LUNDS
UNIVERSITET

INSTITUTIONEN FÖR PSYKOLOGI

Buy some carrots and skip the stick if you want your learning to stick - An experimental study on the retention of reward & punishment learning and the effect of context

Gabriel Köhler & Emil Olsson

Kandidatuppsats VT 2019

Handledare: Johannes Björkstrand
Examinator: Roger Johansson

Abstract

Operant conditioning is a psychological theory about learning through positive and negative reinforcement which has been researched for decades. However, some fundamental components of this theory have not yet been thoroughly researched, such as its interaction with long-term retention and context. These components are essential if we want to understand how operant conditioning applies in everyday life outside a human Skinner box. A computerized task was constructed based on reinforcement learning through operant conditioning. Data from 33 subjects were collected from two separate days of testing. During the first day, subjects learned associations between symbols and monetary outcomes under two different contexts. Five days later, memory retention was measured for these associations. In addition, a context manipulation was executed so that retention was tested in the same or switched context. The study found no significant difference for context the manipulation but a significant interaction between reinforcement type and memory retention. Therefore, our results suggest that context has no general influence on the retrieval of previously established operant responses, but further studies are needed. Long-term retention is proven to be worse after punishment than reward subsequent to reinforcement learning, in favor of the carrot over the stick. However, when measuring long-term retention after five days as in this study, there is a recovery of the negatively reinforced learning after exposure to retrieval cues.

Keywords: Operant conditioning, instrumental learning, long-term retention, context, reinforcement type, reward prediction error

Introduction

It is a well-established fact, and evident to most people that the consequences of our actions have an impact on the probability of their future occurrence. Similarly, we all continuously try to shape the behavior of others by selective application of reward and punishment, although most of the time, this is not a conscious or deliberate plan for behavior modification. It is just a basic aspect of how we interact with one another. We scold at our children when they misbehave and praise them when they are good in the hope of reducing or increasing the likelihood of such behaviors in the future. We nag at our respective other for not doing a proper job of the dishes or reward them with a happy smile when they have uncharacteristically remembered to fill up the car with gas. Such negative and positive consequences evidently affect behavior, as has been extensively studied from the earliest days in the field of psychology. It is well established that negative and positive consequences have an immediate effect on behavior, but the long-term retention of this type of learning has been relatively understudied, particularly in humans. This is surprising since the purpose of applying negative and positive contingencies is not just to increase or decrease their occurrence here and now, but to affect the probability of their future occurrence. Here to investigate these questions, subjects underwent operant conditioning using reward and punishment in two different contexts, and their memory was tested five days later, in either the same or different context than learning took place in.

Operant conditioning

The two different types of learning procedures that are often talked about in psychology are classical conditioning and operant conditioning (Kolb & Whishaw, 2001). Classical conditioning is simply pairing a repeated neutral stimulus with an event. The stimulus and the event are thus automatically associated with each other and produce a response. After some time, the response could be set off with or without the presence of the other stimulus. This type of learning procedure should not be confused with operant conditioning, which is, unlike classical conditioning, based on a more active learning process (Egidius, 1994). In operant conditioning, sometimes referred to as instrumental learning, the learning procedure is based on the consequences of a particular behaviour. The consequences could be either positive or

negative, sometimes referred to as reinforcement with reward or punishment, and this determines the likelihood that a particular behaviour will occur again. Edward Thorndike and B.F. Skinner, where the first two psychologists to study this extensively. They both did many experiments in which animals would try to figure out an action that would result in a reward. The animals therefore exhibited how they had learned how a particular behaviour was associated with a particular consequence by performing the task. To quote Thorndike himself when explaining his law of effect "Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond" (Thorndike, 1911, p. 244). Taking into account how simplifying the behaviorist could be, it should be stated that behaviors possess a far greater complexity than could ever be explained from inside a Skinner box, all depending on the level of analysis. Operant conditioning is not bound to one single brain circuit. The dedicated brain circuits are different and depending on what the learning procedure requires (Kolb & Whishaw, 2001).

Reward prediction error

Since the first experiments made by Thorndike and B.F. Skinner, many others have studied operant conditioning in various formats. New insight into the underlying mechanisms behind reinforcement related behaviour has been discovered. One of the underlying mechanisms that have received much attention is the dopaminergic system (Kolb & Whishaw, 2001). Studies have shown the importance of the dopaminergic system in the so-called reward prediction error. The reward prediction error represents the difference between an actual reward and a predicted reward. The reward could be any positively perceived stimuli and can be provoked by, e.g. an event or an object. A reward prediction error could be either positive or negative. When a positive reward prediction error occurs, the actual reward exceeds what is predicted. When a negative reward prediction error occurs, the actual reward does not meet the level of what is predicted. This can be illustrated by examining the activity of dopamine neurons

using artificial stimulation (Schultz, 2017). Positive reward prediction error signals, i.e., when a reward is better than expected, causes phasic activation of dopamine neurons (Bayer & Glimcher, 2005; Jang, Nassar, Dillon, & Frank, 2019; Matsumoto & Hikosaka, 2009). Negative reward prediction error signals, i.e., when a reward is worse than expected, dopamine neurons to become suppressed in their activity (Tobler, Dickinson, & Schultz, 2003). In real life, a positive reward prediction error is most likely a pleasant surprise and greeted as something that exceeded one´s expectations in a good way. On the other hand, a negative reward prediction error is probably a bad surprise and not likable. Either way, positive or negative, a reward prediction error will always have an element of learning and adaptation to it. The learning and adaptation come from the fact that one has to adjust one´s predictions or behaviours to the objective reality. The difference between the actual reward and one´s expectations about the reward is the error and the consequence that will facilitate learning. Thus, reward prediction error is closely linked to operant conditioning (Schultz, 2017). Dopamine activation is not only an essential component for understanding prediction error but also when explaining mechanisms behind memory formation. The dopaminergic system runs in part through areas of the hippocampus and basal ganglia. These areas have both shown to be crucial for memory (Gerrard, Burke, McNaughton, & Barnes, 2008; Kolb & Whishaw, 2001; Maquet et al., 2000). Not surprisingly, studies have found activation of dopamine neurons to affect memory and learning (Calabresi, Picconi, Tozzi, & Di Filippo, 2007; El-Ghundi, O'Dowd, & George, 2007; Jang et al., 2019; Wickens, Reynolds, & Hyland, 2003)

One study demonstrated the existence of this teaching signal in primates by observing that midbrain dopamine cells encode errors in reward predictions (Samejima, Ueda, Doya, & Kimura, 2005). There is, however, a lot of variables that seem to be incorporated in dopamine reward prediction error, like time, context, probability and magnitude of the expected reward. Then it follows logically that individual differences in subjective valency of these factors are paramount to dopaminergic activity and consequently learning (Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). Further investigations of the human role of dopamine in learning from reward prediction have been carried out by Pessiglione et al. (2006) who constructed a monetary gain task that could measure learning under operant conditioning. Participants had to choose between different arbitrary stimuli that had predetermined high or low outcome probabilities for monetary wins or losses. They found that similar learning occurred after 30 trials using gain

and loss reinforcers, with the remarks of lower internal consistency and higher response time for loss reinforced trials. Subjects treated with L-DOPA, a well-documented precursor to dopamine, showed higher monetary gain than subjects treated with dopamine antagonist haloperidol. But on the other hand, the drug manipulation did not affect learning in the loss condition, thus suggesting dopamine is profoundly influencing behavior towards approaching rewards but not towards avoiding losses, despite the fact that the topographical aspect of this type of learning is very similar. Overlapping but nevertheless, different brain regions seemed to mediate prediction error in the gain compared to the loss conditions. For example, the cerebellum seems to be involved in encoding negative outcomes related to motor learning which is generally categorized as implicit memories (Ernst et al., 2002; Galea, Vazquez, Pasricha, de Xivry, & Celnik, 2011) and the insula when learning from errors, more on this topic to be discussed later (Hester, Murphy, Brown, & Skilleter, 2010). A recent study added readable indicia to the understanding of reward prediction error by investigating the role of reward prediction error for consolidating memories in humans (Jang et al., 2019). The study found that visual information in their experimental task was encoded to a greater extent consequently to positive reward prediction errors compared to negative reward prediction errors, which later had a positive impact on memory retention. Moreover, the study observed that subjects who were more prone to taking risks in the task were also more likely to have a shift in their anticipatory attention, which in turn prompt reward prediction errors. Encoding enhancement through reward prediction error was noticeable both immediately and 24 hours after and thus independent of consolidation, indicating that positive reward prediction errors have a specific role in memory formation.

Long term retention

Operant conditioning with positive and negative reinforcement in a state of short-term memory and immediate responses has been studied extensively in both humans and animals (Ferster, 2002; Kim, Shimojo, & O'Doherty, 2006; Kolb & Whishaw, 2001; O'Doherty, 2004). These findings are, as mentioned, mostly concerned with short-term memory, when information is adhered in memory only briefly and then often forgotten. In long-term memory, there are no limits to how long information adheres in memory (Kolb & Whishaw, 2001). One critical

mechanism for memory and learning is consolidation. For a new memory to become permanent consolidation has to take place. Consolidation is the mechanism for storing new fragile memories into solid and stable long-term memories. Research suggests that it takes time for consolidation to occur, often a couple of hours or days. Therefore, not surprisingly, sleep is an important factor for consolidating new memories. Results from studies indicate that NREM sleep is especially crucial for explicit memory consolidation, and REM sleep is especially crucial for implicit memory consolidation (Gerrard et al., 2008; Kolb & Whishaw, 2001; Maquet et al., 2000).

Studies looking at operant conditioning in relation to long-term memory are quite few, and those studies published on this topic are foremost using rodents or insects. Findings from studies using rodents or insects regarding operant conditioning in relation to long-term memory suggest that positive reinforcement is more advantageous compared to negative reinforcement at facilitating long-term memory. One study measured memory in flies after training for olfactory discrimination. They found that memory persists for up to 24 hours after training in flies who were rewarded with sucrose after correct responses, compared to 4-6 hours in flies who were punished with an electric shock after incorrect responses. Interestingly, the concentration of sucrose and the intensity of the electric shock, magnitude of reinforcement, did not affect the flies memory in a significant way (Tempel, Bonini, Dawson, & Quinn, 1983). The magnitude of reinforcement has been studied in humans with inconsistent findings (Trosclair-Lasserre, Lerman, Call, Addison, & Kodak, 2008). Another study looked at olfactory learning and visual pattern learning in crickets. Results from this study showed that punishment had a substantially faster decay effect on memory than reward did in crickets. According to this study, it is now clearly proved that insects follow the pattern of better long-term memory for positive reinforcement and worse long-term memory for negative reinforcement. However, the underlying mechanisms behind this pattern and if these results are transferable to humans is not well established and should be emphasized in future studies (Nakatani et al., 2009). Despite the fact that the majority of studies have used insects or rodents when investigating operant conditioning in relation to long-term memory, there are a few exceptions with human models. Previous studies have shown some promising clues to the question if reward or punishment during learning effects long-term memory during motor-skill acquisition. Abe et al. (2011)

demonstrated that reward during the training of a motor skill enhances long term memory of that skill but punishment does not. The three conditions were split among three groups and consisted of rewarding feedback, punishing feedback and no feedback during a human motor skill task. The performance was after that tested in all three groups immediately, 6 h, 24 h, and 30 days after motor task training. There was no difference in learning between rewarding, punishing or neutral feedback during training and all three groups saw improvements in performance immediately after training. When tested after 6 h the reward group showed performance maintenance whilst the other two groups showed significant forgetting, but comparing to performance immediately after to 24 hours after the subjects who were rewarded during training performed better on the motor-skill task meanwhile the neutral and punishment group performed similar, thus all improved from 6 h to 24 h after. However, performance was only maintained in the reward group tested one month after training, while performance declined in the neutral and punishment group. Because maintenance of memory and learning for the rewarded group was driven by offline memory gain, it is suggested that reward improves memory consolidation and hence maintains long term retention. Further Rothwell (2011) argues that this fits previous hypothesis that dopamine, which is released after reward, aids the consolidation of memory at a synaptic level. Complementary research to (Abe et al., 2011) have proven enhanced learning from punishing compared to rewarding and neutral feedback under human motor skill learning meanwhile getting concordant results to previous discussed studies regarding increased memory retention from reward compared to punishment (Galea, Mallia, Rothwell, & Diedrichsen, 2015), Hester et al. (2010) also demonstrated enhanced learning from punishment in an associative learning task in which the learning was dependent on the magnitude of the monetary loss which could be predicted by insula activation. Greater monetary loss resulted in faster learning and more insula activation.

Context in learning and memory

Context in relation to operant conditioning is a topic which has recently gained more interest within the scientific field of psychology. However, the amount of studies looking at context in relation to operant conditioning are few, especially those testing humans. A couple of studies have investigated how context affects operant conditioning when rodents have learned to respond to a particular behaviour, through operant conditioning in a specific context,

the response weakens when there is a context switch (Bouton, Todd, & León, 2014; Bouton, Todd, Vurbic, & Winterbauer, 2011; Todd, Winterbauer, & Bouton, 2012). In one study examining renewal of operant conditioning after extinction, they found that the response of a particular learned behaviour was weaker at the beginning of extinction training when extinction was organized in a different context from the original training context (Bouton et al., 2011). Another study tested how well rodents transferred a learned response between two different contexts. The response was directly weakened after the context switch. This study also investigated whether the response would be affected by different types of reinforcers, lever pressing or chain pulling, in different contexts. Results indicate that the weakening in the rodents response after the context switch did not depend on the different reinforcers (Todd, 2013). To confirm this, another study used rodents to investigate whether there was an interaction effect between context, different reinforcers, and amount of training. No interaction was found, but the weakening of the response after the context switch remained (Thrailkill & Bouton, 2015). Therefore, instead of weakening the response with several interactions, a context switch seems to have a more general effect on the response. In addition to the role context plays in the extinction of operant conditioning, some studies have found evidence for that context plays a significant role in the acquisition of operant conditioning (Bouton et al., 2011). More specifically, context seems to play a role in operant conditioning before extinction has occurred. These findings have different interpretations, but perhaps the most prominent one is that the context could have entered into a direct association with the learned response (Thrailkill & Bouton, 2015).

Moreover, the topic of action versus habit in relation to context is important to consider when trying to understand how a direct association could develop within a particular context. Stimulus-response (S-R) association is frequently said to regulate habit learning. Not surprisingly, the formation of new instrumental S-R habits could conceivably be guided through a context-dependent stimulus. The distinction between action and habit origins from what is called the dual-process theories of operant conditioning. These theories state that there is clear evidence for a reward-based system consisting of two separate processes, one facilitating goal-directed actions, and the other creating new habits. The goal-directed actions are directly guided through their consequences and are often called response outcome (R-O) associations. Habits

have a more reflexive component and are guided through broad associations. Habits fall under the category of S-R associations. After sufficient R-O associations, there is a transition to S-R associations, that is, the action is transformed into habit. What is not yet clear is the cause and the duration to which this transition happens. However, the link between S-R associations and its dependence of context are highly relevant when trying to understand operant conditioning along with contexts (Dezfouli & Balleine, 2013). A direct association between a stimulus and a context might explain why a context switch produces a weaker response. It has also been suggested that a context switch produces a greater prediction error which could lead to a necessary behavioural change, in other words, a different response (Rosas, Todd, & Bouton, 2013). Although there has been a large number of well-validated learning tasks carried out in psychological research the last decades, some vital critique has been brought up recently. Wimmer, Li, Gorgolewski, & Poldrack, (2018) questioned the ecological validity of many such tasks by their agglutination of training into one massed session, when in reality most types of long-lasting learning occur after repeated training implicit or explicit on a task throughout many occasions. In their study Wimmer et al. (2018) found diverse neural correlates in reinforced learning for massed versus spaced repetition. Behaviorally massed training facilitated short-term learning, but decayed overtime meanwhile spaced training resulted in preserved memory weeks after. The memory maintenance as results of massed training sessions seemed more dependent on short-term memory performance and could largely be explained by individual differences in working memory capacity. Repetition is the mother of learning it is said and that it should be spaced out through time is nothing that has gone many students unremarked. However, research from categorical learning indicates that using larger variety of examples back and forth have a more substantial impact on long-term memory than timewise spaced repetition in itself (Kang & Pashler, 2012).

Psychological theories about operant conditioning have, for many years, tried to convey the real-life applications of reward and punishment in facilitating and motivating desired behaviours. Some notable implications can be found in social policies, citizen consumption, education, dementia and cognitive impairment (Ben-Elia & Ettema, 2009; Piatkowski, Marshall, & Krizek, 2019). Moreover, there are a plethora of clinical and hence humanitarian applications of these desired understandings but to name a few: “a common feature of addiction

is an increased sensitivity to reward and a diminished sensitivity to punishment that manifests as a failure to learn from or disregard negative or aversive outcomes." (Hester et al., 2010 p. 7); long-term retention of preference conditioning is impaired in schizophrenia as substantiated by an inability to maintain stimulus-reward relationships over time (Herbener, 2009).

In short, through this study we aimed to investigate the effects of operant conditioning on long-term retention in humans and if the same retention can be influenced by context. We hypothesize that context influences retrieval of previously established operant responses and that reinforcement type influences memory retention for operant responses.

Method

Participants

After initial pilot testing on three participants, two of them were excluded after only finishing one out of two days. One of them continued and completed the full study. Nothing operative was changed from the initial pilot testing to the rest of data collection sessions. When all data was collected, the experiment added up to a total amount of 33 subjects. Participant age ranged from 20 – 32 (M = 23.09, SD = 2.76) of which were 15 females and 18 males. The assortment of participants was selected through convenience and all subjects were students at different faculties at Lund University. Most of the recruitment was made through conversations around campus where students were asked to participate. To verify participation, students had to complete an online form in which information about the experiment's procedure was listed with time options for the appointments. Every participant received at least one cinema ticket for completing the full experiment which they were explicitly informed about beforehand.

Behavioral task

The experiment was arranged during two testing days, which took about 30 minutes each. Both days were separated into two test blocks, and one test block consisted of 128 trials, adding up to a total amount of 256 trials for each testing day. The task is a reconstruction from the monetary gain task used by Pessiglione et al. (2006) with some variations. During the task, participants got to see two symbols side by side on the computer screen on each trial. The

symbols were arbitrary in their form and intended to be entirely novel for the participants, meaning no symbol were supposed to have any previous associations connected to it. The symbols were characters from the Agathodaemon font, as has been used in previous studies (Pessiglione et al., 2006). Each trial also contained a background image displaying an environment behind the symbols. The background image in combination with a sound which would naturally occur in that environment played back through headphones constituted the context. E-Prime version 3 (Psychology Software Tools, Pittsburgh, PA) was used to program the paradigm and enable the desired visual makeup. The experiments were run on standard stationary computers with accompanying keyboard and headphones.

Participants were told that their objective was to choose between two symbols appearing on the screen (see *Figure 1*). Based on their choice, they would be able to win or lose points. The points were presented in the form of images of money where 1 point = 1 Swedish Krona. The subjects were explicitly informed that it was not real money they won or lost but only points. However, they were told that they should try to win as many points as possible and try to avoid losing points as much as possible. To motivate performance, the subjects were also informed that the five people in the study who collected the most points would be rewarded with an extra cinema ticket. In order to choose one of the symbols, either (key c) had to be pressed with the left index finger to select the left one, or (key b) had to pressed with the right index finger to select the right one. There was no time limit to choose between the two symbols; nonetheless, subjects were instructed to answer within a relatively short amount of time, 1-3 seconds approximately. After the choice was made, a gray box appeared around the selected symbol during one second followed by the outcome displayed on the screen for 1,5 seconds. The outcome meaning possible gain (+5 kr), loss (-5 kr) or neither gain or loss (0 kr). After completing the two test rounds during day 1, consisting of 256 trials and taking 30 minutes, subjects were done until day 2. As mentioned, during the first-day subjects received immediate feedback after each choice and got to see if they won or lost points. However, during the second day, the subjects were informed that the feedback had been removed, but that the task remained the same and that they should try to earn as many points as possible. In the instructions for the second day, it was also stressed that if they remembered a particular pattern for the symbols, the same pattern would apply this time. If they did not remember or notice any particular pattern for the symbols during the first day,

they were told to trust their gut feelings and try their best.

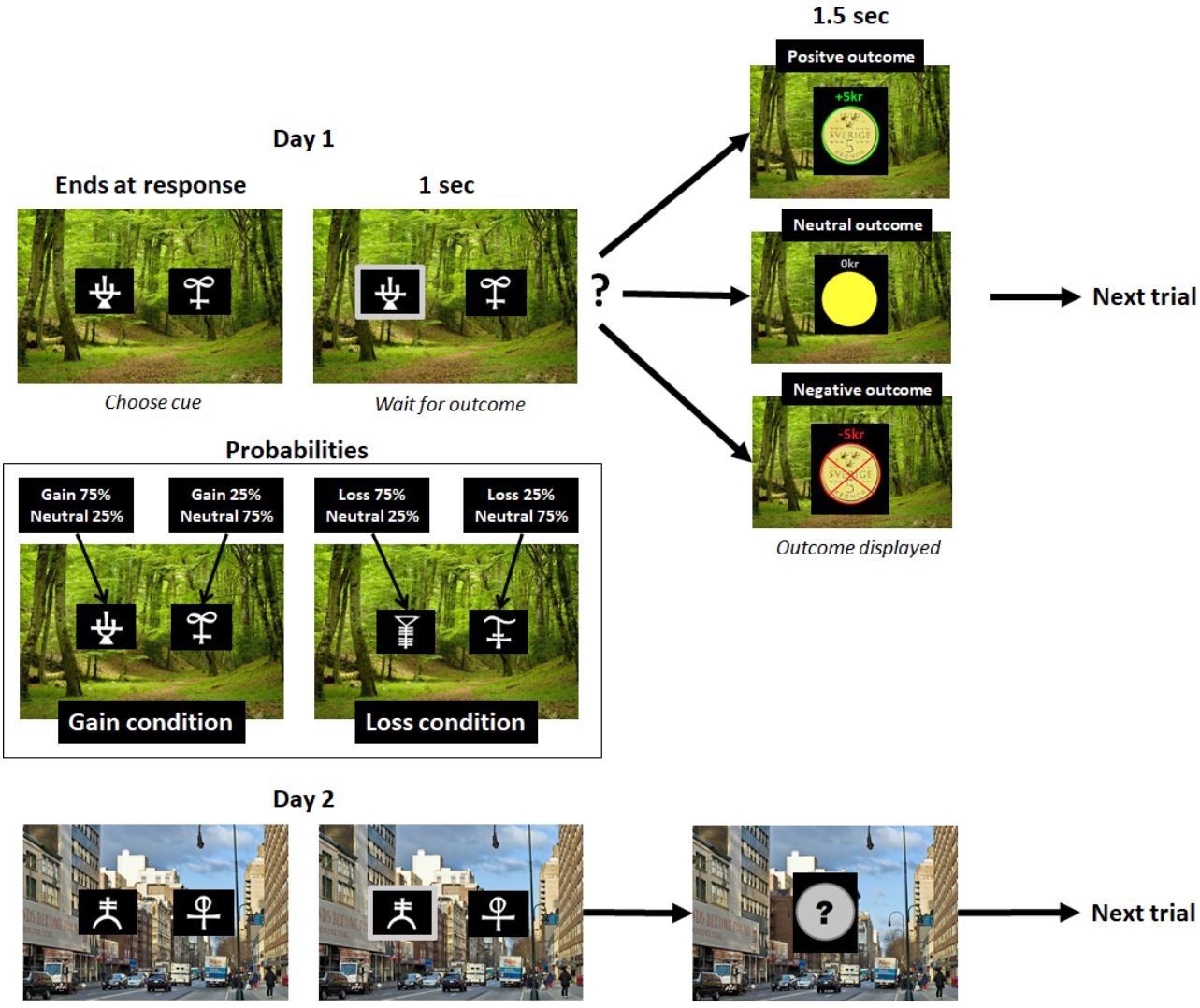

Figure 1. Instructions for the behavioral task. All possible outcomes after symbol choice are displayed in the top right. Same outcomes applied to day 2, but a question mark replaced the feedback. The middle picture describes the outcome probabilities in each condition.

Design

As this study primarily aimed to investigate potential effects on learning and long-term retention of two reinforcement types under two types of contexts, a within groups design was considered appropriate. In this design, every subject was exposed to every condition, stimuli and context an equal amount of times. The design of the behavioral task was created around two different contexts and eight pairs of symbols. The two separate contexts were two real-world pictures, either a forest background, called context A, or a city background, called context

B. To increase immersiveness and salience of the different contexts an appropriate soundscape was played in headphones, i.e., during training in context A subjects listened to a soundtrack of forest type sounds (birds chirping and wind blowing in trees) and during training in context B they listened to city-type sounds (noise of traffic, indistinct chatter in background, etc.). One test round consisted of four pairs of symbols in one context. Symbols, which were presented in pairs, had specific outcome probabilities in accordance with their Reinforcement type (see Figure 1). Four pairs constituted the gain conditions (Gain pair A-D), and four pairs served as loss conditions (Loss pair A-D). Similar to previous studies (Pessiglione et al., 2006), in the gain condition, one of the symbols was set at 75% gain probability and 25% neutral probability, whereas the other symbol in the same pair was set at 25% gain probability and 75% neutral probability. In the loss condition, one of the symbols was set at 75% loss probability and 25% neutral probability, whereas the other symbol in the same pair was set at 25% loss probability and 75% neutral probability. When testing for retention, the context manipulation was executed simply by testing the Gain and Loss pairs in either the same or different context. As exemplified in Figure 2, pair A and C would remain in the same context as they were originally presented; meanwhile, pair B and D would switch context. The specific stimuli assigned to be in in the reward or punishment condition was counterbalanced across subjects. Similarly, specific stimuli assigned to the low/high probability position was counterbalanced across subjects. Also, the specific stimulus pairs presented in the same/different context during the retention test was counterbalanced. Furthermore, the order in which the contexts were presented during acquisition and re-test was counterbalanced across subjects. This ensures that the behavioral effects cannot be assigned to any particular stimulus or be a consequence of order effects

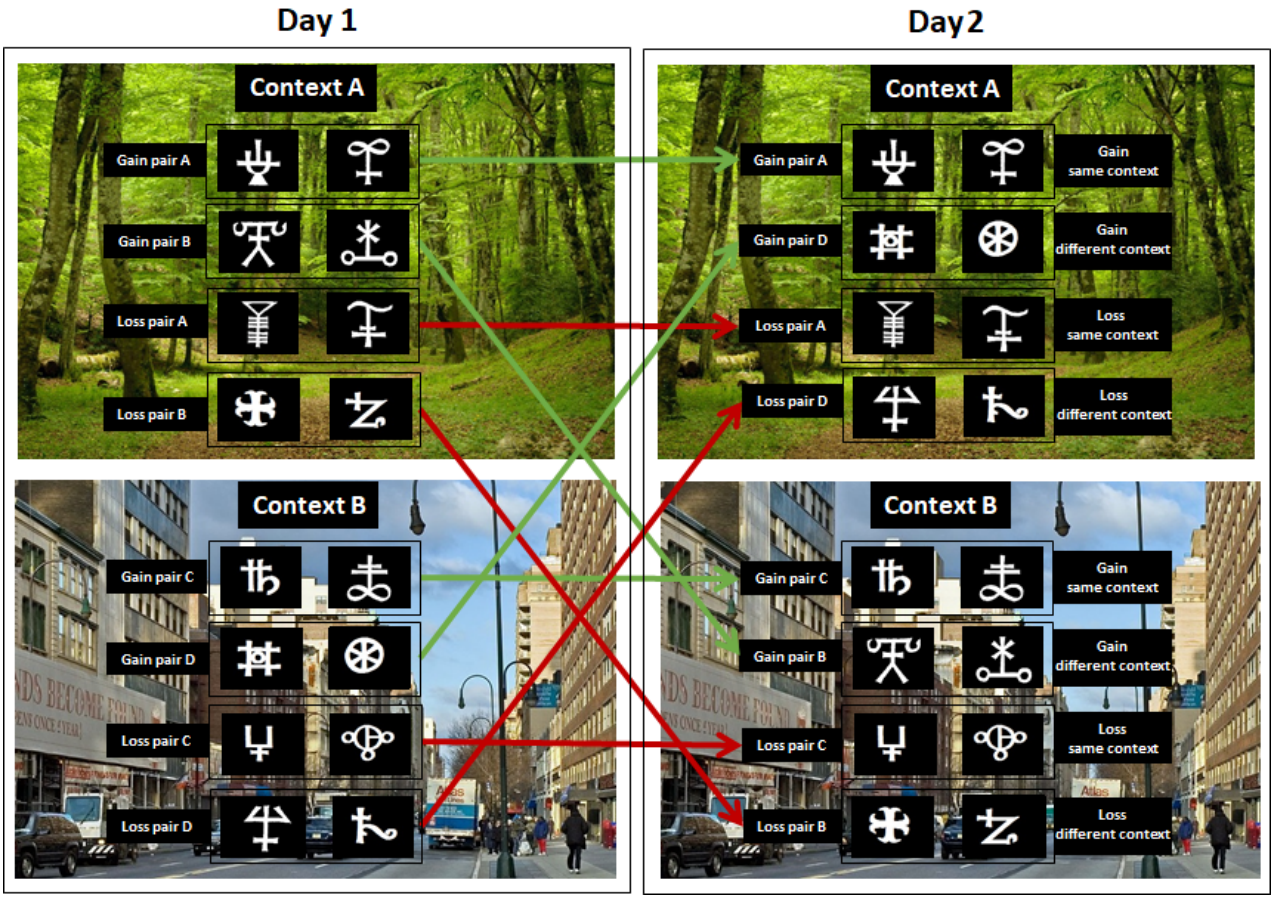

Figure 2. Design of the behavioral task. Arrows display how the context manipulation was executed by switching context for half of the symbol pairs

Procedure

The design of the experiment required participants to be available for about 30 minutes during two separate days. In addition to that, the two separate days were arranged to have no more and no less than four days between them. Testing sessions for each participant took place either Wednesday + Monday, Thursday + Tuesday or Friday + Wednesday, anytime between 9.00 am - 6.00 pm. The scheduled testing time did not have to be the same during both days. All testing took place in a room at the psychological institution at Lund University. The room was equipped with five stationary computers including headphones and was organized in a way that allowed for testing on up to five subjects simultaneously without any two subjects visually exposed to each other when performing the task. Upon arrival, all participants at the session were briefly instructed together by the experiment leaders about the procedure and had to read

a detailed explanation of the experiment before signing the informed consent. Subjects were thereafter asked to keep their phones in silent mode before being placed individually by a computer and put on provided headphones where they were urged to remain quiet throughout the experiment and raise their hands if they had any questions. The more task-specific instructions were shown at the screen and the task was started manually by an experiment leader when subjects were ready. Successfully completing the first day, subjects were told without further instructions that they would perform a similar task five days later. During the second day, subjects were informed verbally that they would have to perform two very similar tasks which they had done during the first day. Thereafter the same routine followed with the onscreen instructions, including information explained under the "Behavioral task" section, before the task was started. Those were that they would not receive any feedback, that the same pattern that applied today one still applies and that the score is counted in so that performance affects their winning chances. Upon completing the behavioral task the second day, participants were asked to fill in an online form containing the accompanying explicit memory assessments about details in the behavioral task along with two personality questionnaires. But these were not analyzed for the present paper and will not be further discussed. Thereafter when all subjects in the room at the time had filled out the questionnaires a more extensive debriefing was held, and participants had the opportunity to ask questions.

Data Analysis

For this thesis we focused only on the behavioral data, meaning how accurate individuals were in choosing the optimal choice across the experiment, i.e., choosing the stimulus with a high probability of winning points in the gain condition and the stimulus with a low probability of losing points in the losing condition. Data were analyzed using repeated measures ANOVAs evaluating potential effects of Reinforcer type (Gain; Loss) and at retest, Context (Same; Different) as well as how behavior changes across trials. Greenhouse-Geisser corrections were applied when assumptions of sphericity were violated. For the interpretation of interaction effects, we used analyses of simple main effects. When analyzing data during acquisition (day 1) when no context manipulation had yet occurred, data were collapsed across contexts. To increase the reliability of the behavioral measure we used averaged behavioral responses across

two trials; thus data from each phase of the experiment was divided into 16 bins (hereafter referred to as trials) since we felt this provided a good balance between gains in reliability and still retaining a good degree of temporal resolution. All analyses were performed using JASP version 0.9.2.

Ethics

The experiment was not associated with any particular discomfort. In the purpose of scientific analysis, the project intended to collect and record information about participants, including responses and results on the experiment task as well as answers to personality questionnaires. To ensure participants anonymity, all collected data were pseudonymized and stored on password-protected hard drives and computers so that only people associated with the project could have access to the information. Neither would any personal information be shared to any third party. All participants were informed of the general arrangement and handling of personal anonymity in the study such as anonymity and agreed to participate by signing an informed consent under full voluntary conditions, meaning they could abort the study and ask for their data to be withdrawn at any time without giving a reason why. Furthermore, all participants were debriefed verbally after they completed the study and were able to ask questions to the experiment leaders. They were also notified about where to find the results from the study and had attained contact information to the researchers when signing the informed consent.

Results

To evaluate the initial acquisition of behavioral responses on day 1 we performed a 2x16 repeated measures ANOVA with factors Reinforcement type (Gain; Loss) and Trial (1-16). The analysis showed a significant main effect of Trial, but no main effect of Reinforcement type and no Reinforcement type x Trial interaction (see Table 1). As can be seen in *Figure 3*. accuracy increases across the training session with no differences between gain and loss trials.

Table 1. Within Subjects Effects day 1. Learning with Main and interaction effects.

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p
Reinforcement type	None	0.333	1.000	0.333	1.807	0.188
	Greenhouse-Geisser	0.333	1.000	0.333	1.807	0.188
Residual	None	5.896	32.000	0.184		
	Greenhouse-Geisser	5.896	32.000	0.184		
Trial	None	4.538 [a]	15.000 [a]	0.303 [a]	13.118 [a]	< .001 [a]
	Greenhouse-Geisser	4.538 [a]	9.084 [a]	0.500 [a]	13.118 [a]	< .001 [a]
Residual	None	11.070	480.000	0.023		
	Greenhouse-Geisser	11.070	290.675	0.038		
Reinforcement type ✻ Trial	None	0.270 [a]	15.000 [a]	0.018 [a]	0.678 [a]	0.807 [a]
	Greenhouse-Geisser	0.270 [a]	8.290 [a]	0.033 [a]	0.678 [a]	0.717 [a]
Residual	None	12.767	480.000	0.027		
	Greenhouse-Geisser	12.767	265.295	0.048		

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

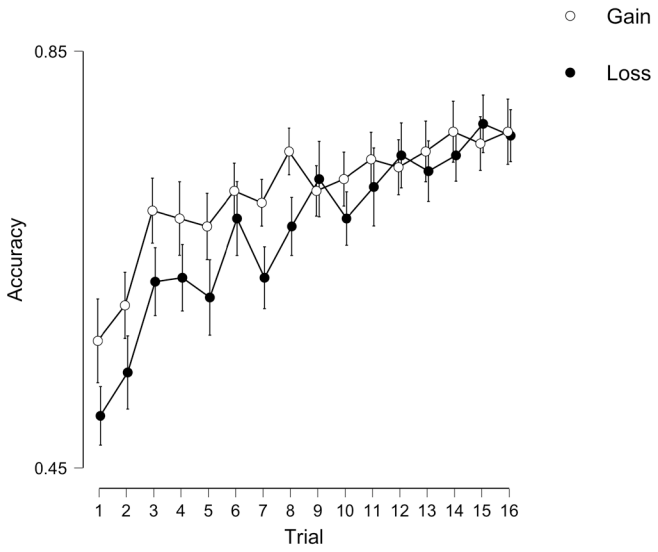

Figure 3. Learning curve Day 1 per Reinforcement type represented by the mean accuracy of choosing the symbol for the optimal probability of maximizing score expressed in percentage on the y-axis by the number of trials on the x-axis. Points and error bars denote means and SEMs

To evaluate retention day 2, we performed a 2x2x16 repeated measures ANOVA with factors: Reinforcement type (Gain; Loss), Context (Same; Different), and Trial (1-16). The results are displayed in Table 2. There was no main effect of either Reinforcement type, Context or Trial. There was nearly a significant interaction effect of Reinforcement type x Context x Trial, although this did not pass the alpha-level after applying correction for violations to sphericity (see Table 2). Analysis of simple main effects did not reveal any distinct pattern of effects to explain these results

and are therefore not presented. We did find a significant Reinforcement type by Trial interaction with a medium effect size (see *Figure 4* and Table 2) showing reduced retention for the Loss condition. Analysis of simple main effect with Reinforcement type as the simple main effect factor and Trial as the moderator factor shows a significant main effect of Reinforcement type for the first trial only (see Table 3) where accuracy is higher for Gain condition as compared to the Loss condition. Furthermore, there appears to be a recovery effect in the loss conditions as analysis of simple main effects with Trial as the simple main effect factor and Reinforcement type as the moderator factor shows a main effect of Trial in the Loss condition but not in the Gain condition, (see Table 4). Thus, although accuracy is lower for the Loss condition at the beginning of the retention test, behavioral responses recover and accuracy increases across trials, and differences in accuracy between Gain and Loss is present during the first trial only. Notably, accuracy increases in the loss condition despite the absence of any feedback.

Table 2. Within Subjects Effects day 2. Main and interaction effects with Reinforcement type x Context x Trial analysis.

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2
Reinforcement type	None	0.219	1.000	0.219	0.262	0.612	0.008
	Greenhouse-Geisser	0.219	1.000	0.219	0.262	0.612	0.008
Residual	None	26.746	32.000	0.836			
	Greenhouse-Geisser	26.746	32.000	0.836			
Context	None	0.086	1.000	0.086	0.183	0.671	0.006
	Greenhouse-Geisser	0.086	1.000	0.086	0.183	0.671	0.006
Residual	None	15.058	32.000	0.471			
	Greenhouse-Geisser	15.058	32.000	0.471			
Trial	None	0.306	[a] 15.000	[a] 0.020	[a] 0.853	[a] 0.618	[a] 0.026
	Greenhouse-Geisser	0.306	[a] 8.477	[a] 0.036	[a] 0.853	[a] 0.563	[a] 0.026
Residual	None	11.475	480.000	0.024			
	Greenhouse-Geisser	11.475	271.274	0.042			
Reinforcement type ✻ Context	None	0.006	1.000	0.006	0.008	0.930	0.000
	Greenhouse-Geisser	0.006	1.000	0.006	0.008	0.930	0.000
Residual	None	23.541	32.000	0.736			
	Greenhouse-Geisser	23.541	32.000	0.736			
Reinforcement type ✻ Trial	None	0.757	[a] 15.000	[a] 0.050	[a] 2.039	[a] 0.012	[a] 0.060
	Greenhouse-Geisser	0.757	[a] 8.362	[a] 0.090	[a] 2.039	[a] 0.040	[a] 0.060
Residual	None	11.872	480.000	0.025			
	Greenhouse-Geisser	11.872	267.576	0.044			
Context ✻ Trial	None	0.158	[a] 15.000	[a] 0.011	[a] 0.422	[a] 0.973	[a] 0.013
	Greenhouse-Geisser	0.158	[a] 9.388	[a] 0.017	[a] 0.422	[a] 0.928	[a] 0.013
Residual	None	11.979	480.000	0.025			
	Greenhouse-Geisser	11.979	300.408	0.040			
Reinforcement type ✻ Context ✻ Trial	None	0.600	[a] 15.000	[a] 0.040	[a] 1.919	[a] 0.020	[a] 0.057
	Greenhouse-Geisser	0.600	[a] 8.614	[a] 0.070	[a] 1.919	[a] 0.052	[a] 0.057
Residual	None	10.009	480.000	0.021			
	Greenhouse-Geisser	10.009	275.656	0.036			

Note. Type III Sum of Squares

Table 3. Simple Main Effects - Reinforcement type. Day 2 with Trial as the moderator factor.

Level of Trial	Sum of Squares	df	Mean Square	F	p
1	0.371	1	0.371	6.655	0.015
2	0.057	1	0.057	0.921	0.344
3	0.189	1	0.189	2.855	0.101
4	4.735e -4	1	4.735e -4	0.005	0.943
5	0.038	1	0.038	0.483	0.492
6	0.030	1	0.030	0.383	0.540
7	0.047	1	0.047	0.581	0.452
8	0.057	1	0.057	0.595	0.446
9	0.057	1	0.057	0.786	0.382
10	0.012	1	0.012	0.146	0.705
11	0.002	1	0.002	0.022	0.883
12	0.023	1	0.023	0.404	0.529
13	0.057	1	0.057	0.867	0.359
14	0.008	1	0.008	0.088	0.768
15	0.017	1	0.017	0.251	0.620
16	0.008	1	0.008	0.098	0.756

Note. Type III Sum of Squares

Table 4. Simple Main Effects – Trial. Day 2 with Trial as the moderator factor.

Level of Reinforcement type	Sum of Squares	df	Mean Square	F	p
Gain	0.341	15	0.023	1.113	0.341
Loss	0.722	15	0.048	1.703	0.047

Note. Type III Sum of Squares

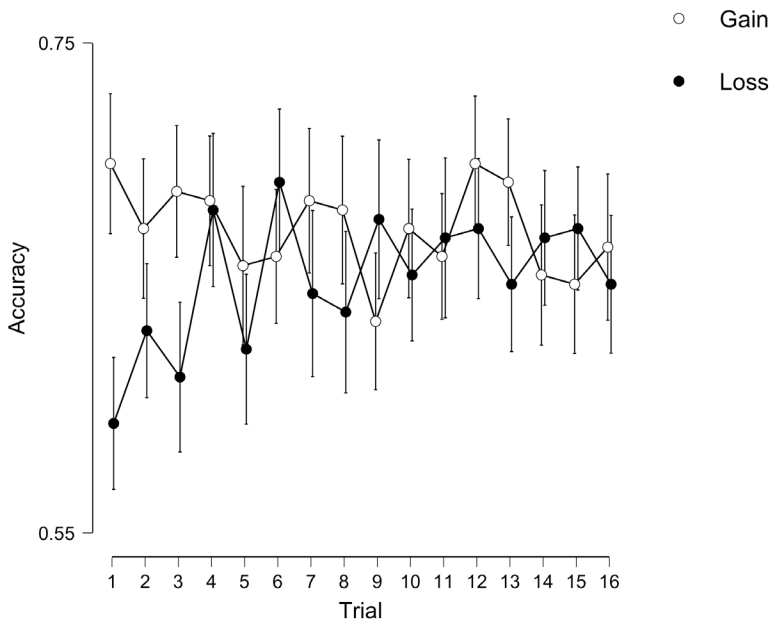

Figure 4. Performance by Reinforcement type day 2 with the recovery of Loss condition. Mean accuracy of choosing the symbol for the optimal probability of maximizing score expressed in percentage on the y-axis and number of trials on the x-axis. Points and error bars denote means and SEMs

Since the differential effect of Reinforcement type on retention seems to be restricted to the first trial of the retention test, we performed a follow-up analysis to further explore the effect of reward and punishment on retention by comparing accuracy at the end of acquisition to the beginning of the retention test. For this analysis, we collapsed accuracy scores across the context conditions since we did not find any significant effect of context on retention in the analysis presented above. Thus, we performed a 2x2 repeated measures ANOVA with factors Reinforcement type (Gain; Loss) and Time (Last trial during acquisition (day1); First trial during retention (day 2) (see Table 5). We found no main effect Reinforcement type, but we did find a significant main effect of Time with a very large effect size, indicating that across both conditions accuracy is lower day 2 as compared to day 1, and a significant Reinforcement type x Time interaction with a large effect size, indicating that the drop in accuracy from day 1 to day 2 is larger in the Loss condition as compared to the Gain condition. Analysis of simple main effects with Reinforcement type as the simple main effects factor and Time as the moderator factor confirmed a significant difference in accuracy only on the first trial of the retention test but no difference during the last trial of acquisition (see Table 6). Similarly, analysis of simple main effects with Time as the simple main effects factor and Reinforcement type as the moderator factor shows the main effect of Time is significant in both the gain and loss conditions (see Table 7). Thus, although accuracy drops from acquisition to retention in both conditions, the drop is greater in the loss condition. As can be seen from descriptive statistics for these measurements displayed in Table 8, in the gain condition accuracy drops from 77,3% during the end of acquisition to 70,1% during the first trial of the retention test, a difference of 7,2%, whereas in the loss condition accuracy drops from 76,9% to 59,5%, a difference of 17,4%.

Table 5. Within Subjects Effects. Main and interaction effects for follow-up analysis with factors Reinforcement type (Gain; Loss), and Time (acq 16th trial; ret 1st trial).

	Sum of Squares	df	Mean Square	F	p	η^2
Reinforcement type	0.100	1	0.100	2.138	0.153	0.063
Residual	1.490	32	0.047			
Time	0.500	1	0.500	23.415	< .001	0.423
Residual	0.683	32	0.021			
Reinforcement type ✻ Time	0.086	1	0.086	5.659	0.023	0.150
Residual	0.488	32	0.015			

Note. Type III Sum of Squares

Table 6. Simple Main Effects - Reinforcement type. Time as the moderator factor.

Level of Time	Sum of Squares	df	Mean Square	F	p
Acq 16th Trial	2.367e -4	1	2.367e -4	0.007	0.934
Ret 1st Trial	0.186	1	0.186	6.655	0.015

Note. Type III Sum of Squares

Table 7. Simple Main Effects – Time. Reinforcement type as the moderator factor.

Level of Reinforcement type	Sum of Squares	df	Mean Square	F	p
Gain	0.085	1	0.085	4.997	0.033
Loss	0.501	1	0.501	25.687	< .001

Note. Type III Sum of Squares

Table 8. Descriptive Statistics for the last trial of acquisition and first trial of retention for the gain and loss condition separately.

	Gain 16th trial Acq	Gain 1st trial Ret	Loss 16th trial Acq	Loss 1st trial Ret
Mean	0.773	0.701	0.769	0.595
Median	0.750	0.750	0.750	0.625
Std. Deviation	0.215	0.223	0.180	0.221
25th percentile	0.688	0.500	0.625	0.438
50th percentile	0.750	0.750	0.750	0.625
75th percentile	1.000	0.875	0.938	0.750

Note. Descriptive statistics

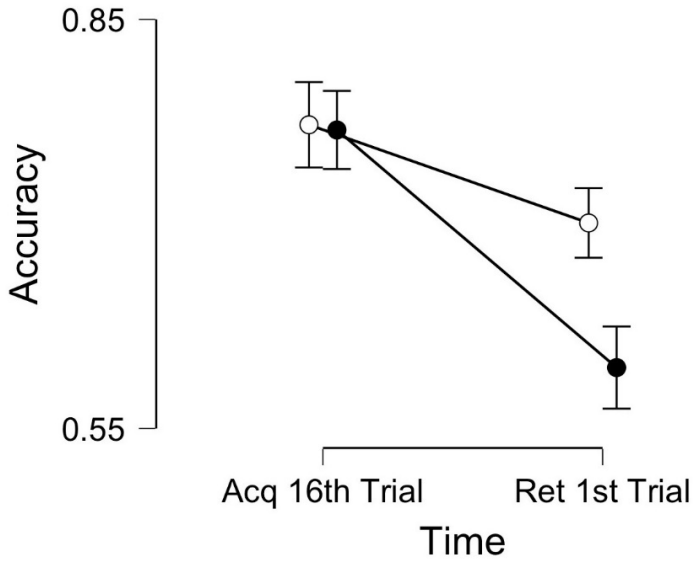

Figure 5. Retention by Reinforcement type, comparing accuracy on the beginning of day 2 with accuracy on ending of Day 1. Points and error bars denote means and SEMs.

Discussion

The purpose of this study was to investigate the long-term retention of operant reward and punishment learning and how it is influenced by context in human subjects. The first hypothesis was that operant conditioning is in part context-dependent, and therefore, retention should be better in the same context where learning took place as opposed to in a different context. Secondly, retention was hypothesized to be better for rewarded- than punished reinforcement. We did not find support for the hypothesis that context influences retrieval of previously established operant responses. We did find support for the hypothesis that reinforcement type influences memory retention for operant responses. The effect was in the predicted direction in that retention was poorer for responses established through punishment as opposed to rewards. Notably, there was a recovery effect underlying retention for responses reinforced with punishment. Thus, the differential effect is only present during the beginning of the retention test as data suggests responses established through punishment recover after a couple of non-reinforced presentations, whereas responses established using rewards remain stable.

Considering the effect of positive and negative reinforcement on learning, most of the existing literature has focused on its effect on short-term models of memory and learning. The sparse amount of literature regarding the effect of reinforcement learning on long-term memory is mostly from studies with rodents or insects. It is unclear to what degree these findings are transferable to humans but should nevertheless not be withdrawn from the equation as learning comprises basic mechanisms among all species and seems to operate somewhat similarly on a behavioral and biological level. Results from studies with rodents and insects suggest that learning through positive reinforcement is more advantageous than negative reinforcement at facilitating retention and long-term memory (Tempel et al., 1983; Nakatani et al., 2009). The few human studies in this area also support this idea (Abe et al., 2011).

In this study, results from day 1 showed a significant main effect of trials, but no difference for reinforcement type on learning. In other words, learning for both gain and loss conditions got better as more trials were completed. This confirms that the task was designed in a manner that enabled for learning to occur, and later have the possibility to measure memory

retention. The results from day 1 are primarily measures of short-term memory and learning. Previous research has indicated that positive and negative reinforcement has little to no difference in their impact on short-term memory and learning when measured immediately after training (Abe et al., 2011). Not particularly unexpected followed the fact that there was no significant difference between learning due to consequences of rewarding or punishing reinforcements from first to the last trial on acquisition day as displayed by the similar learning curves in *Figure 3*. However, this stands in contrast to one study reviewed which found enhanced learning from negative reinforcement, although this seemed to be limited to a skill acquisition paradigm were the negative reinforcement had to be directly related to the actual performance and was not dependent on monetary loss (Galea et al., 2015).

This study has placed a large part of its focus on long-term memory retention for operant conditioning and the effect of context. By testing participants during no feedback five days after the acquisition we tried to answer if long-term retention actually had been established, if it was best facilitated by rewarding or punishing reinforcement types and if retention was superior in the same context. Acknowledging the absence of a significant context effect on retention leaves us unable to accept the hypothesis that context influences retrieval of previously established operant responses. The hypothesis emerged from the research of Thrailkill & Bouton (2015), proving retention for operant learning is sensitive to- and weakened by context switch in rodents. One explanation for why no significant context effect was found is that subjects were not able to transition from R-O associations to S-R associations. Previous studies highlighted the importance of context when forming S-R associations and why they can have an impact on learning and memory (Dezfouli & Balleine, 2013). Another interpretation could be that the context manipulation was not too prominent to have an effect in our experiment. Supporting this theory would be what remained the same beyond the background picture and soundscape after context switch, namely the experiment room, the experiment leaders, the computer and so forth. In reality, context expands to more than what is currently in front of us. Additionally, there may be a distinct difference between humans and rodents in terms of limits to the amount of perceivable information that is categorized into a context, and what experiences we attribute to a context (Sapolsky, 2017). But it would not necessarily be wise to exaggerate the magnitude of these particular results taken into account that we almost found a significant interaction with

context, although follow-up analyzes did not show any clear pattern that allows us to interpret this. Just as for the effect of reinforcer type, the context effects could be limited to early trials because of subsequent recovery. If so, it is likely more difficult to detect these effects, as they are only evident in a few trials. It may be that the study has too little power to reveal these patterns clearly. Alternative explanations in the case for a potential inapplicability of context dependency to reinforcement learning may be extracted from research on fear conditioning. Suggesting experiences of fearful stimuli tends to be easily generalized upon because of conspicuous evolutionary reasons rendering us watchful to threats in our surroundings (Asok, Kandel, & Rayman, 2018).

A somewhat unpredicted but not necessarily implausible finding was the recovered retention of the loss reinforced trials on day 2. Unlike day 1, results from day 2 showed a significant Reinforcement type by Trial interaction. Indicating better retention for the gain conditions than the loss conditions at the beginning of the retention test, which in part is consistent with previous findings (Abe et al., 2011; Wimmer et al., 2018). As seen in *Figure 4*, the accuracy for loss reinforced value associations after five days caught up to the same level as those reinforced by gain after a few trials. In the absence of feedback, it could be argued that there was no evident incentive not to stick to the same choices except if the mere exposure of the stimuli served as retrieval cues. An interesting question is then raised about the applicability of retrieval cues in negative and positive memories. In this case, the result was dissociable in the sense that only retention for the negatively reinforced trials recovered slightly when considering the accuracy decline on both reinforcement types from day 1 to day 2 as to be further discussed later on. Could this be related to weaker consolidation of negative memories over positive memories, and if so, would the recovery effect disappear after a prolonged time, let us say if long-term retention were tested instead after 30 days or six months? One possible explanation for why subjects remembered gain conditions better than loss conditions, at the beginning of the second day, is that positive reinforcement generates positive reward prediction errors. When an outcome is better than expected, a positive reward prediction error occurs and signals dopamine neurons to activate (Jang et al., 2019). When an outcome is worse than expected, a negative reward prediction error occurs and signals dopamine neurons to become suppressed (Bayer & Glimcher, 2005; Matsumoto & Hikosaka, 2009). The activation of dopamine neurons, especially in the hippocampus and basal

ganglia, have shown to be important for memory integration (El-Ghundi et al., 2007). Thus, a considerable explanation for the significant Reinforcement type by Trial interaction during the second day is that a positive reward prediction error occurred during the first day, which activated dopaminergic neurons in important memory areas, and caused subjects in our study to have better memory and performance for gain conditions at the beginning of the second day. When comparing behavioral performance at the end of training day 1 to the beginning of the retention test day a significant drop in accuracy for both conditions is exposed. Thus, memory retention is not complete in either condition. However, the drop-in accuracy is significantly larger for the loss condition as compared to the gain condition, whereas performance during the end of behavioral training is highly similar for both conditions. This supports the conclusion that the retention gap between rewarded and punished responses is not due to differences in the initial acquisition of these responses

Before we consider to speculate in too descriptive terms about how we remember, we may take an analytical approach and ask ourselves what and why we remember. It has been said that memory exists for the purpose of predicting the future by the ability to learn from the past (Schacter & Madore, 2016). If we then ask why we predict the future, we may find the first clues to explain what we remember. Reinforcement learning states that we predict in order to strengthen the odds to reach positive outcomes as opposed to negative outcomes as a consequence of our constant interaction with environmental stimuli. Then what we remember in the most basic behavioral sense is governed by the degree of which a stimulus is followed by a desirable or undesirable outcome. What we do not yet fully comprehend is how and in what circumstances these opposite types of outcomes create and maintains memories, in this study we have contributed with some fresh clues. Since research in this area embraces fundamental human behavior new findings will hopefully continuously contribute to important applications in most thinkable fields from medical and clinical practice to business and education.

Limitations

As a whole, this study operated as intended with no major complications. However, in retrospect there are a few things we suspect could have been administered differently to fully explore how operant conditioning influences memory and its contextual interaction. Starting with the study design, we speculated if reinforcement type probabilities were set too difficult for

subjects to learn the value associations within the task properly. Performance during learning was highly variable, and since memory was under the scope for this study, one postulate could arguably have been that subjects properly should learn every value association before analyzed for long term retention. Most subjects did not establish sufficient learning to at least one of the eight pairs on day 1. This will undoubtedly have contributed with noise to the analysis. Including only subjects showing sufficient learning to all eight pairs would have yielded too few subjects for meaningful analysis. Thus, we accepted this as a source of noise for the present study, but future studies should take this into account and aim to achieve higher rates of learning during the initial training session. As mentioned before, we used pairs of symbols with a 75% to 25% probability of delivering a rewarding, neutral or punishing outcome; perhaps those were set to arbitrary. One approach could be to exclude subjects if their performance on the learning task was not significantly better than simulated random behavior to ensure that subjects were actively engaged in the task (Jang et al., 2019). Another factor to debate is whether the magnitude of reinforcement should have been set differently. Our reinforcement values were decided when formulating the experiment to be graded positive (+5kr), neutral (0kr) or negative (-5kr). Although we did see an effect, one could argue a different magnitude of reinforcement could have made participants feel that more was at stake and thus eliciting a different effect (Trosclair-Lasserre et al., 2008). Moreover, we cannot but wonder if the context design was too weak for an effect to occur. Context is a complex topic which converges around several environmental factors including more than just a background picture and related sounds. Nonetheless, previous studies with similar research questions have used even simpler context designs, compared to ours, and seen an effect, although not in humans (Thrailkill & Bouton, 2015).

Future research

Preexisting this study was a lack of published research when it comes to the possible effects of operant conditioning on long-term memory and the possible dependency of context in retention for pre-established operant responses. In human studies, this is a scientific void that we have just begun to scratch the surface of. One future approach would be to examine how different time frames could be applied. This could be hours, days or weeks after learning in order to test differences between reinforcement types and their influence on long-term retention. An interesting idea would be to test performance without feedback day 1, such as 30 minutes after training. If the effect is not there, then one can with greater certainty say that it is just memory consolidation that

differs between reward and punishment. But if there is a difference already after 30 minutes, then it is probably other processes that are the cause. For future research, it would be wise to ensure learning has taken place before measuring long-term retention. It would also be of interest to investigate potential dissimilar effects from different magnitudes of reinforcement values. Future research could also try to incorporate a more complete context experience. This could for example be possible through the use of virtual reality in order to create a more immersive real-world experience. Complementing the behavioural data with biological markers as could be derived from the use of brain-imaging techniques like fMRI would be paramount to further explore the underlying mechanisms in the effect of operant conditioning on long-term memory.

Conclusion

No evidence was found to back up the hypothesis that context influences retrieval of previously established operant responses, but we speculate that this may be related to specifics in our research design and conclude that further studies are needed to generalize. Our findings indicate that long-term retention is worse for negatively reinforced learning than positively reinforced, however, when measuring long-term retention after five days as in this study, there is a recovery of the negatively reinforced learning after exposed to retrieval cues. Buy some carrots and skip the stick if you want your learning to stick.

References

Abe, M., Schambra, H., Wassermann, E. M., Luckenbaugh, D., Schweighofer, N., & Cohen, L. G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Current Biology: CB*, *21*(7), 557–562.

Asok, A., Kandel, E. R., & Rayman, J. B. (2018). The Neurobiology of Fear Generalization. *Frontiers in Behavioral Neuroscience*, *12*, 329.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Ben-Elia, E., & Ettema, D. (2009). Carrots versus sticks: Rewarding commuters for avoiding the rush-hour—a study of willingness to participate. *Transport Policy*, *16*(2), 68–76.

Bouton, M. E., Todd, T. P., & León, S. P. (2014). Contextual control of discriminated operant behavior. *Journal of Experimental Psychology. Animal Learning and Cognition*, *40*(1), 92–105.

Bouton, M. E., Todd, T. P., Vurbic, D., & Winterbauer, N. E. (2011). Renewal after the extinction of free operant behavior. *Learning & Behavior*, *39*(1), 57–67.

Calabresi, P., Picconi, B., Tozzi, A., & Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neurosciences*, *30*(5), 211–219.

Dezfouli, A., & Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Computational Biology*, *9*(12), e1003364.

Egidius, H. (1994). *Natur och kulturs psykologi lexikon*. Natur och kultur.

El-Ghundi, M., O'Dowd, B. F., & George, S. R. (2007). Insights into the role of dopamine receptor systems in learning and memory. *Reviews in the Neurosciences*, *18*(1), 37–66.

Ernst, M., Bolla, K., Mouratidis, M., Contoreggi, C., Matochik, J. A., Kurian, V., … London, E. D. (2002). Decision-making in a risk-taking task: a PET study. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *26*(5), 682–691.

Ferster, C. B. (2002). Schedules of reinforcement with Skinner. 1970. *Journal of the Experimental Analysis of Behavior*, *77*(3), 303–311.

Galea, J. M., Mallia, E., Rothwell, J., & Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nature Neuroscience*, *18*(4), 597–602.

Galea, J. M., Vazquez, A., Pasricha, N., de Xivry, J.-J. O., & Celnik, P. (2011). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: the motor cortex retains what the cerebellum learns. *Cerebral Cortex* , *21*(8), 1761–1770.

Gerrard, J. L., Burke, S. N., McNaughton, B. L., & Barnes, C. A. (2008). Sequence reactivation in the hippocampus is impaired in aged rats. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(31), 7883–7890.

Herbener, E. S. (2009). Impairment in long-term retention of preference conditioning in schizophrenia. *Biological Psychiatry*, *65*(12), 1086–1090.

Hester, R., Murphy, K., Brown, F. L., & Skilleter, A. J. (2010). Punishing an error improves learning: the influence of punishment magnitude on error-related neural activity and subsequent learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(46), 15600–15607.

Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0597-3

Kang, S. H. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it

Promotes Discriminative Contrast: Spacing promotes contrast. *Applied Cognitive Psychology*, *26*(1), 97–103.

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology*, *4*(8), e233.

Kolb, B., & Whishaw, I. Q. (2001). *An introduction to brain and behavior*. Worth Publishers.

Maquet, P., Laureys, S., Peigneux, P., Fuchs, S., Petiau, C., Phillips, C., … Cleeremans, A. (2000). Experience-dependent changes in cerebral activation during human REM sleep. *Nature Neuroscience*, *3*(8), 831–836.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841.

Nakatani, Y., Matsumoto, Y., Mori, Y., Hirashima, D., Nishino, H., Arikawa, K., & Mizunami, M. (2009). Why the carrot is more effective than the stick: different dynamics of punishment memory and reward memory and its possible biological basis. *Neurobiology of Learning and Memory*, *92*(3), 370–380.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, *14*(6), 769–776.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045.

Piatkowski, D. P., Marshall, W. E., & Krizek, K. J. (2019). Carrots versus Sticks: Assessing Intervention Effectiveness and Implementation Challenges for Active Transport. *Journal of Planning Education and Research*, *39*(1), 50–64.

Rosas, J. M., Todd, T. P., & Bouton, M. E. (2013). Context change and associative learning. *Wiley*

Interdisciplinary Reviews. Cognitive Science, *4*(3), 237–244.

Rothwell, J. (2011). Motor learning: spare the rod to benefit the child? *Current Biology: CB*, *21*(8), R287–R288.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*(5752), 1337–1340.

Sapolsky, R. M. (2017). *Behave: The biology of humans at our best and worst*. Penguin.

Schacter, D. L., & Madore, K. P. (2016). Remembering the past and imagining the future: Identifying and enhancing the contribution of episodic memory. *Memory Studies*, *9*(3), 245–255.

Schultz, W. (2017). Reward prediction error. *Current Biology: CB*, *27*(10), R369–R371.

Tempel, B. L., Bonini, N., Dawson, D. R., & Quinn, W. G. (1983). Reward learning in normal and mutant Drosophila. *Proceedings of the National Academy of Sciences*, Vol. 80, pp. 1482–1486. https://doi.org/10.1073/pnas.80.5.1482

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. Macmillan.

Thrailkill, E. A., & Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *Journal of Experimental Psychology. Animal Learning and Cognition*, *41*(1), 69–80.

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(32), 10402–10410.

Todd, T. P. (2013). Mechanisms of renewal after the extinction of instrumental behavior. *Journal of Experimental Psychology. Animal Behavior Processes*, *39*(3), 193–207.

Todd, T. P., Winterbauer, N. E., & Bouton, M. E. (2012). Contextual control of appetite. Renewal of inhibited food-seeking behavior in sated rats after extinction. *Appetite*, *58*(2), 484–489.

Trosclair-Lasserre, N. M., Lerman, D. C., Call, N. A., Addison, L. R., & Kodak, T. (2008). Reinforcement magnitude: an evaluation of preference and reinforcer efficacy. *Journal of Applied Behavior Analysis*, *41*(2), 203–220.

Wickens, J. R., Reynolds, J. N. J., & Hyland, B. I. (2003). Neural mechanisms of reward-related motor learning. *Current Opinion in Neurobiology*, *13*(6), 685–690.

Wimmer, G. E., Li, J. K., Gorgolewski, K. J., & Poldrack, R. A. (2018). Reward Learning over Weeks Versus Minutes Increases the Neural Representation of Value in the Human Brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(35), 7649–7666.