Methods

# Materials and Methods

## Participants:

19 healthy volunteers (mean-age: 25.47±3.53; range 20 – 33 years old; 12 female subjects) 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) plus the amount left in the experiment bank or a finishing bonus (for elaboration on the paying method see the appendix section).

**Exclusion:** Seven participants were excluded from the study. Three participants excluded due to technical problems, another three because of insufficient compliance to the schedule of tasks, and one that had repeated bad performance from the beginning of the experiment.

After the 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).

**Additional data:**

Results of a parallel experiment on reward conducted in the lab was added to the analysis. 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 a 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" task, 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 to deliver punishment. Unknown to the subjects, the probabilities were set to be in a hierarchical structure with the probabilities of [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 the preliminary lab meeting subjects played 13 training blocks of overall 388 trials (that were excluded from analysis). Outside the lab subject 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. Load white noise
2. Monetary loss

Each participant went through a screening interview by phone to suffice the experiment criteria and if successful the subject was invited to the 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.

For the monetary-loss group, we conducted a preliminary task beforehand to create an experimental bank 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) although our task was unique. 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 (eq 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. Except of creating an experimental bank, the aim of this WoF task was 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. This was also used to create a sense of agency, to increase the engagement of subjects in the task (Taub et al. 2020), due to the possibility that since the reduction from the bank 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, participants might not consider the "big picture" and therefore disregard the loss on a single trial. The sense of endowment and agency over the bank of money increases the aversiveness of the monetary loss, therefore magnifying the effect on the subject (*Delgado et al 2011*; *Tricomi et al., 2004; Zink et al., 2004;* *De Martino et al., 2009*).

The loud white noise group was provided with earphones from the lab (Miracase MBTO106). To make sure that subjects are wearing the earphones throughout the games and listening to the delivery of punishments, we added "attention colors task" between trials. Randomly, every few trials appeared a screen with six rectangles of different colors. Then, one of the color names randomly asserted, 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 no feedback was administered, while a set of punishments was delivered, the task appeared randomly between noises.

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 delivery of aversive audio stimuli in a web-based experiments, Seow & Hauser (2021) showed that they are reliable for inducing affective states similar to in-lab studies, with the right technical measures.

For the White Noise group, a sound calibration was made beforehand 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 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 amount of intensity of the noise 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 "attention colors task". We wanted to keep volume to the minimum necessary for the subject to hear the names of the colors asserted but not higher so that they could not hear it without earphones. For each subject, while wearing the earphones, we played the names of colors in the lowest volume and increased it until the subject said she hear them clearly, but not more than that.

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

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

**APP SCHEDULE:**

The schedule of the app included a morning session of 2 consecutive blocks and an evening session 2 consecutive blocks. Before each session subjects had to put on the heartrate sensor and complete 5 minutes of rest time measurement. Each session took approximately 20-25 minutes. This routine was kept for 10 Consecutive days, following by a rest day and finally another experiment day meant as a summery test for all stimuli.

**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, after the choice was made, visual feedback appeared for 1s indicating the outcome of subject's choice (punishment or no-punishment) and immediately afterwards the next trial started, again with two thumbs on the screen. If the subject did not choose within 3 seconds an alert message appeared asserting that it is too slow.

Importantly, after they choose, subjects can see the outcome of their choice and learn the value of that stimulus. A punishment outcome was seen as a red arrow pointing down inside a circle with black background (and 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 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).

A picture containing shape

Description automatically generated

**Figure X:**

1. **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.
2. **Two stimuli each trial.** On each trial subject held the phone horizontally and chose between 2 circles with styled backgrounds.
3. **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.
4. **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, a curtain covered the outcomes to conceal them from the subject. This means that the training phase is over for these set of stimuli, and the **testing** phase starts. This way we can test how much those stimuli that were learned retained in memory. Although outcomes of choices were not presented, they were still stored, and 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 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 the number of times that the hidden outcomes conveyed punishments. This way we kept the incentives for choosing the better stimulus without revealing the outcomes.

**Monitoring subjects remotely.** To make sure that subjects are following the schedule and the instructions the 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 on other parameters of subjects' performance. To make sure that subjects are doing the tasks on time we looked at the times that the tasks were executed. To make sure that subjects are not just choosing randomly we tested the reaction time (we flagged it if it was below 1400 ms) and tested also for side bias to see if there is a tendency to choose one side more than the other (no side bias should occur).

## Analysis

# Frameworks. We analyzed the behavioral data using Rstudio (Version 1.1.456 – © 2009-2018 RStudio, Inc.; R version 4.0.3), and we processed the physiological data (HR) using MATLAB (version 9.5.0.944444 (R2018b)).

**Additional data.** First, we added to the analysis data from a parallel experiment on reward conducted in the lab. 19 subjects performed the same learning game in parallel schedule, but with rewards instead of punishment. This enabled us to compare both punishment groups with reward.  
**Preprocessing**. We preprocessed data by excluding the lab session trials from analysis, it serves for us as training phase. To calculate the RT we subtracted the time the stimulus presented from the time the choice was made by the subject.

**Binomial test.** To analyze how significant our results are we used the binomial test. The binomial test is meant to evaluate how much the observed data deviates from a null hypothesized binomial distribution. It is dealing with a dichotomous variable, just like our subjects in the experiment that their performance can be above chance level or not. Our question was is how likely is the data that we observed of subjects' success on each group if it came from a p=0.5 binomial distribution:

H0: p = 0.5

H1: p > 0.5

Now we need to calculate the probability to get the number of successes we got according to the null hypothesis using the binomial distribution:

A picture containing text, gauge

Description automatically generated

This gives us the p value of the null hypotheses and then we can reject it or not.