulated sample shows a strong preference for the alternative choice during extinction the average rat behavior seems to change towards omissions. Thus, in terms of generative effects, the average performance of the computational model is overall differing from the average rat behavior. To investigate this difference I further analyzed the behavior of the rats on a subject level in order to obtain a more detailed picture of the behavioral variability expressed between subjects. The traditional approach of averaging data across subjects has the downside of potentially masking behaviors. Thus, analyzing the individual subjects might uncover hidden behavioral patterns expressed during extinction.

Just as expected, the subject-level analysis uncovered several types of behaviors which were concealed by the average analysis. For the rest of this section I will present the results from the subject-level analysis which was focused mainly on two criteria - the type of transition from acquisition to extinction and the preferentially expressed choice during extinction.

4.2.1 Smooth vs. abrupt transitions

The analysis of the average behavioral variability suggested that, on average, the animals are behaving according to the canonical view of extinction (*Nevin, 2012; Podlesnik and Shahan, 2009, 2010*). Namely, they persist on choosing the right arm before gradually change their behavior towards omissions. To further explore the soundness of this observation, I assessed the smoothness of the transitions upon introducing context B (Supplementary Images). 15 rats were included in the analysis since one rat was excluded due to an onset of strong extinction-like behavior during the last 20 trials of acquisition. To quantify the presence of abrupt and smooth transitions, I focused only on the first 5 trials after the first presentation of context B. A transition was considered abrupt if the rat emitted at least 3 non-reinforced responses in the first 5 trials. Otherwise, the rat was considered to show a clear preference for the learned choice and thus, the transition was classified smooth (or canonical). In total 3 rats (20%) exhibited an abrupt transition (*Figure 10, A left*) while the majority of rats (12 rats, 80% of all subjects) showed a canonical, smooth transition (*Figure 10, A right*). Thus, even though some rats were changing their behavior abruptly at the onset of extinction, a smooth transition remained the predominant behavior in the sample.

The simulated data showed even a stronger preference for the canonical transition since all 5 simulated subjects (100% of the sample) transitioned smoothly into the extinction phase (*Figure 9, C right*). The lack of variation in the transition type might be due to the difference in the inhibitory and the excitatory rates of learning. Since the inhibitory rate of learning is slower, a greater number of trials will be needed to balance out the excitatory associations which are guiding the preference for the right arm.

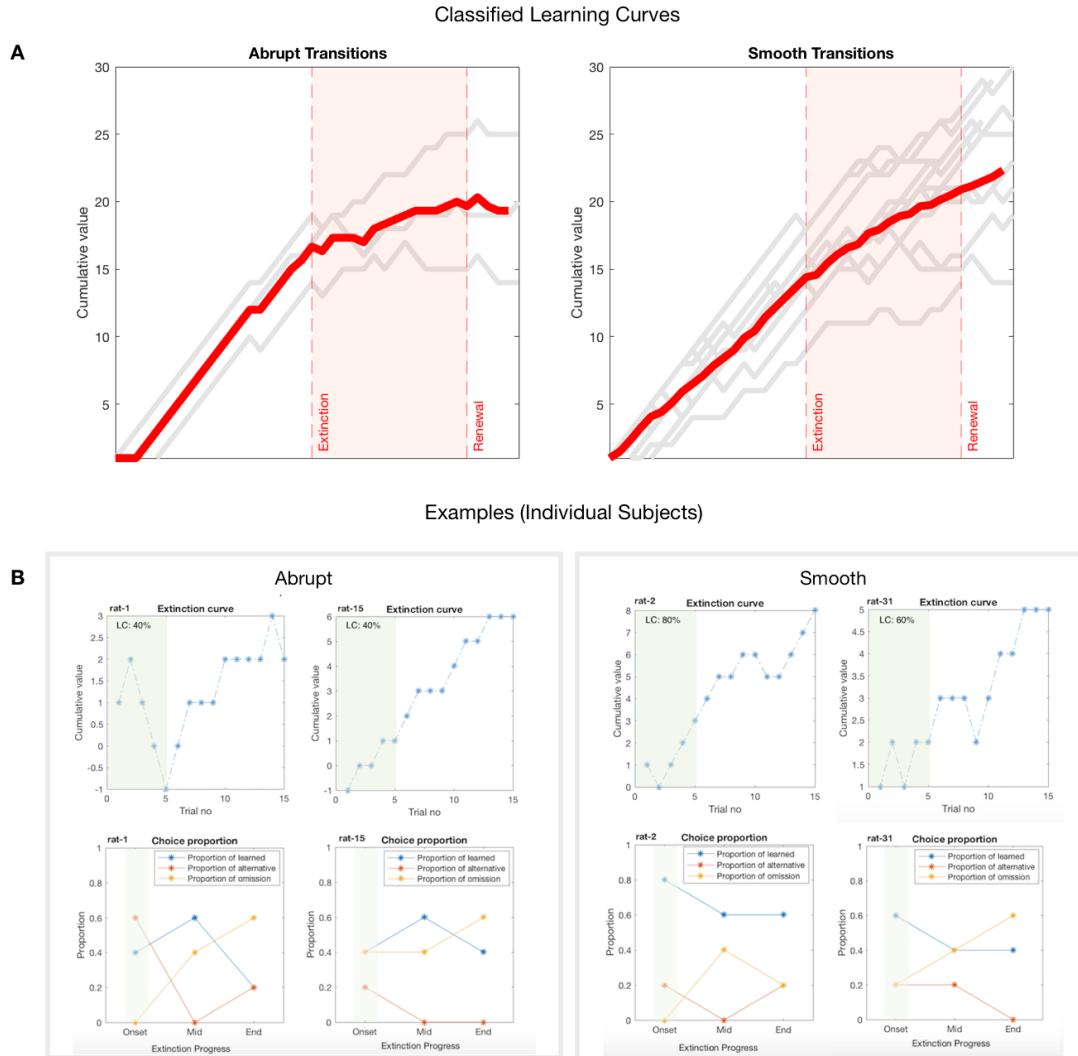


Figure 10: Smooth and abrupt transitions in rats. A) Learning curves classified by transition type. Average curves shown in red. B) Extinction curves and choice preference of subjects exhibiting abrupt (left panel) or smooth (right panel) transitions. With green are marked the trials used to determine the transition type. If the proportion of learned choices (LC) is 60% or higher the transition is considered as smooth.

4.2.2 Omission dominant vs. alternative dominant

As already established on a sample level, the rats are expressing a gradually increasing preference towards omissions compared to the alternative choice during extinction. This observation is not surprising since it is also in line with the canonical view of extinction. However, the simulated rats expressed a strong preference for the alternative choice compared to omissions. This raised the question of whether such a behavior can be observed in real rats. To answer this question, I performed a subject-level analysis of choice variability during extinction.

An alternative dominated behavior was defined by the following two criteria: 1) An increasing proportion of alternative choices in the mid and final stages of extinction (last 10 trials). 2) Higher preference for the alternative choice compared to omissions in the mid and final stages of extinction (last 10 trials).

A total number of 12 rats was analyzed, three were excluded because they showed resistance to extinction. The analysis showed that 11 rats (91.66%) expressed an omission dominant behavior while only 1 rat (8.34%) showed an alternative dominant behavior. A two-sample t-test showed that this rat is exhibiting a significant preference for the alternative choice during the last 10 trials of extinction compared to omissions ($t = 3$, $p = .04773$). At the same time, 4 simulated subjects (80%) showed a similar alternative dominant behavior (*Figure 11*). Again, they all showed a significant preference for the alternative choice over omissions during the last 10 trials of extinction ($t = 60.673$, $p < .001$; $t = 4.6406$, $p = .02172$; $t = 11.946$, $p < .01$; $t = 7.2381$, $p < .01$).

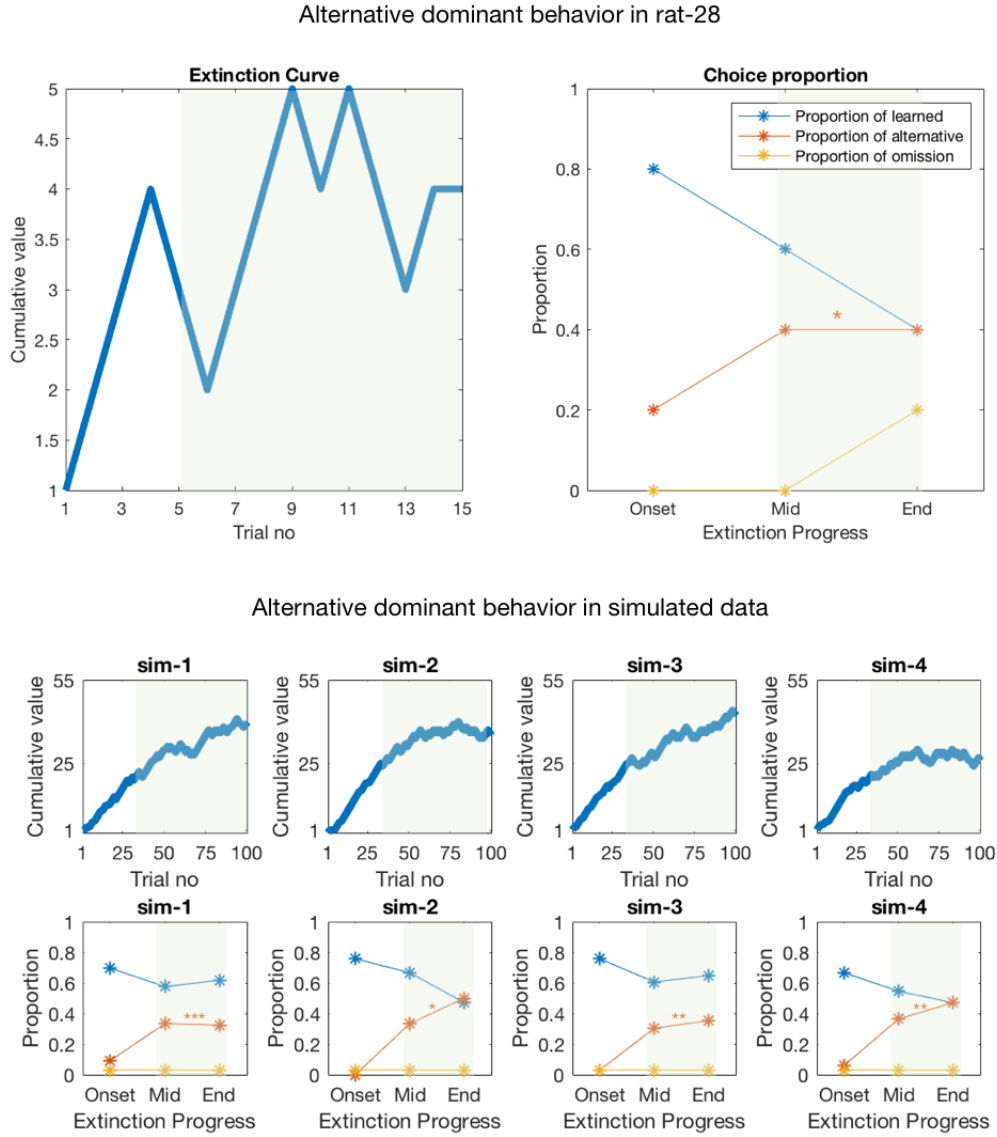


Figure 11: Alternative dominant behavior during extinction in rats (top) and simulated data (bottom). A two-sampled t-test showed that rat-28 exhibits a greater preference for the alternative choice than for omissions ($t = 3$, $p = .04773$). The same dynamics were observed in 4 simulated subjects. Marked in green are the trials used to identify the type of behavioral dominance (last 10 trials of extinction for rats and last 67 for simulated rats)

The fact that there is a rat expressing an alternative dominant behavior shows that this type of behavior is a part of the behavioral repertoire of rats during extinction. Thus, the associative model seems to be at least partially accounting for the variety of behaviors during extinction as far as our data can show. However, this

finding raises the question of why the alternative dominant behavior would occur in first place. By the time of the onset of extinction, animals already know that choosing the alternative arm is not bringing any added value. Furthermore, choosing the alternative arm equals more energy expenditure for no chance of reward. To answer these questions and confirm the existence of alternative dominant behavior in rats, more data and further work will be needed.

4.2.3 Resistance to extinction

Finally, the analysis showed that 3 rats were not expressing a decrease in their preference for learned response during extinction (*Figure 12, top row*). Two of these rats showed a significantly higher preference for the learned choice compared to the non-reinforced choices during extinction: rat-12 ($\chi^2 = 8.5333$, $p = .001744$) and rat-23 ($\chi^2 = 8.5333$, $p = .001744$). The third rat did not show a statistically meaningful preference towards the learned choice but nevertheless, the proportion of learned choices did not decrease at any point of extinction for this subject (*Figure 12, top row, rat-26*). Notice that upon expecting the cumulative learning curves for these rats, one can see that the positive slope formed during acquisition remains overall unchanged in the course of extinction (*Figure 12, top row, left*). This suggests that extinction did not occur for these rats, instead they persisted on going for the right arm even when the rewards ceased.

Similarly, one of the simulated subjects exhibited the same behavior (*Figure 12, bottom row*). Just like rat-12 and rat-23, this subject expressed a significant preference for the learned choice over the non-reinforced choices ($\chi^2 = 27.38$, $p < .001$) during extinction. Again, the learning curve seems to remain with a stable positive slope almost through the entire extinction period. Notice that at the end of the extinction phase the curve is starting to flatten. This suggests that extinction might be taking more trials to occur, possibly because of a strong reinforcement of the excitatory synapses during acquisition. Such a behavior could be a byproduct of the difference between the learning rates of the excitatory and the inhibitory synapses.

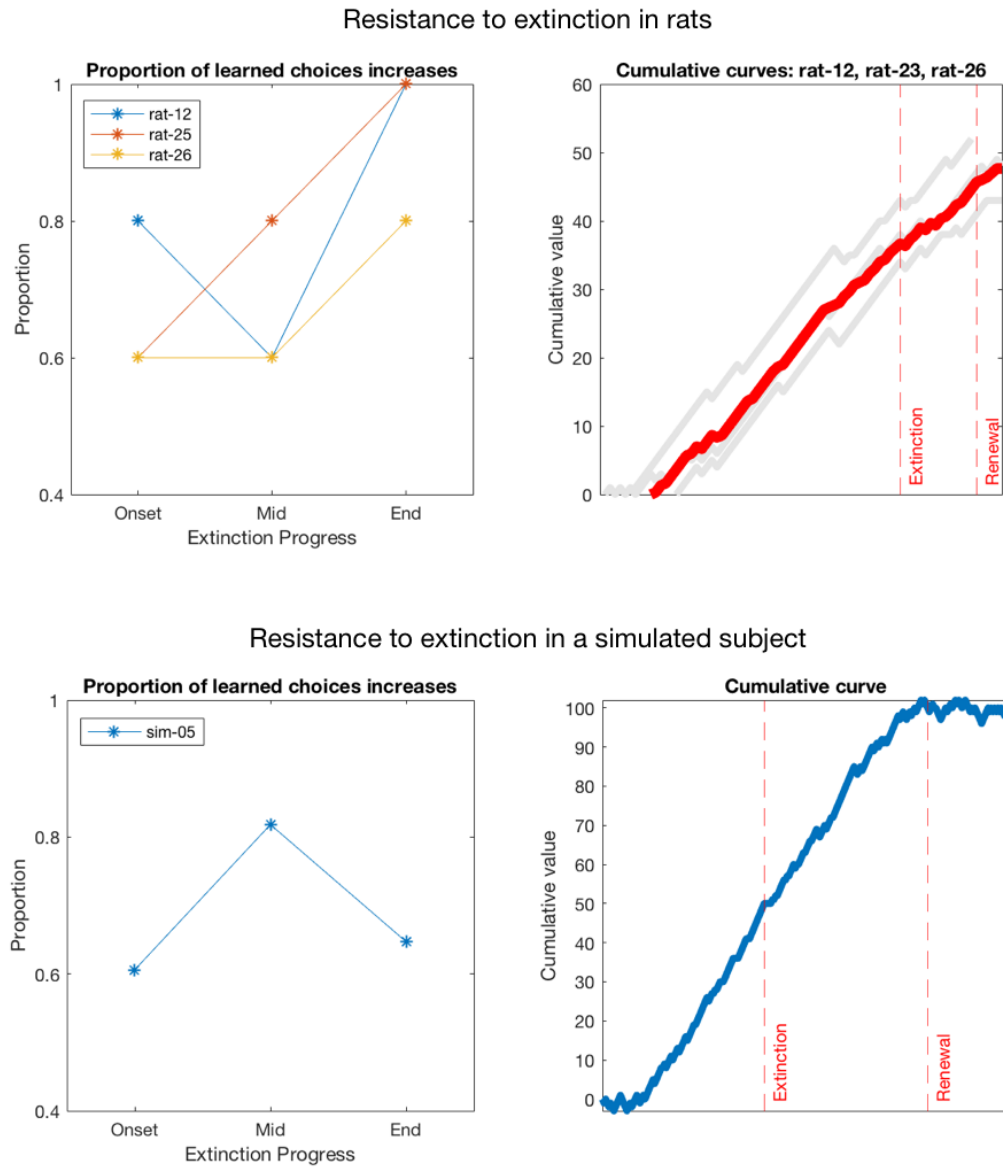


Figure 12: Subjects showing resistance to extinction - top row (rats), bottom row (simulation). Three rats did not show a decrease of their preference for the learned response during extinction. Notice that the positive slope of the average learning curve remains unchanged throughout extinction. The same dynamics are observed in one of the simulated subjects (bottom row).

5 Discussion

I have analyzed the behavior of 16 rats ($N=16$) performing a T-maze experiment designed to investigate context-dependent extinction. A between-phases analysis of

the choice variability uncovered a significant decay of the learned responses during extinction, with a gradually increasing preference towards omissions on a sample level. Additionally, the results showed an absence of the renewal effect on a sample level. To investigate the generative effects of extinction, I further focused on the extinction phase by analyzing data from single animals. The subject level of analysis uncovered a rich variability of behaviors expressed during extinction some of which were previously concealed by the averaged data: 1) 80% of the analyzed subjects showed a smooth transition to extinction while 20% expressed an abrupt change of behavior upon the onset of context B. 2) 91.66% of the subjects showed a dominant preference for omissions as the rate of the learned responses decreased. However, one subject expressed a strong preference for the alternative choice in comparison to omissions. 3) For three subjects (18.75%) extinction did not occur since the rate of the learned response did not decrease at any point during extinction. These subjects did not change their preference towards any of the non-reinforced behaviors during extinction. Consistent with a previous study on pigeons (*Donoso et al., In review*), I demonstrated that a computational model based on an associative learning combined with a winner-takes-all decision-making process, can account for a complex variability of behavior, similar to that observed in the data.

The choice of an associative architecture was driven by the need for in-depth understanding of the basic mechanisms needed to produce extinction and extinction-induced variability. An associative model provides such benefits as transparency, high level of explainability and simplicity. The success of this model to produce some of the extinction-induced effects observed during extinction in rats suggests that an associative network is a sufficient condition for the emergence of these particular phenomena. However, the associative account failed to explain the emergence of omission dominant behavior and thus, it might be the case that to produce the full range of behaviors observed in the data one would need to account for the involvement of higher cognitive functions as well.

5.1 Associative model accounts for the extinction of a previously reinforced response

The sample level analysis of the simulated data showed that the proportion of learned choices during extinction is significantly decreasing just like in the data obtained from rats. Note that the size of the simulated sample might have influenced the power of the statistical findings. Therefore, there is a possibility that a larger sample might produce results as significant as the ones observed in the rat data. The average cumulative curves illustrated in Figure 9E shows that the decrease occurs with a delay compared to the behavioral data. The simulated data preserves the slope of the learning curve at the beginning of the extinction phase. Then, in later trials, the significant decrease of the slope becomes visible. This delayed expression of extinction learning may be accounted for by the choice of an inhibitory learning rate which is slower than the excitatory learning rate. Thus, the model needs more trials to balance out the accumulated excitatory associations with the inhibitory synaptic weights. As a result, it is an expected behavior for the model to continue expressing the reinforced behavior before the excitatory and inhibitory synapses are balanced out and non-reinforced behaviors start to emerge.

5.2 Choice variability on a subject level

Research on extinction is usually conducted by studying population averages. However, the practice of pooling and averaging across subjects has the tendency to conceal the complexity of behavior during a learning task (*Gallistel et al., 2004; Gershman et al., 2010; Larrauri and Schmajuk, 2008; Lissek et al., 2015*). The results presented here are also representative of the existence of a masking effect as a result of pooling data from several subjects. For instance, by examining the average learning curve of the rat sample, one might easily overlook the existence of abrupt transitions since the average curve is perfectly smooth at the onset of extinction (*Figure 9, E*). Similarly, one might assume that all subjects are exhibiting extinction judging by the averaged data (*Figure 9*). A further problem stems from the fact that most computational models assume that the grand averages account for the full

range of behavioral variability represented in the individual subjects (*Gallistel et al., 2004*). This assumption might limit models from giving a comprehensive account of extinction learning, especially with regard to the behavioral variability observed across subjects. Furthermore, such an assumption can potentially downplay the success of some computational models which are able to account for those behaviors masked by the grand average.

5.3 Abrupt transitions in an associative model?

The results from the analysis showed that the computational model failed to produce abrupt transitions at the onset of extinction and thus, the associative account described here fails to explain the emergence of this behavior. However, producing abrupt transitions was set to fail by design in associative models. This is due to the fact that establishing an excitatory or inhibitory association is an accumulative process and therefore, by the onset of extinction, there is already a solid accumulation of excitatory input. At this point, the only way to change the learned behavior would be to balance out the excitatory input with a comparable amount of inhibitory input. However, this can happen only through the same process of accumulation which requires a certain number of trials. Therefore, abrupt transitions cannot be explained by the associative model proposed by *Donoso et al., In review*. This suggests that abrupt transitioning might be a product of a higher level of processing that possibly involves the experience of the animal or even an abstract-rule learning.

5.4 On the emergence of alternative dominant behavior

It has been demonstrated that extinction goes beyond extinguishing a previously reinforced response; it can also drive the emergence of non-reinforced behaviors (*Antonitis, 1951; Grow et al., 2008; K. M. Lattal and Lattal, 2012*). In line with this, the analysis of the rat data showed that the reinforced response was replaced not only by omissions as the canonical view of extinction suggests (*Nevin, 2012; Podlesnik and Shahan, 2009, 2010*) but also in some cases by alternative responses. One rat exhibited such a preference for alternative responses during extinction even

though this particular response was never reinforced. This observation raises the question of why an animal would persist on an unrewarded and purposeless behavior rather than omissions. Explaining this phenomenon would be a difficult task for the framework of reinforcement learning, since this type of behavior does not provide any added value relative to an omission and even requires more energy for execution. On the other hand, the results presented here show that a parsimonious model based on an associative network and a winner-takes-all decision-making process can account for the puzzling emergence of a persistent, alternative dominant behavior. This persistent behavior is likely to be dependent on the fact that the context is treated as another stimulus and therefore, it can establish excitatory and inhibitory associations with specific responses (*Donoso et al., In review*). This way, the change of context generates imbalances in the net inputs to the motor units thus allowing the emergence of persistent alternative choices during the extinction phase. Interestingly, the associative model was not able to produce an omission dominant behavior. This could be due to the limited number of simulated subjects used in the analysis. It might be the case that a higher number of simulations could reveal more types of behavior, including omission dominant behavior. On the other hand, the inability to produce a stable omission dominant behavior might suggest that there is some kind of a higher-cognitive function at play which cannot be accounted for by the model described used here.

5.5 Resistance to extinction

The most important factor found to influence the resistance to extinction is the schedule of reinforcement (*Boren, 1961; Lewis and Duncan, 1956; Shull and Grimes, 2006*). It has been shown that the fewer trials are reinforced during acquisition, the more the subjects are persisting on the learned behavior even in the absence of reinforcers (*Boren, 1961; Lewis and Duncan, 1956; Shull and Grimes, 2006*). The analysis presented in this thesis showed that two rats were showing a statistically significant resistance to extinction, which might be in connection with the reinforcement schedule used in the experiment - multiple acquisition sessions with rewards

being gradually decreased from 100% to 30%. This gradual decrease in the amount of rewarded trials might have contributed to the resistance effect observed in some rats. However, the fact that one of the simulated subjects exhibited a similar resistance cannot be explained by the same account since the provision of reward was continuous in the virtual experimental design. Rather, the emergence of such behavior in the model can be explained by an extreme strengthening of the excitatory synapses during acquisition. This can happen by a long chain of correct trials in which the subject reaches the right arm directly and in a very short time. The less time the subject needs to reach the arm, the less the excitatory synaptic weights will decay. Thus, forming a chain of constant accumulative increase in the excitatory input. Additionally, the inhibitory learning rate is lower than the learning rate of the excitatory synapses, and therefore, in the case of strong excitation it will take a while before the inhibitory weights are large enough to influence the learned behavior. Thus, the model demonstrates the key role of strong excitatory associations in the resistance to extinction phenomenon. This finding might provide the basis of a mechanistic account for some cases of psychological disorders where a conditioned stimuli might acquire a strong emotional valence (appetitive or aversive) and this valence may persist even in the face of repeated and prolonged non-reinforced presentations (*Redish et al., 2007*). Such persistence of behavior is especially relevant in phobias, obsessive compulsive disorder and addiction (*Redish et al., 2007*). Note that the current simulated data offers a limited number of observations and thus, further research with a larger data set is needed to confirm the soundness of these findings.

5.6 Limitations

The current model has several limitations stemming from various sources including variety of motor actions and implementation. The model has three executors and thus, can perform one of three actions - moving forward, left or right. However, the agent is not able to alter its speed or halt its movement at any point of the simulation. This limits the behavioral repertoire of the agent. Another limitation

might be the choice of a T-maze environment since the T-maze is providing a path-forming environment which is naturally navigating the agent towards the arms. The ultimate test for such a model would be a Morris water maze (*Morris, 1981*) where such influences would be eliminated. Last but not least, a natural limitation of the current work is the size of the used data set.

5.7 Future Directions

In order to confirm the validity of the findings reported here, an extended study with a more substantial amount of data should be conducted. A renewal effect was not found on a sample level in both simulated and behavioral data however, the work presented here did not look for renewal effect on subject level. Thus, future work remains to establish whether, and to what extent, the renewal effect was reproduced by the model.

The current implementation of the model used constant but different learning rates for the excitatory and the inhibitory synapses. Future work might continue in direction of comparing the performance of the model under different learning rate configurations. Such an effort might show that different configurations can produce various behavioral patterns observed in animals.

Last but not least, the current work can be extended in a direction of testing the model in different environments. As already mentioned the Morris water maze (*Morris, 1981*) would provide the ultimate testing environment for the model.

6 Conclusion

The current work demonstrated that an associative model with a winner-takes-all decision making process (*Donoso et al., In review*) can, at least partially account, for the richness of the behavior observed during a T-maze experiment, designed to study the extinction in rats. While the model provides an associative account for the eliminative capacity of extinction, it is also capable of reproducing the emergence of some non-reinforced behaviors during extinction as well as the phenomenon of resistance to extinction. Even though, the associative model used here provides a parsimonious, mechanistic explanation to these phenomena, this cannot refute the possibility of having higher cognitive capacities involved as well. Further research and more data are needed to validate and possibly extend the results presented here.

References

- Allen, C., & Bekoff, M. (1999). *Species of mind: The philosophy and biology of cognitive ethology*. MIT Press.
- Antonitis, J. J. (1951). Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of Experimental Psychology*, *42*, 273–281. <https://doi.org/10.1037/h0060407>
- Bernal-Gamboa, R., Nieto, J., & Uengoer, M. (2018). Removing but not adding elements of a context affects generalization of instrumental responses. *Learning Behavior*, *46*(3), 256–264. <https://doi.org/10.3758/s13420-017-0307-9>
- Boren, J. J. (1961). Resistance to extinction as a function of the fixed ratio. *Journal of Experimental Psychology*, *61*(4), 304–308. <https://doi.org/10.1037/h0040208>
- Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–86. [https://doi.org/10.1016/s0006-3223\(02\)01546-9](https://doi.org/10.1016/s0006-3223(02)01546-9)
- Bouton, M. E. (2011). Learning and the persistence of appetite: Extinction and the motivation to eat and overeat. *Physiol Behav.*, *103*(1), 51–8. <https://doi.org/10.1016/j.physbeh.2010.11.025>
- Bouton, M. E. (2019). Extinction of instrumental (operant) learning: Interference, varieties of context, and mechanisms of contextual control. *Psychopharmacology*, *236*(1), 7–19. <https://doi.org/10.1007/s00213-018-5076-4>
- Bouton, M. E., Todd, T. P., Vurbic, D., & Winterbauer, N. E. (2011). Renewal after the extinction of free operant behavior. *Learn Behav*, *39*(1), 57–67. <https://doi.org/10.3758/s13420-011-0018-6>
- Bouton, M. E., Westbrook, R. F., Corcoran, K. A., & Maren, S. (2006). Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biological Psychiatry*, *60*(4), 352–360. <https://doi.org/10.1016/j.biopsych.2005.12.015>
- Byrne, J. H. (2013). Learning and memory: Basic mechanisms. In L. R. Squire, D. Berg, F. E. Bloom, S. du Lac, A. Ghosh, & N. C. Spitzer (Eds.), *Fundamental*

- neuroscience (fourth edition)* (pp. 1009–1027). Academic Press. <https://doi.org/10.1016/B978-0-12-385870-2.00047-0>.
- Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction*, *97*(2), 155–167. <https://doi.org/10.1046/j.1360-0443.2002.00014.x>
- Donoso, J. R., Packheise, J., Pusch, R., Lederer, Z., Walther, T., Uengoer, M., Lachnit, H., Güntürkün, O., & Cheng, S. (In review). Emergence of complex dynamics of choice due to repeated exposures to extinction learning.
- Eisenberg, M., & Dudai, Y. (2004). Reconsolidation of fresh, remote, and extinguished fear memory in medaka: Old fears don't die. *Eur J Neurosci*, 397–403. <https://doi.org/10.1111/j.1460-9568.2004.03818.x>.
- Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, 1481–1489. <https://doi.org/10.1007/978-1-4419-1428-6>
- Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the USA*, *101*(36), 124–131. <https://doi.org/10.1073/pnas.0404965101>
- Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychol Rev*, *117*, 197–209. <https://doi.org/10.1037/a0017808>.
- Gottfried, J. A., & Dolan, R. J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nature Neuroscience*, *7*, 1144–1152. <https://doi.org/10.1038/nn1314>
- Grow, L. L., Kelley, M. E., Roane, H. S., & Shillingsburg, M. A. (2008). Utility of extinction-induced response variability for the selection of mands. *Journal of applied behavior analysis*, *41*(1), 15–24. <https://doi.org/10.1901/jaba.2008.41-15>
- Hartley, C. A., & Phelps, E. A. (2012). Extinction learning. *See N.M. (eds) Encyclopedia of the Sciences of Learning*. <https://doi.org/10.1007/978-1-4419-1428-6>

- Larrauri, J. A., & Schmajuk, N. A. (2008). Attentional, associative, and configural mechanisms in extinction. *Psychol Rev*, *115*, 640–676. <https://doi.org/10.1037/0033-295X.115.3.640>
- Lattal, K. M., & Lattal, K. A. (2012). Facets of pavlovian and operant extinction. *Behav Processes*, *90*(1), 1–8. <https://doi.org/10.1016/j.beproc.2012.03.009>
- Lattal, K. A., Peter, C. S., & Escobar, R. (2013). Operant extinction: Elimination and generation of behavior. *APA handbook of behavior analysis*, *2*, 77–107. <https://doi.org/10.1037/13938-004>
- Lengersdorf, D., Stüttgen, M. C., Uengoer, M., & Güntürkün, O. (2014). Transient inactivation of the pigeon hippocampus or the nidopallium caudolaterale during extinction learning impairs extinction retrieval in an appetitive conditioning paradigm. *Behavioral Brain Research*, *265*, 93–100. <https://doi.org/10.1016/j.bbr.2014.02.025>
- Leslie, J. C., Shaw, D., Gregg, G., McCormick, N., Reynolds, D. S., & Dawson, G. R. (2005). Effects of reinforcement schedule on facilitation of operant extinction by chlordiazepoxide. *Journal of the Experimental Analysis of Behavior*, *84*(3), 327–338. <https://doi.org/10.1901/jeab.2005.71-04>
- Lewis, D. J., & Duncan, C. P. (1956). Effect of different percentages of money reward on extinction of a lever-pulling response. *Journal of Experimental Psychology*, *52*, 23–27. <https://doi.org/10.1037/h0043090>
- Lissek, S., Glaubitz, B., Güntürkün, O., & Tegenthoff, M. (2015). Noradrenergic stimulation modulates activation of extinction-related brain regions and enhances contextual extinction learning without affecting renewal. *Front. Behav. Neurosci.*, *9*(34). <https://doi.org/10.3389/fnbeh.2015.00034>
- Llaving, G. A., Hagopian, L. P., Long, E. S., & O'Connor, J. (2004). Response-class hierarchies and resurgence of severe problem behavior. *The Psychological Record*, *54*, 621–634. <https://doi.org/10.1037/13938-004>
- Méndez-Couz, M., Becker, J. M., & Manahan-Vaughan, D. (2019). Functional compartmentalization of the contribution of hippocampal subfields to context-

- dependent extinction learning. *Frontiers in Behavioral Neuroscience*. <https://doi.org/10.3389/fnbeh.2019.00256>
- Morris, R. G. (1981). Spatial localization does not require the presence of local cues. *Journal of Experimental Psychology*, *12*(2), 239–260. [https://doi.org/10.1016/0023-9690\(81\)90020-5](https://doi.org/10.1016/0023-9690(81)90020-5)
- Neuringer, A., & Kornell, N. (2001). Stability and variability in extinction. *Journal of Experimental Psychology Animal Behavior Processes*, *27*, 79–94. <https://doi.org/10.1037/0097-7403.27.1.79>
- Nevin, J. A. (2012). Resistance to extinction and behavioral momentum. *Behavioral Processes*, *90*(1), 89–97. <https://doi.org/10.1016/j.beproc.2012.02.006>
- Nieto, J., Uengoer, M., & Bernal-Gamboa, R. (2017). A reminder of extinction reduces relapse in an animal model of voluntary behavior. *Learning Memory*, *24*(2), 76–80. <https://doi.org/10.1101/lm.044495.116>
- Osvath, M., & Osvath, H. (2008). Chimpanzee (*pan troglodytes*) and orangutan (*pongo abelii*) forethought: Self-control and pre-experience in the face of future tool use. *Animal Cognition*, *11*(4), 661–674. <https://doi.org/10.1007/s10071-008-0157-0>
- Pearce, J. M., & Hall, G. (1980). A model for pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. <https://doi.org/10.1037/0033-295X.87.6.532>
- Podlesnik, C. A., & Shahan, T. A. (2009). Behavioral momentum and relapse of extinguished operant responding. *Learn Behav*, *37*(4), 357–364. <https://doi.org/10.3758/LB.37.4.357>
- Podlesnik, C. A., & Shahan, T. A. (2010). Extinction, relapse, and behavioral momentum. *Behavioral Processes*, *84*(1), 400–411. <https://doi.org/10.1016/j.beproc.2010.02.001>
- Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychol Rev*, *114*, 784–805. <https://doi.org/10.1037/0033-295X.114.3.784>

- Rescorla, R., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 64–99.
- Shull, R. L., & Grimes, J. A. (2006). Resistance to extinction following variable-interval reinforcement: Reinforcer rate and amount. *J Exp Anal Behav*, 85(1), 23–29. <https://doi.org/10.1901/jeab.2006.119-04>
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century.
- Stollhoffa, N., Menzel, R., & Eisenhardt, D. (2005). Spontaneous recovery from extinction depends on the reconsolidation of the acquisition memory in an appetitive learning paradigm in the honeybee (*apis mellifera*). *Journal of Neuroscience*, 25(18), 4485–4492. <https://doi.org/10.1523/JNEUROSCI.0117-05.2005>
- Suddendorf, T. (2006). Behavior. foresight and evolution of the human mind. *Science*, 312(5776), 1006–1007. <https://doi.org/10.1126/science.1129217>
- Todd, T. P. (2013). Mechanisms of renewal after the extinction of instrumental behavior. *Journal of Experimental Psychology. Animal Behavior Processes*, 39(3), 193–207. <https://doi.org/10.1037/a0032236>
- Todd, T. P., Vurbic, D., & Bouton, M. E. (2014). Mechanisms of renewal after the extinction of discriminated operant behavior. *J Exp Psychol Anim Learn Cogn*, 40(3), 355–368. <https://doi.org/10.1037/xan0000021>
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373. <https://doi.org/10.1037/h0020071>