The Hebrew University of Jerusalem
The Faculty of Humanities
Cognitive and Brain Sciences Department

Probing learning from primary aversive stimulations over multiple days

By: Aharon Ran Cohen

Supervised by: Dr Eran Eldar

Jerusalem

September 2021

Acknowledgments:

This work was done under the guidance of Dr Eran Eldar. First and foremost, I would like to express my sincere gratitude and appreciation to my supervisor and my lab P.I. Dr Eran Eldar. Thank you for the patience and guidance through the whole process of designing and developing the experiment, collecting, analyzing data, and eventually writing this thesis. I feel humble and grateful that I had you as my instructor and was lucky to be part of your new lab. I also want to thank my lab colleagues: Levi Solomyak for allowing me to use the data he collected in his experiment on reward and the help in analyzing data, Lena Gorina and Paul Sharp for helping me in the writing process, Arne Gouwy and Alon Erdman for helping me in the data analysis. I could not have made it this far without all your graceful help. I would like to dedicate this thesis to my mother Shoshana that passed away recently, she was my spring of life, motivation and strength and I still receive power from her memory.

Table of contents:

Abstract

Studies suggest that people's sensitivity to reward may change from day to day, and such changes could impact mood and mental disorders. However, it has yet to be investigated whether changes in punishment sensitivity could play a similarly important role. Of particular importance may be the sensitivity to primary aversive stimulation such as pain or loud noises (as opposed to secondary, conditioned punishment like monetary loss) since this is the most basic and natural form of punishment in humans and animals alike. However, experimentally testing learning from primary aversive stimulation is difficult even in a single day lab experiment, let alone in an experiment that aims to assess and re-assess such learning over multiple days, during which people may habituate to the experimental punishment. The goal of this project was to develop an experimental paradigm that can probe learning from primary punishment over multiple days. For this purpose, we developed a phone-based punishment learning task wherein punishment is administered via headphones. The task allows quantifying both sensitivity and responsivity to punishment, with the former manifesting in the degree to which subjects tend to avoid experimental stimuli previously associated with punishment, and the latter in heartrate responses to punishment as measured through a wearable sensor. The experiment was conducted outside of the laboratory and was monitored remotely by the experimenter. 6 subjects performed an experimental session on their mobile phone twice a day, for 12 days. For comparison, another group of 6 subjects performed the experiment with secondary punishments, which consisted of monetary losses. On each session, subjects had to learn via trial-and-error the probabilities with which choices of different images (textured circles) led to punishment. Primary punishment consisted of a loud noise played through headphones put on during the sessions. To ensure that participants received the

punishments properly, a listening task was administered randomly between trials. On each session subjects learned about new stimuli. To assess how well learned information was remembered, on some trials subjects were asked to choose between images they had already learned about on previous days. To evaluate whether the experiment successfully probed learning from punishment over the 12 days of the experiment, we examined the degree to which subjects successfully avoided images that were associated with larger probability of punishment. As an implicit and more direct measure of the emotional impact of the punishment, we also measured subjects' heartrate responses to outcomes in the learning task. Subjects' choices showed that they successfully avoided stimuli that were more likely to be associated with loud noise, even a few days after learning, which indicates successful and enduring conditioning. Moreover, almost no decline in choice accuracy was observed during the progress of the experiment, which suggests low habituation to the experimental punishment. Analysis of subjects' heartrate showed consistent responsivity to punishment outcomes in both study groups. The small sample size of this preliminary study did not allow conclusive comparison between the primary and secondary punishment groups. However, visual inspection suggested that choice sensitivity and heartrate responsivity to loud noise were at least as strong and enduring as sensitivity and responsivity to monetary loss.

These results confirm that the experimental procedure that we developed constitutes a valid and effective paradigm for investigating sensitivity and responsivity to primary punishment outside of the lab over multiple days. Furthermore, using this paradigm, future studies could probe the relationship between temporal mood fluctuation and aversive learning mechanisms.

Introduction

Conceptually, punishment differs from reinforcement in the direction of change in the probability of choosing a behavior. While Reward increases the likelihood of a behavior to occur, punishment decreases it (Azrin and Holtz 1966; Johnston, J. M. 1972). Naively, we might say that punishment is the negative equivalent of reward, therefore they should have symmetrical effects on people.

Nevertheless, it seems like aversive reinforcement is unique in its mechanism and effects, as we can observe in the famous phenomenon referred to as loss-aversion bias in which people prefer to avoid punishment over gaining an equivalent reward (Kahneman & Tversky 1979). In other words, the absolute value of a coin lost is perceived as greater than a coin gained (Rasmussen & Neweland, 2008).

Many studies have demonstrated the differences between reward and punishment (Jean-Richard-Dit-Bressel et al 2018), which involve many aspects such as their effect on learning and behavior, different neuronal systems that underlie aversive and appetitive learning, and clinical psychiatric disorders that may be implicated by them (Schultz 2007; Cools et al 2008; Costafreda et al 2008; Galea et al 2015; Eldar et al. 2016a; Steel et al 2016; Toshikazu et al. 2018; Jean-Richard-Dit-Bressel et al 2018; Aylward et al 2019; Michely et al 2020).

Although punishment learning mechanism has an important role in learning processes and extensive potential implications for psychiatric disorders (Jean-Richard-Dit-Bressel et al. 2018; Wise & Dolan 2020), we still know little about its neuronal and computational processing, and its precise influence on human behavior (Jean-Richard-Dit-Bressel et al. 2018; Wise & Dolan 2020). Thus, the investigation of punishment effects on behavior and learning, and its implication on mental disorders is crucial and promising.

The most fundamental form of punishment for humans and animals alike is primary aversive stimulus. It is naturally perceived to have negative value without any conditioning. An example which is frequently used in research is the delivery of an electric shock. The recipient is immediately repelled even if it is the first time it ever encountered this stimulus. In contrast, a secondary punishment is aversive only after having been conditioned with a primary aversive stimulus. Monetary loss is a widespread example of a secondary punishment frequently used in human experiments (Franzoi, S. L. 2015; Jean-Richard-Dit-Bressel et al 2018).

Although studies with primary punishments are common, there are not many longitudinal experiments probing their repeated stimulation over time. Moreover, although monetary loss can be highly aversive to humans as electric shock (Delgado et al 2006), it is still dependent on reinforcement learning related systems rather than on biological fear (Delgado et al 2011; Palminteri et al 2015) and therefore might have different effects than primary punishment. Thus we thought it was important to be able to compare both types of punishment. In order to create a conditioned response for each one of them separately (Delgado et al 2011), we used a between group experimental design.

Individuals often differ in their perception of reward and punishment (Carver and White 1994). Studies suggest that this receptivity to affective outcomes is implicated in different forms of psychopathology (Johnson et al 2003).

Eldar et al (2018) divides this receptivity into two parts: sensitivity and responsivity. They define Sensitivity to punishment as the transformation from objective value into subjective utility and responsivity as the modulation of the magnitude of physiological responses to the stimulation (which may reflect how much attention is given to the dimension of punishment). They found that responsivity to reward prediction errors changes from day to day and these changes predicts subsequent mood fluctuations.

Although responsivity to punishment may also fluctuate over days, we do not currently know how its dynamics resemble or differ from responsivity to reward. Moreover, studies have also suggested that learning in a short timescale result in a learning process distinct from the learning process in a long timescale learning (ligaya K. 2016; Wimmer & Poldrack 2017; Eldar et al 2018). Therefore it is highly informative to investigate primary punishment both in short and long period of time.

In order to explore these dynamics we sought to develop a platform that enables to investigate how punishment sensitivity and responsivity varies by day as manifested in the degree to which subjects tried to avoid images associated with punishments and their heartrate responses to the aversive stimulus, respectively.

Although using electric shock as punishment is common in regular studies, using it in a remote longitudinal experiment has two downsides: the first is the lack of experimental control (Reips 2000), meaning that it is hard to operate it at the home of the subject and we have no good way to ensure that the procedure of delivering the electric shock would not be compromised. The second concern is that subjects will habituate to the stimulus with time (Mcsweeney & Roll 1998). In order to address these issues, we used a different kind of primary aversive stimulus, loud white noise. A loud noise stimulus can resolve the first concern because it can be delivered and controlled via headphones. Additionally, Sperl et al (2016) presented a comparison between electric shock and loud white noise as an unconditioned stimulus and concluded that the latter had greater valence of unpleasantness, less extinction of conditioned response, and a better recall of the conditioned response after 24h.

To carry out a longitudinal experiment on learning from primary aversive punishment we used a novel mobile platform that can be used by subjects outside of the laboratory. Online experiments using auditory aversive conditioning has proven reliable for studies (Seow & Hauser 2021) and has many advantages (although also a few challenges - see Reips 2000).

Subjects' data was uploaded and stored in a secured location every few hours. To make sure the data we collect are indeed reliable, the data were reviewed by the experimenter regularly and a few reliability tests were made such as checking the reaction time was standard, side bias, performance, and tasks time schedules. We will elaborate on these measurements in the methods section.

The aim of this study was to validate that the design we developed is reliable to investigate learning from punishment over multiple days, meaning that subjects learn the probabilities of different stimuli throughout the 12 days of the experiment. To do that, we tracked the daily performance and heartrate responses of subjects to evaluate the effects of the aversive stimulus. The results also gave us a first glimpse about the differences between a primary punisher and a secondary punisher, and between them and reward. Given the scope of the study, our sample size was too small to draw conclusions about group differences. However, this novel experiment will hopefully provide a good basic design for future studies to investigate the variability of sensitivity and responsivity to punishment over multiple days, and its involvement in behavior, mood fluctuations and psychiatric disorders such as depression and anxiety.

Methods

Participants.

19 healthy volunteers (mean-age: 25.47±3.53; range 20 – 33 years old; 12 female) were recruited through social media advertisements and participated in a 12-day experiment. Subjects were divided randomly to the white noise punishment group (9 subjects) and the monetary loss group (10 subjects). They conducted a trial-and-error game through an android app installed on their phones. Subjects who did not have phones that are compatible with the experiment application were provided with a phone owned by the lab. All subjects underwent a screening process to exclude any motor, hearing or vision impairments, history of psychiatric disorders and use of psychiatric medications and drugs. The experiment was approved by the Israeli Helsinki committee and all subjects signed a consent form accordingly. All subjects were paid by the hour (40 shekels). Additionally, each group received additional

payment at the end of the experiment. For the monetary loss group, subjects received the amount left in their experimental bank and for the white noise group, subjects got a finishing bonus (300 shekels).

Exclusion. Seven participants were excluded from the study. Three participants were excluded due to technical problems. Another three because of insufficient compliance to the schedule of tasks, meaning that there were more than two consecutive days without performing the tasks. One participant was excluded due to repeated bad performance (choices success rate was at chance level) from the beginning of the experiment, and therefore we believe did not understand the probabilistic nature of the task. After exclusion, 12 volunteers' (mean-age: 25.58±2.96; range 20 – 31 years old; 8 female) data were taken for the analysis (6 in each of the study groups).

Comparison with reward learning data. Results of a parallel experiment on reward conducted in the lab was added to the analysis for comparison with the punishment learning groups. 19 participants performed a trial-and-error learning game in which they needed to choose between different images with different probabilities of delivering reward. The experiment continued for 12 days, and the design was symmetrical to our punishment experiment.

Mobile application. for the longitudinal learning game we used an Android application that was developed for experimental purposes by the lab P.I. Dr Eran Eldar. It was programmed in Java using the Android Studio programing environment

(Google, Mountain View, CA, version 3.5.2). The original game used rewards as incentive, and we adjusted it for punishment. All the data collected by the app was stored locally on the phones as SQLite databases and was uploaded every few hours to a Gmail cloud storage space, designated for it.

For generating the white noise and cutting sound duration (0.5 seconds) we used the sound editor software Audacity (http://audacityteam.org/) version 2.4.2.

For the "Wheel of fortune", that was used to create the initial amount of money for subjects in the monetary loss group, we used and edited the example code from https://github.com/zarocknz/javascript-winwheel (Copyright (c) 2016 Douglas McKechie).

Stimuli. The punishment learning game contained 64 images (round fractals with styled backgrounds). Each stimulus has its own probability for delivering punishment. Unknown to the subjects, the probabilities were set to be associated with one of the four probabilities [0, .33, .66, 1] where 0 represents no chance of getting punished and 1 means that punishment will be delivered 100% of the time.

In a preliminary lab meeting subjects played 13 training blocks of overall 388 trials (that were excluded from analysis). Outside the lab subjects played 44 blocks of overall 3088 trials (4 blocks per day). Altogether, the experiment consisted of 3476 trials, in which 2211 of them were learning trials (with feedback) and 1265 were testing trials (without feedback).

Punishments were either the loss of a coin (worth approximately 0.2 shekels) for the monetary-loss group or the delivery of a loud white noise that lasted 0.5 s (mean dB=92.68±0.69; range 92 – 93.6 dB) for the noise group.

Procedure.

Subjects were randomly assigned (before screening interview) into the two study groups of two types of punishments:

1. Loud white noise (mean age: 25.66±2.58; 4 females)
2. Monetary loss (mean age: 25.5±3.56; 4 females)

Each participant went through a screening interview by phone regarding the experiment criteria and if meeting the criteria the subject was invited to a preliminary lab meeting.

Preliminary lab meeting. After signing a consent form, the experimenter explained to the subjects the task and its schedule throughout the following 11 days of the experiment.

Wheel of fortune. For the monetary-loss group, due to the possibility that since the reduction of money with each loss of a coin was meager (0.2 shekels, ~ 0.06 dollars) we might lose the loss-aversion effect and the motivation to avoid punishments would be sparse, although throughout the experiment the sum of potential loss is great. Nevertheless, participants might not consider the "big picture" and disregard the loss on a single trial. Therefore we wanted to magnify the effect of money loss on subjects by creating a sense of endowment to make the subject value more the initial sum of money. We also wanted to create a sense of agency, to increase the engagement of subjects in the task (Taub et al. 2020). The sense of endowment and agency over the money increases the aversiveness of the monetary loss and magnifies the effect on the subject (Delgado et al 2011).

Therefore, instead of just providing subjects with an initial amount of money, we conducted a preliminary draw beforehand to create an initial winning from which the subject can further lose money. This design was similar to other monetary loss studies such as Delgado et al (2011) and Steel et al (2016; 2020). The task included spinning a Wheel of Fortune (WoF) in a computer program with different amounts of optional gains (400-1000 shekels). Unknown to the participants, the win was fixed on 600 shekels (equivalent to ~182 dollars) in order to equalize the amount of money for all subjects. This was their bank of money from which they could lose or keep during the following sessions, and therefore should do their best to avoid punishments.

Colors task. The loud white noise group was provided with headphones from the lab (Miracase MBTO106). To make sure that subjects are wearing the headphones throughout the games and listening to the delivery of punishments, we added the "colors task" between trials. Randomly, every few trials appeared a screen with six rectangles of different colors. Then, one of the color names was randomly played, sometimes to the left ear and sometimes to the right ear, and the subject needed to press the matching color-rectangle. Also, on testing trials, where punishments were given only after 10 trials and therefore a sequence of punishments might be delivered, the colors-task appeared randomly between noises.

App installation. After that, for both groups, we installed the application of the experiment on the subject's phone (or, if it was not compatible, the phone was delivered by the lab – "Redmi Note 9 Pro"). Next, the volume calibration was conducted for the white noise group.

Volume Calibration. For the White Noise group, the intensity of the sound was calibrated to set the system's volume to the range predetermined by the experimenter (92-95 dB). To check the sound volume we used a sound meter monitor "UT353 Mini Sound Level Meter". As in Sperl et al (2016) study, the range was between 92-95 dB for every subject in the noise group (mean dB=92.68±0.69; range 92 – 93.6 dB). The variability in the dB volume is due to the different phone systems and from the noise intensity perceived by each subject. If the noise was unbearably intense for the subject, we lowered the volume but maintained a minimum of 92 dBs. In addition to the white noise volume calibration, we also needed to calibrate the volume of the " colors task". We wanted to keep volume to the minimum necessary for the subject to hear the names of the colors but not higher so that they could not hear it without wearing the headphones on top of their ears. For each subject, while wearing the headphones, we played the names of colors in the lowest volume and increased it until the subject said she hears them clearly.

After that, we instructed the subjects on how to put the wearable heartrate sensor on their body, a "Polar H10 " device monitor that processes heartbeat rates. Subjects had to wear the sensor whenever they played the experiment game, including a five-minute baseline measurement at rest before starting the games.

Next, we explained the tasks to the subjects in detail, including the structure of the game.

App schedule. The app asks the subjects to perform the task in a predetermined schedule. The daily schedule included a morning session of 2 consecutive blocks of trials and an evening session of 2 consecutive blocks. Each block had 72 trials that included a set of 3 images with different probabilities of delivering punishment. The images were paired with each other on every trial. Some sets of images had the probabilities [0 .33 .66] and others had [.33 .66 1]. The probability of a particular stimulus was constant through the whole experiment. Every few blocks a new set of images was presented but sometimes a set of images already learned before was presented. As mentioned before, on learning trials, the outcome of a choice was revealed to the subjects. After 2 blocks of repeatedly presented with a set of images and their revealed outcomes (learning trials; 64 times for each stimulus on average), those images appeared on consecutive blocks between learning trials, without revealing their outcomes (testing trials). Every third trial in those blocks, appeared a testing trial for images already learned while the rest of the trials on that block were learning trials for other images. This routine was kept for 10 Consecutive days, followed by a rest day and finally the last experiment day, meant as a summery test for subjects about all the stimuli they learned about. On that last day, subjects performed two more sessions (morning & evening) where only testing trials were given and images from different blocks were mixed and paired with each other.

Before each session (2 consecutive blocks) subjects had to put on the heartrate sensor and complete 5 minutes measurement of rest. Each session took approximately 20-25 minutes overall.

Delayed games. If subjects did not execute a block or a session until it was too close in time (4 hours) to the scheduled time of the next session, the app postponed it to that next scheduled time. By that, all sessions were postponed and the experiment prolonged.

The game. Each trial started with subjects holding the phone horizontally and placing their thumbs on each side of the screen. Only then, two images were presented to them, one to the left of the screen and the other to the right. To avoid punishment, subjects had to remove their finger from the image they thought will most probably deliver punishment. Avoiding one stimulus meant that the stimulus on the other side of the screen got chosen. On the learning trials, ~0.5s after the choice was made, visual feedback appeared for 1s for the monetary loss group and 1.5s for the white noise group (due to the 0.5s of white noise played) indicating the outcome of subject's choice (punishment or no-punishment) and ~2.5s afterwards the next trial started, again with two thumbs on the screen. If the subject did not choose within 4 seconds an alert message appeared asserting that the subject was too slow.

Importantly, after they choose, subjects can see the consequence of their choice and therefore can learn, by trial-and-error the stimulus probability of producing punishment. A punishment outcome was seen as a red arrow pointing down inside a circle with black background (that in the noise group was paired with a loud white noise) and no-punishment outcome was seen as a blank black circle. It was emphasized to the subjects, that the game is probabilistic and therefore a "bad" stimulus can sometimes not deliver punishment (although it is still the worst choice) and a "good" stimulus can sometimes deliver punishment (although it is still the best choice).

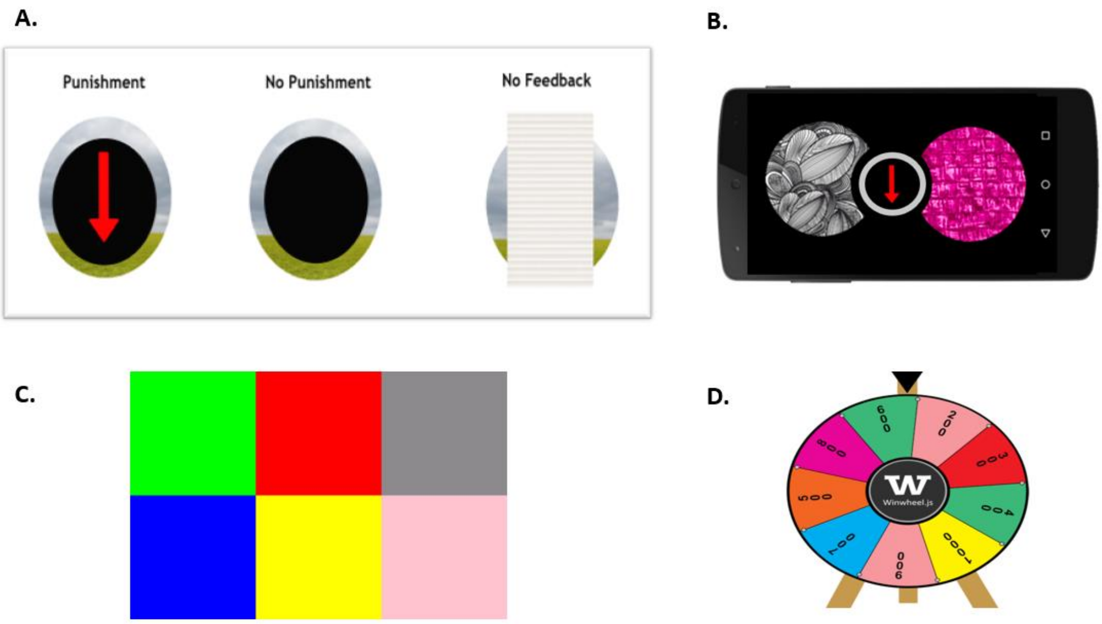

Figure 1:

A. **Possible Feedbacks.** The two outcomes that the subject receive in learning trials after choice was made. a red arrow pointing down inside a circle with black background indicates punishment and a blank black circle for no-punishment. On testing trials the outcome was covered by white curtain.

B. **Two stimuli each trial.** On each trial subject held the phone horizontally and chose between 2 circles with styled backgrounds.

C. **The color attention task.** From time to time, in the white noise group while playing the game, an attention task appeared where subjects needed to choose the colored rectangle that matches the color asserted in the earphones.

D. **Wheel of Fortune.** Before the beginning of the experiment, in the monetary loss group, subject spined a wheel of fortune to create an experimental bank.

After enough times that a set of stimuli was repeated, and subjects had the chance to learn the probabilities of a set of stimuli, a curtain covered the outcomes to conceal them from the subjects. This means that the learning trials are over for these set of stimuli, and the testing trials begin. This way we can test how much those stimuli that were learned were retained in memory by subjects. Although outcomes of choices were not presented, they were still stored, and their sum was presented to the subject after every 10 trials. In the money loss group a message appeared saying that the subject lost an amount of X coins in the last 10 trials, and in the white noise group a message saying that outcomes were punishment X times in last 10 trials. In addition to the message, in the white noise group, a loud white noise was delivered as many times as the hidden outcomes conveyed punishment. This way we kept the incentive for choosing the better stimulus without revealing the individual outcomes.

Accuracy. On each trial when a subject avoided the stimulus that had more chance of producing punishment between the two stimuli, we considered it an accurate choice, even though the stimulus that was chosen eventually delivered punishment, and vice versa. Accuracy rate was the number of correct choices out of all trials.

Monitoring subjects remotely. To make sure that subjects are following the schedule and the instructions they got and that the data we get is reliable, we checked frequently (at least once a day) the data that was uploaded to the cloud storage. Besides the color attention task that was already mentioned above we looked at other parameters of subjects' performance:

a. Task execution – we looked at the times that the tasks were executed. If subjects didn't complete the task, we noted them that and asked for improvement. If this behavior maintained for more than 2 consecutive days, we ended the experiment, and the subject was excluded.
b. Reaction time (Rt) – a very low Rt (we flagged it below 1400 ms following results from past experiments conducted in the lab) may indicate subjects are not paying attention to stimuli and are choosing randomly. If subjects averaged Rt was flagged, we did the same procedure as in a'.

c. Side bias – side bias is when subjects are inclined to choose one side over the other. Because stimuli are distributed randomly between sides on each trial, no side bias should occur and the probability of choosing one of the sides should be 0.5. If subjects were biased 5% more towards one side, we did the same procedure as in a'.

d. Accuracy rate – low accuracy rate in the task may reflect that subjects are not engaged in it. Therefore, if accuracy rate was under 70%, we did the same procedure as in a'.

All these tasks together confirmed the reliability and integrity of task execution. Subjects who finished the 12 days of the experiment showed almost no violations of these criteria on average during the days of the experiment.

Analysis.

Frameworks. We analyzed the behavioral data using RStudio (Version 1.1456 - © 2009-2018 RStudio, Inc.; R version 4.0.3), and we processed the physiological data (HR) using MATLAB (version 9.5.0.944444 (R2018b)).

Binomial test. To analyze how significant our results were we used the binomial test. The binomial test is commonly used to test hypotheses about the probability of success in a sequence of a dichotomous variable observations. It evaluates the statistical significance of deviances of the sample data from a theoretically expected distribution (binomial). Given the small sample size (6 subjects per group), we reasoned that an effect would be significant only if it categorically appears in 6 out of

6 subjects. The probability of such an outcome can be evaluated using a binomial test where the null hypothesis is that the probability of observing the effect in a given subject is 50%:

H_0: p = 0.5

H_1: p > 0.5

We needed to calculate the probability to get the number of successes we got (or more extreme number) according to the null hypothesis using the binomial distribution:

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

This gives us the p value of the null hypotheses that determines whether we can reject it (Howell 2006).

Results[1]

General Performance. All groups performed well with more than 80% accuracy, without large differences between groups (White Noise (WN): 0.826± 0.9, Monetary Loss (ML): 0.817±0.4, Reward (R): 0.831±0.6). All groups performed above chance level (one-tailed binomial test: WN: p=.0206, ML: p=.0206, R: p=.000018). With respect to reaction time though, we observed a clear trend for faster responses in the WN

[1] In figures 2-5, in order to show all data points and lines clearly the Y axis scale does not start from zero. Therefore the reader should consider the proportionality of the relationship between data points and between lines.

group compared to the ML group, but testing more subjects is necessary to confirm this difference (WN: 1628±63ms, ML: 1867±202ms, R: 1707±130ms; Figure 2).

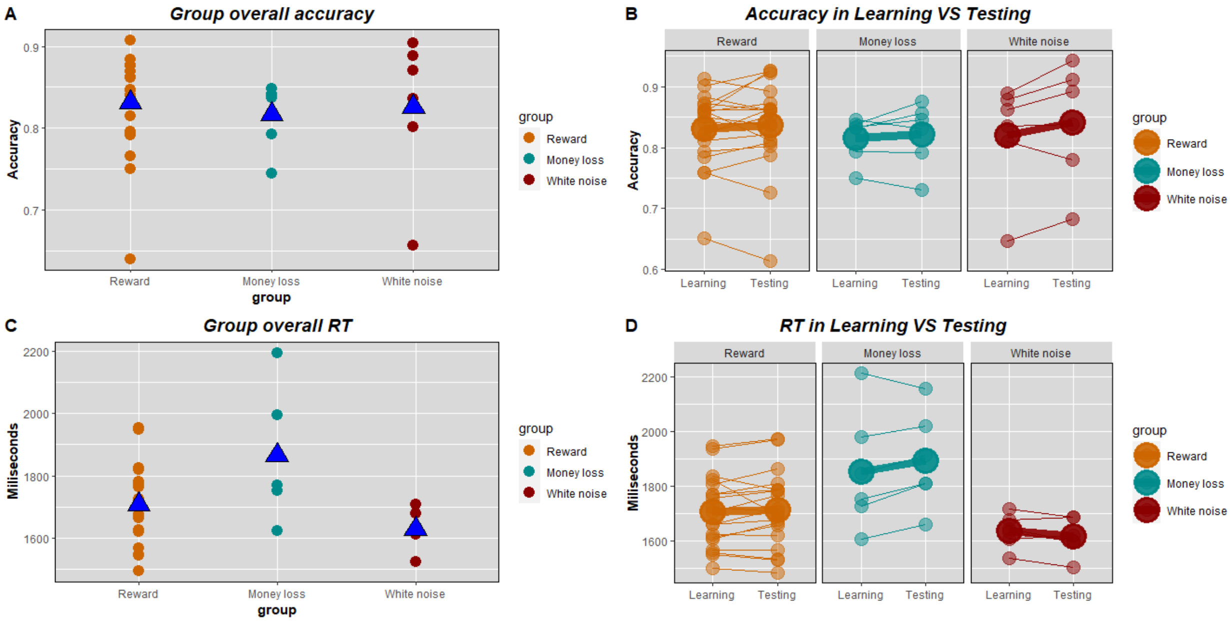

Figure 2:

a. Overall accuracy rate. The rate of correct choices according to designated probabilities from all trials. Y axis is the accuracy rate and X axis is the group condition. Dotted points are participants' averages and blue triangles are group means.

b. Accuracy on learning trials vs testing trials. Y axis is the accuracy rate and X axis is the learning/testing trials condition. Thicker lines and thicker dots represent group average.

c. Mean reaction time. Reaction time is the amount of time between the appearance of stimuli and the choice time. Y axis is the reaction time in milliseconds and X axis is the group condition. Dotted points are participants' averages and blue triangles are group means.

d. Reaction time on learning trials vs testing trials. Reaction time is the amount of time between the appearance of stimuli and the choice time. Y axis is the reaction time in milliseconds and X axis is the learning/testing trials condition. Thicker lines and thicker dots represent group average.

Additionally, we analyzed the differences in accuracy and reaction time between learning trials and testing trials. Overall, subjects in all groups maintained high performance in learning and testing trials alike which suggests the retention of the knowledge acquired on learning trials (one-tailed binomial test: WN: p=.0206, ML:

p=.0206, R: p=.000018). We also observed a small trend in the white noise group of higher accuracy and lower Rt levels in the testing trials compared with learning trials (accuracy: 0.821 - 0.841, Rt: 1635 - 1615), yet again testing more subjects is necessary to confirm this observation. No similar difference was observed in reward and monetary-loss groups (Figure 2).

Changes in accuracy over time. because over time habituation can lead to differences in accuracy over days, we examined how many choices were accurate as a function of days. For both punishment groups, accuracy on learning trials fluctuated from day to day but without a clear trend and remained above chance for every day (one-tailed binomial test: WN: p=.0206, ML: p=.0206). Similar results were observed for the accurate choices in testing trials for stimuli learned in different days (one-tailed binomial test: WN: p=.0206, ML: p=.0206; Figure 2) which indicates a good consolidation of the acquired information all along the days of the experiment.

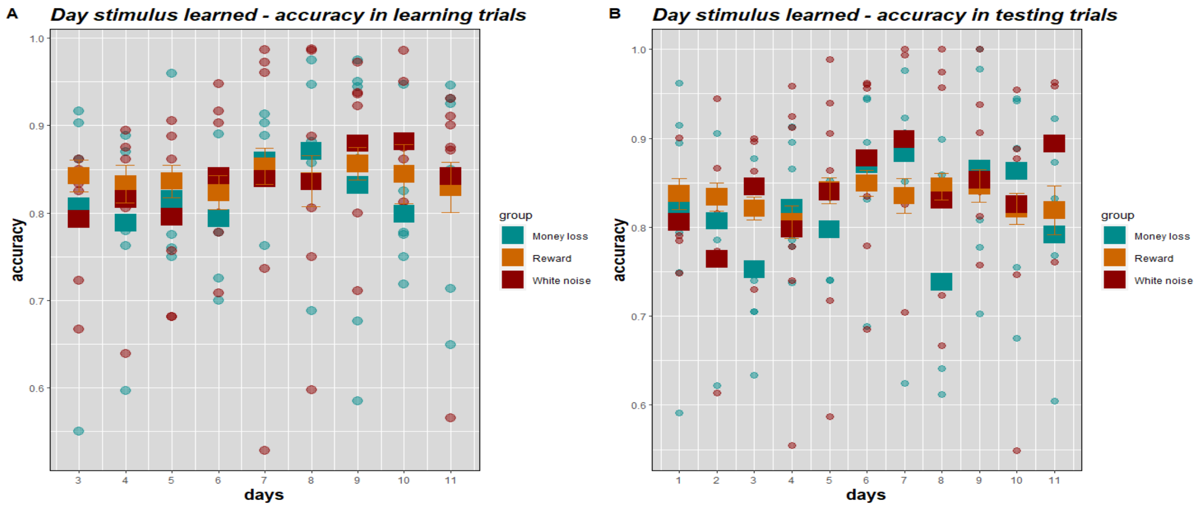

Figure 3:

A. Accuracy in learning trials over the day stimulus was learned. The rate of correct choices in learning trials of stimuli as a function of the day they were learned. On the X axis is the day that a stimulus was learned and on the Y axis the average accuracy of its learning trials. The dots are subjects' averaged accuracy on all stimuli that were learned that day. Squares represent group average. For the reward group instead of subjects' data points we added SE to the averages.

B. Accuracy in testing trials over the day stimulus was learned. Similar to panel A, but for accuracy in testing trials.

Learning curve. Averaging over all sessions for each trial in the learning game gave us a learning curve for each group that indicates how learning changed throughout the games. The curves start from over 70% accuracy and not from chance level because some games consisted of stimuli that were already learned in past games. We observed a gradual improvement in group averaged accuracy as the game proceeded, which suggest that learning continued throughout the duration of the games (Figure 4).

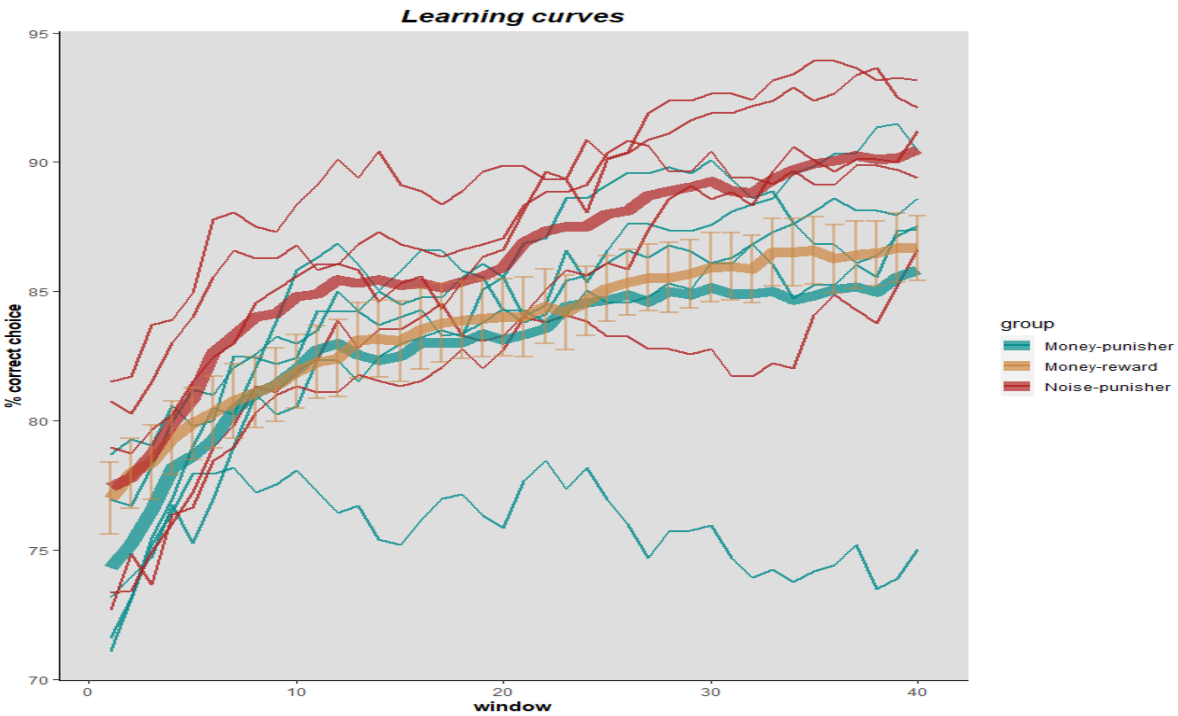

Figure 4: Learning curves. how learning changes throughout a learning game? On the Y axis the percentage of accurate choices, on the X axis are the sliding windows, where each window corresponds to a chunk of 10 trials (window 1 = trial 1-10, window 2 = trial 2-11, window 3 = trial 3-12 and so on). Thin lines are subjects' curves and thick lines are group curves. For the reward group instead of subjects' curves we added SE to each data point.

Stimuli recall. Looking at the proportion of accurate choices on testing trials as a function of how many hours passed since stimuli were learned revealed that in both punishment groups, long after learning the stimuli, accuracy levels in testing remained high on average even 75 hours after learning (Figure 4). Nevertheless, these results were not conclusive (Figure 5).

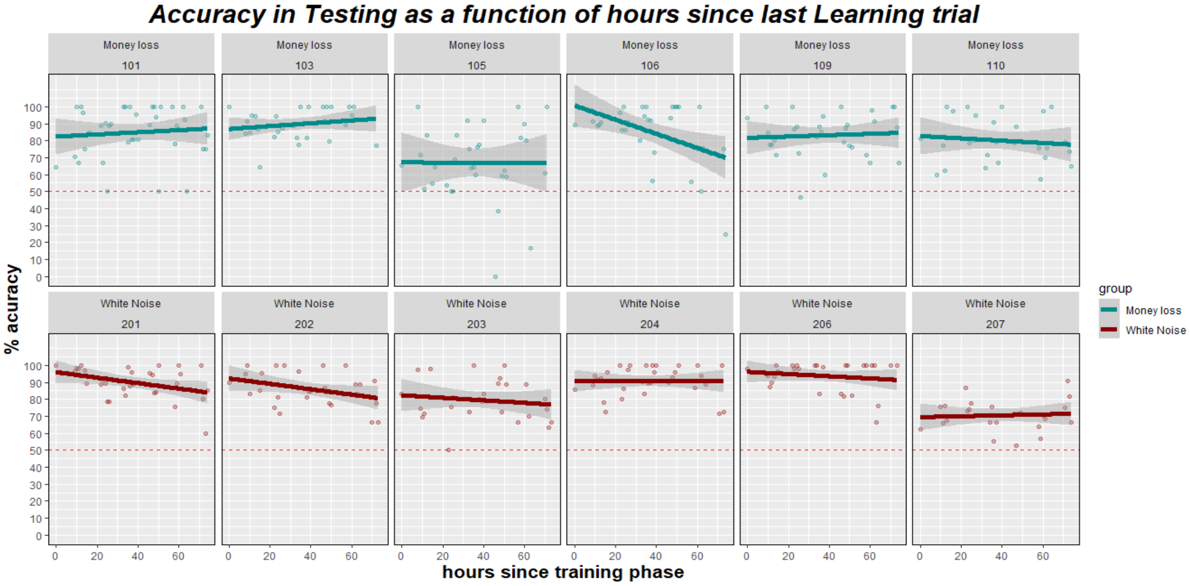

Figure 5: percentage of correct choices according to designated probabilities on testing phase as a function of how many hours passed since stimuli were learned for each subject on each group. Y axis is averaged accuracy and on the X axis the number of hours since learning.

Heartrate Responses. Looking at the implicit physiological impact to the aversive stimulus we checked for the patterns of heartrate responses during the 1 second before and 10 seconds after feedback was given. Similar to what was shown in Eldar et al (2018, see Figure 6), looking at the change in heart rate response, we observed in both groups a deceleration in heartrate responses a second after outcome was given (one-tailed binomial test: WN: $p=.0206$, ML: $p=.0206$) and a clear trend of modulation by the type of outcome received (punishment/no-punishment) after 3-5 seconds (Figure 6).

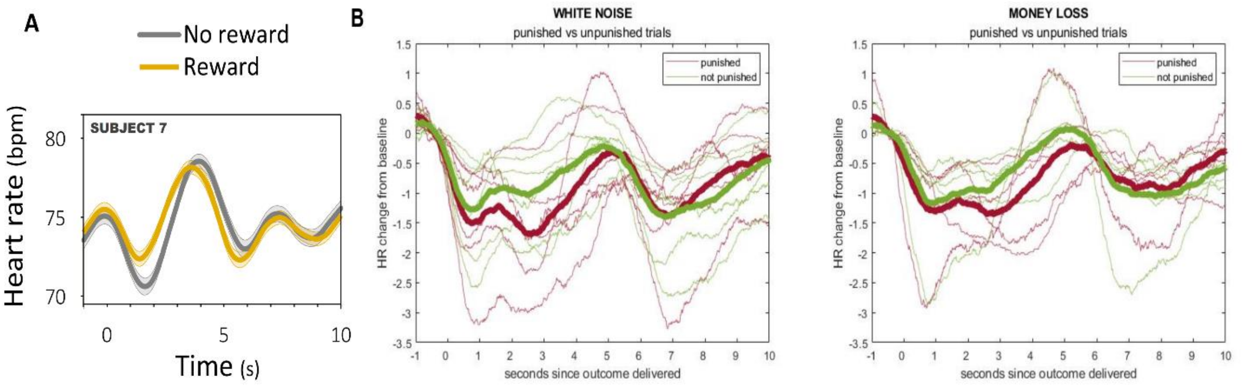

Figure 6:

A. Heartrate responses to reward vs no-reward trials of a subject from Eldar et al 2018. An example of subject's heartrate responses from 1 second before feedback given till 10 seconds after.

B. Heartrate responses to punished vs non-punished trials of both groups. The change in heartrate responses from baseline - averaged responses for the second before outcome presented, over the 10 seconds after feedback. Thin lines are subjects and thick lines are group averages. On the Y axis we see the change of heartrate responses from baseline. On the X axis is the time in seconds since outcome presented.

Self-reported questionnaire. As subjects finished the experiment, they filled a self-report questionnaire that requested them to rate (between 0 to 3, where 0 was "no agreement at all" and 3 was "highly agreeing") how they agree to an assertion about their experience during the experiment. Both groups rated high their motivation to avoid punishments while playing the games (WN: 2.5, ML: 2.4) and that in testing trials their choices were based on memory rather than guessing (WN: 2.5, ML: 2.4). Interestingly, they rated lower their success in identifying the worst images that are likely to produce punishments during the learning trials (WN: 2, ML: 1.6).

Colors task. All subjects of the noise group maintained a high rate of success in the Color-attention task (mean: 0.962±0.02, range: 0.924 – 0.995). This means that they had their headphones on during the task and the delivery of punishments.

Discussion

The aim of this project was to develop an experimental framework for probing aversive instrumental learning over multiple days with a focus on primary punishments. For this objective we used a mobile platform that can be used outside the lab by subjects and can be monitored remotely. We wanted to be able to reliably assess and reassess the daily changes in behavioral and physiological effects of the punishment.

Results showed overall high accuracy in choices between stimuli throughout the days of the experiment, both in learning and testing trials. These results prove that subjects were conditioned by the aversive stimulus on the stimuli that were presented to them during the experiment. Moreover, subject retained this information even after learning has ended.

Analysis also shows that these observations are not dependent on a specific day of learning or testing but rather this is consistent throughout the days of the experiment. This means that sensitivity to punishment did not systematically decrease throughout the experiment.

Therefore, habituation to the aversive stimulus was not clearly evident. This conclusion is also encouraged by the self-reported questionnaire in which subjects rated very high their motivation to avoid punishments.

Another variable that might be important here is the reaction time. Fast reaction time may indicate higher motivation (Touré-Tillery & Fishbach, 2014). Interestingly, we observed a trend of fast reaction time for the white noise group with the progress of the experiment which may suggest the maintenance of high motivation.

Looking at heartrate responses data, that is indicative of the implicit impact of the administered stimulus, suggest a clear trend in which the effect of the WN punishment is to that of the ML punishment and the Reward. Eldar et al (2018) operationalized responsivity to reward as the intensity of the physiological responses to the appetitive stimulus. Therefore, the resemblance between the primary punishment and the secondary punishment and reward suggests the heartrate responses as a way to operationalize responsivity to primary punishment.

Unfortunately, due to time limits, we had a relatively small sample size in this study which is not enough for drawing conclusions from the comparisons between the study groups. Nonetheless, we can see a few data trends that were worth mentioning, although expanding the sample size is needed to confirm the trends that were suggested here. Overall the WN group tended to have better accuracy rates and faster reaction time. Also, it looks like their acquired information lasts longer. Moreover, looking visually at heartrate responses we can see sharper slopes for the WN group compared to the ML group. All these trends may stem from the evolutionary fundamental aversiveness of the primary punishment.

Limitations and confounds. As mentioned above, this study has a relatively small sample size and therefore more data will need to be collected to verify some of the observations.

An important confound in our design is the fact that primary and secondary punishment were not numerically equivalent, meaning that we did not adjust the aversiveness of a loud white noise to the aversiveness of the loss of a coin (0.2 shekels) for every subject. The different intensities of aversiveness between types of punishments may have affected the results we got. Therefore we suggest that future studies will account for this and balance the aversiveness of both aversive stimuli.

Another possible confound is the fact that subjects in the monetary loss earned the initial amount of money that was in their experimental bank too easily (a draw from a wheel of fortune) and therefore might not appreciate the losses as their own losses. Although we tried to correct it by adding the WoF draw it might be better to create more effortful preliminary task and make subjects "own" their money.

Conclusion.

We conclude that the experiment we developed offers a useful tool for probing the sensitivity and responsivity of primary aversive stimulation over multiple days. Therefore, researchers may use this paradigm in future studies focusing on the relationship between aversive instrumental learning and behavioral and mood changes over time and compare the data with similar reward studies.

Bibliography

Aylward, J., Valton, V., Ahn, WY. *et al.* Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nat Hum Behav* **3,** 1116–1123 (2019).

Azrin, N. H., & Holz, W. C. (1966). Punishment. In W. K. Honig (Ed.), Operant behavior: Areas of research and application (pp. 380–447). New York: Appleton-Century-Crofts.

Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. Journal of Personality and Social Psychology. 1994;67:319–33.

Cools, R., Robinson, O. & Sahakian, B. Acute Tryptophan Depletion in Healthy Volunteers Enhances Punishment Prediction but Does not Affect Reward Prediction. Neuropsychopharmacology 33, 2291–2299 (2008).

S.G. Costafreda, M.J. Brammer, A.S. David, C.H. Fu, Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies Brain Res Rev, 58 (2008), pp. 57-70

Delgado Mauricio R, Li Jian, Schiller Daniela and Phelps Elizabeth A. 2008 The role of the striatum in aversive learning and aversive prediction errors. Phil. Trans. R. Soc. B3633787–3800.

Delgado, M. R., Jou, R. L., & Phelps, E. A. (2011). Neural systems Underlying AVERSIVE conditioning in humans with primary and secondary reinforcers. Frontiers in Neuroscience, 5. doi:10.3389/fnins.2011.00071

M. R. Delgado, C. D. Labouliere, E. A. Phelps, Fear of losing money? Aversive conditioning with secondary reinforcers, Social Cognitive and Affective Neuroscience, Volume 1, Issue 3, December 2006, Pages 250–259

Eldar, E., Hauser, T.U., Dayan, P. & Dolan, R.J. (2016a) Striatal structure and function predict individual biases in learning to avoid pain. Proc Natl Acad Sci U S A, 113, 4812–4817.

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of Reward Learning Signals Predicts Mood Fluctuations. *Current Biology*, *28*(9), 1433–1439.e7.

Franzoi, S. L. (2015). Psychology: a discovery experience. South-Western Cengage Learning.

Galea, J., Mallia, E., Rothwell, J. et al. The dissociable effects of punishment and reward on motor learning. Nat Neurosci 18, 597–602 (2015).

Howell, D. C. (2006). *Statistical Methods for Psychology* (6th ed.). Wadsworth Publishing.

Jean-Richard-dit-Bressel, Philip & Killcross, Simon & Mcnally, Gavan. (2018). Behavioral and neurobiological mechanisms of punishment: implications for psychiatric disorders. Neuropsychopharmacology. 43. 1. 10.1038/s41386-018-0047-3.

Johnston, J. M. (1972). Punishment of human behavior. American Psychologist, 27(11), 1033–1054.

Johnson SL, Turner RJ, Iwata N. BIS/BAS levels and psychiatric disorder: an epidemiological study. Journal of Psychopathology and Behavioral Assessment. 2003;25(1):25–36.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–292.

Iigaya K. Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. eLife. 2016; 5: e18073

Mcsweeney, F.K., Roll, J.M. Do animals satiate or habituate to repeatedly presented reinforcers?. *Psychonomic Bulletin & Review* **5,** 428–442 (1998).

Michely J, Rigoli F, Rutledge RB, Hauser TU, Dolan RJ. Distinct Processing of Aversive Experience in Amygdala Subregions. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020 Mar;5(3):291-300.

Michelle Taub, Robert Sawyer, Andy Smith, Jonathan Rowe, Roger Azevedo, James Lester,The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment,Computers&Education,Volume,147,2020.

Palminteri, S., Khamassi, M., Joffily, M. *et al.* Contextual modulation of value signals in reward and punishment learning. *Nat Commun* **6,** 8096 (2015).

Rasmussen, E. B., & Newland, M. C. (2008). Asymmetry of reinforcement and punishment in human choice. *Journal of the experimental analysis of behavior*, *89*(2), 157–167.

Ulf-Dietrich Reips, Chapter 4 - The Web Experiment Method: Advantages, Disadvantages, and Solutions,Editor(s): Michael H. Birnbaum,Psychological Experiments on the Internet,Academic Press,2000,Pages 89-117.

Sperl MFJ, Panitz C, Hermann C, Mueller EM. A pragmatic comparison of noise burst and electric shock unconditioned stimuli for fear conditioning research with many trials. Psychophysiology. 2016 Sep;53(9):1352-65.

Seow, T. X. F., & Hauser, T. (2021). Reliability of web-based affective auditory stimulus presentation. In: bioRxiv.

Steel, A., Baker, C.I. & Stagg, C.J. Intention to learn modulates the impact of reward and punishment on sequence learning. *Sci Rep* **10,** 8906 (2020).

Steel, A., Silson, E., Stagg, C. et al. The impact of reward and punishment on skill learning depends on task demands. Sci Rep 6, 36056 (2016).

Touré-Tillery, M., & Fishbach, A. (2014). How to Measure Motivation: A Guide for the Experimental Social Psychologist. *Social and Personality Psychology Compass*, *8*(7), 328–341.

Toshikazu Kuroda, Carlos R. X. Cançado & Christopher A. Podlesnik (2018) Relative effects of reinforcement and punishment on human choice, European Journal of Behavior Analysis, 19:1, 125-148, DOI: 10.1080/15021149.2018.1465754

Wimmer G.E., Poldrack R.A. Reinforcement learning over time: spaced versus massed training establishes stronger value associations. bioRxiv. 2017;

Wise, T., Dolan, R.J. Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. Nat Commun 11, 4179 (2020).

Wolfram Schultz. Behavioral dopamine signals. Trends in Neurosciences. Volume 30, Issue 5,2007. Pages 203-210,ISSN 0166-2236.

Abstract (Hebrew)

מחקרים מראים כי הרגישות לפרסים אצל בבני אדם עשויה להשתנות מיום ליום ושינויים כאלו עלולים להשפיע על מצבי רוח ובעיות נפשיות. אולם, עדיין לא נחקר האם שינויים ברגישות לעונש עשויים לשמש באותו תפקיד חשוב. חשובה במיוחד אולי היא הרגישות לגירוי שלילי ראשוני, כגון כאב או רעש חזק(בניגוד לעונש שניוני כמו הפסד כספי), כיון שזוהי הצורה הבסיסית והטבעית ביותר של עונש, הן בבני אדם והן בבעלי חיים כאחד. אולם, ביצוע מחקר על למידה מגירוי שלילי ראשוני הוא מורכב אפילו בניסוי של יום אחד במעבדה, לא כל שכן בניסוי שמטרתו לבדוק את הלמידה לאורך ימים רבים, שבמהלכו נבדקים עלולים לפתח הביטואציה לגירוי השלילי. מטרת הפרויקט שלנו הייתה לפתח פרדיגמה ניסויית שיכולה לחקור למידה מעונש ראשוני לאורך ימים רבים. לצורך כך, פיתחנו מטלת למידה מעונש מבוססת מובייל שמחלקת עונשים דרך אוזניות. המטלה מאפשרת לכמת גם רגישות וגם תגובתיות לעונש, כאשר רגישות לעונש מתממשת בנטייה של הנבדקים להמנע מהעונש שיוחס לגירוי השלילי, ותגובתיות לעונש מתממשת בתגובות קצב הלב לגירוי השלילי שנמדדות ע"י חיישנים לבישים. הניסוי נערך מחוץ למעבדה ונוטר מרחוק ע"י הנסיין. שישה נבדקים ביצעו פעמיים ביום סשן של הניסוי בטלפון הסלולרי שלהם למשך 12 יום. לצורך השוואה, קבוצה נוספת של שישה נבדקים ביצעה את הניסוי עם עונש שניוני, בדוגמת הפסד כספי. בכל סשן היה על הנבדקים ללמוד דרך ניסוי וטעייה את ההסתברות של תמונות שונות (עיגולים מסוגננים) להוביל לעונש. העונש הראשוני היה רעש לבן חזק שנוגן דרך אוזניות שנלבשו במהלך הסשנים. בכדי להבטיח שהנבדקים קיבלו את העונש באופן תקין, מטלת שמיעה וקשב ניתנה באופן רנדומלי בין

הטריילים. בכל ססן, הנבדקים למדו על גירויים חדשים. בכדי לבדוק עד כמה ידע חדש שנלמד גם נשמר בזיכרון, בחלק מהטריילים הנבדקים התבקשו לבחור בין תמונות שכבר נלמדו בימים קודמים. בכדי להעריך עד כמה הניסוי בדק בהצלחה למידה מעונש לאורך 12 ימי הניסוי, בדקנו את רמת ההצלחה של הנבדקים להימנע מתמונות שמקושרות עם סיכויים גבוהים יותר לעונש. כאמצעי מרומז אך יותר ישיר של ההשפעה הרגשית של העונש, מדדנו גם את תגובות קצב הלב של הנבדקים לפידבקים (עונש/לא עונש) במטלת הלמידה. הבחירות של הנבדקים המטלה הראו שהם הצליחו להימנע מגרויים שהיו יותר מקושרים עם הרעש הלבן החזק, אפילו כמה ימים לאחר הלמידה, מה שמצביע על התניה מוצלחת ומתמשכת. בנוסף, כמעט לא נצפתה ירידה בדיוק של הבחירות במהלך ההתקדמות של הניסוי, מה שמצביע על הביטואציה נמוכה לעונש הניסויי. ניתוח קצב הלב של הנבדקים, הראה תגובתיות עקבית לעונש בשתי קבוצות הניסוי. גודל הדגימה הקטן בניסוי הראשוני הזה לא איפשר השוואה קונקלוסיבית בין קבוצות העונש הראשוני והשניוני. אולם, בדיקה ויזואלית מראה כי רגישות לבחירה ותגובתיות קצב לב לעונש הרעש הלבן היו חזקים ומתמשכים לפחות כמו רגישות ותגובתיות להפסד הכספי. התוצאות הללו מאשרות כי הפרוצדורה הניסויית שפיתחנו מהווה פרדיגמה תקפה ואפקטיבית לחקירה מחוץ למעבדה של רגישות ותגובתיות לעונש ראשוני לאורך ימים רבים. בנוסף, באמצעות הפרדיגמה הזאת, מחקרים עתידיים יוכלו לחקור את היחסים בין שינויים טמפורליים במצב הרוח ומנגנונים של למידה מעונש.