The Effects of Social Reward on Reinforcement Learning

Anila D’Mello

Georgetown University
Abstract

Learning and changing behavior based on feedback, referred to as reinforcement learning, is an important method by which people plan actions in order to maximize reward. This study aimed to examine the effects of social versus non-social feedback on reinforcement learning. Subjects completed three versions of a reinforcement learning task in which feedback differed. These included a visual condition in which correct and incorrect feedback was presented visually, a neutral condition in which correct and incorrect feedback was presented by a human voice, and a reinforced condition in which feedback was presented by human voice saying positive and negative social statements. Learning strategies such as exploitative versus explorative behaviors and the effects of positive versus negative feedback on learning were measured. The type of feedback received was not found to have any significant effect on early learning strategies. In addition, there was no difference in the effects of positive versus negative feedback on reinforcement learning.
The Effects of Social Reward on Reinforcement Learning

Individuals are constantly learning from the feedback they receive in order to maximize benefits and rewards. Also referred to as reinforcement learning, this is a process by which organisms acquire the ability to map situations with actions that maximize resulting rewards (Sutton & Barto, 1998). As individuals maneuver through their environments, they must learn to modify their behavior appropriately to maximize their rewards. In particular, individuals learn to maximize reward by learning from previous choices and either exploiting previously correct choices (exploitative strategy) or avoiding previously incorrect choices (explorative strategy) (Frank, Seeberger, & O’Reilly, 2004). These exploitative and explorative learning behaviors can be differentially affected by correct versus incorrect feedback. In order to measure this process, many reinforcement-learning paradigms have attempted to modulated reinforcement or feedback in order to study the effects of feedback on behavior. Reinforcement learning paradigms, then, can be an interesting way to examine learning strategies and behaviors across individuals.

When it comes to reinforcement learning and exploitative or explorative learning strategies, probabilistic correct and incorrect feedback has been shown to be an especially strong force behind behavioral changes (Solomon, Smith, Frank, Ly, Carter, 2011). Studies using probabilistic feedback attempt to capture the ability of an individual to maximize reward by consistently choosing highly reinforced probabilities over less consistently reinforced probabilities, a process that is vital in maximizing reward. In addition, using probabilistic feedback such as described above enables a parsing out of differences between the effects of positive versus negative feedback on learning behavior. For example, certain studies suggest that individuals learn more from correct feedback and are therefore more likely to consistently exploit stimuli that were previously given correct feedback (Frank, et al., 2004). Findings such
as these give us insight into typical learning patterns and strategies with which individuals learn to operate in their environments.

Apart from learning strategies and patterns, reinforcement learning can also provide insight into how the nature of feedback affects behavior. For example, studies show that reward rather than mere feedback is an equally if not more powerful force behind behavioral change and decision-making. In reinforcement learning studies, monetary reward can often increase subjects’ choice of highly reinforced stimuli. In addition, adding as opposed to detracting monetary reward has been shown to have effects on learning strategy (Lin, Adolphs, & Rangel, 2011). Interestingly, this effect is not confined to monetary reward. Social reward, like monetary reward, can serve a similar purpose in reinforcement learning tasks. In tasks in which smiling and frowning faces are presented as reinforcement, social reward seems to activate many of the same neural substrates as monetary reward. In fact, in social reinforcement learning studies, subjects learned better and responded more quickly to those who provided frequent positive social reinforcement (Jones, Somerville, Li, Ruberry, Libby, Glover, Voss, Ballon, & Casey, 2011). Studies such as these imply that social reinforcement is viewed as rewarding in learning paradigms. Surprisingly, not many studies have modulated social reward in order to examine behavioral changes. Auditory social reward, for example, has been used in conjunction with smiling and frowning faces in reinforcement learning paradigms (Lin, et al., 2011). Interestingly, the effects of auditory social reward in reinforcement learning paradigms have not been examined alone. Because auditory social feedback is especially naturalistic, mimicking the praise and criticisms students and adults often receive in relation to their choices, any effect of social feedback on learning could have implications for teaching and learning strategy.

In this study, the effects of social reward were examined by comparing social feedback to non-social feedback in different versions of Michael Frank’s (2004) reinforcement learning task.
Frank’s reinforcement learning paradigm provides a good opportunity to easily modify feedback in order to include a social feedback condition. Expanding on Frank’s study, this study aimed to examine a few different aspects of reinforcement learning across different feedback conditions. Frank’s task consists of two parts: a learning phase and a testing phase. Early learning strategies and the effects of positive and negative feedback on learning are examined in the learning and testing phase respectively. Specifically, during the learning phase, two indices of learning strategy were examined: exploitative and explorative behavior. Exploitative behavior was defined as choosing the same stimulus after receiving positive feedback while explorative behavior was defined as choosing a different stimulus after receiving negative feedback. In addition, analysis of the testing phase outlined potential differential effects of positive versus negative feedback on stimulus choice. This study followed Frank’s basic reinforcement learning’s task while exploring the effects of social versus non-social and positive versus negative feedback on learning strategies.

An examination of the effects of social reward on reinforcement learning has wide-ranging implications. Understanding learning strategies in different populations has ramifications for education policy and targeted therapies that can hone in on specific learning patterns in different populations. For example, individuals on the autism spectrum exhibit impairments in social, motor, and communication domains. If conceptualized as a disorder of learning, understanding the conditions and situations under which learning occurs in these individuals could shed light on therapies that increase learning (Solomon et. al, 2011). Examining auditory social reward in particular could side-step difficulties with face-processing or social interactions often associated with autism. In addition, examining the effect of different feedback conditions on reinforcement learning and learning strategy could inform education policy and classroom behavior, providing teachers with an understanding of how best to motivate students to learn.
Methods

Participants

Twenty-one Georgetown University students between the ages of 18 and 22 years participated in the study (13 Female and 8 Male). Out of these, complete data was available for 17 subjects (10 Female and 7 Male).

Research Design

A 3 X 3 repeated measures design was utilized with feedback condition as one factor and reinforcement probability as the second factor. There were three feedback conditions: A text condition with visually presented feedback in which subjects saw “correct” and “incorrect” on the screen, a neutral condition with feedback delivered by a human voice in which subjects heard “correct” and “incorrect”, and a reinforced condition in with feedback delivered by a human voice in which subjects heard positive and negative social statements. In the reinforced condition, subjects heard 9 different positive and negative social statements presented in random order in the place of “correct” and “incorrect” feedback (see Appendix). There three reinforcement probabilities were: AB (reinforced at 80/20), CD (reinforced at 70/30), and EF (reinforced at 60/40). The task itself was comprised of two parts: a learning phase and a testing phase. Each subject preformed the task under each of the three feedback conditions, administered one week apart, in a counterbalanced order across subjects.

Procedure

Subjects sat down in front of a computer screen and were told that they would be participating in a short computer task of no longer than 30 minutes examining reasoning and decision making abilities. Subjects were informed that the study would be completed in three parts, separated by a minimum of 7 days in between each session. In each session, subjects completed either a version of a reinforcement learning task with visual feedback, a neutral
version of the reinforcement learning task with auditory feedback, and a version of the reinforcement learning task in which reinforcement consisted of auditory social reward as described in the section above. These order of these sessions were counterbalanced across subjects. At each session, subjects participated in two separate phases of the task: a learning phase and a testing phase. During the learning phase, participants saw two black Japanese symbols appear simultaneously on the screen. Throughout this phase, participants saw a total of 3 pairs of symbols (AB, CD, and EF) presented randomly (see Figure 1). Subjects were told to choose one of the two symbols and were given feedback (correct/incorrect) subsequent to their choice. One symbol was correct while the other was incorrect. Unbeknownst to the subject, feedback was probabilistic as follows: In AB trials, correct feedback followed a choice of A in 80% of trials while correct feedback followed the choice of B in 20% of trials. In CD trials, a choice of C led to correct feedback 70% of the time while a choice of D led to correct feedback in 30% of cases. Lastly, in EF trials a choice of E was followed by correct feedback 60% of the time while choosing F led to correct feedback in 40% of cases. Choosing the correct symbol was therefore slightly more challenging in CD and EF trials than in AB trials. Participants had to meet criteria in order to move on to the testing phase. Criteria was defined as being able to choose the correct symbol 65%, 60%, and 45% of the time in AB, CD, and EF trials respectively. Subjects repeated blocks of 60 trials until reaching criteria. After 6 blocks of 60 trials, if subjects had not yet met criteria, they were automatically moved on to the testing phase. The experimenter stayed in the room throughout the learning phase and left once the subject moved onto the testing phase. During the testing phase, participants saw novel combinations of symbols containing either an A or a B in each case in addition to previously tested stimulus pairs (see Figure 2). No feedback was provided.
In a post-test survey, subjects were asked to rate their familiarity with the characters on the screen. This was included to ensure that any effects were not caused by a previous knowledge of Japanese characters that would give certain subjects an advantage in the task.

**Results**

All data were analyzed using the Statistical Package for the Social Sciences (SPSS).

**Learning Phase**

In order to maintain consistency between subjects, only the first block of 60 trials was used as this was completed by all subjects. In addition, analyzing learning behaviors in the first 60 trials enabled an examination of early learning behavior. Other completed blocks were excluded from the analysis. Win-Stay and Lose-Shift behaviors were used as measures of exploitative versus explorative behavior respectively.

**Win-Stay Behavior**

Win-Stay behavior was defined as the percentage of times a subject chose the same stimulus after receiving positive feedback. This was calculated by dividing the number of times a subject “stayed” with the same stimulus after receiving positive feedback by the total number of times the subject received positive feedback. Win-Stay behavior was analyzed using repeated measures Analysis of Variance with feedback condition and reinforcement probability as within-subject factors. There was no significant main effect of feedback condition, $p=.469$, and no significant feedback condition X reinforcement probability interaction, $p=.736$. Therefore, the feedback received did not have any effect on subjects’ Win-Stay behavior; subjects were no more likely to Win-Stay in the social feedback. There was a significant main effect of reinforcement probability as subjects had higher Win-Stay percentages in higher probability conditions (AB, CD) than in lower reinforcement probability (EF) [main effect of reinforcement probability, $F(2,32)=5.453$, $p=.009$]. Subjects were more likely to exploit correct stimuli in the
highest reinforcement probability (AB) than in the middle reinforcement probability (CD), $F(1,16)=9.723$, $p=.007$. Subjects were also more likely to exploit correct stimuli in the middle reinforcement probability (CD) than in the lowest reinforcement probability (EF), $F(1,16)=7.372$, $p=.015$.

*Lose-Shift Behavior*

Lose-Shift behavior was defined as the percentage of times a subject chose a different stimulus after receiving negative feedback. This was calculated by dividing the total number of times a subject “shifted” to a different stimulus after negative feedback by the total number of times the subject received negative feedback. Lose-Shift behavior was also analyzed using repeated measures ANOVA with feedback condition and reinforcement probability as within-subject factors. There was no significant main effect of feedback condition, $p=.793$, and no significant main effect of reinforcement probability, $p=.242$. Therefore, neither the type of feedback received nor the probability of reinforcement had any effect on a subject’s explorative behavior. There was also no significant feedback condition X reinforcement probability interaction, $p=.677$; feedback condition had no effect across different reinforcement probabilities.

*Testing Phase*

During the testing phase, Choose A and Avoid B behavior was measured. Choose A was defined as the percentage of times a subject chose the most often reinforced stimulus (A) and Avoid B was defined as the percentage of times a subject avoided the least reinforced stimulus (B). Choose A and Avoid B were used to determine whether the subject had learned more from positive or negative feedback during the early learning phase. Choose A and Avoid B behavior were analyzed using repeated measures ANOVA. There was no significant main effect of feedback condition on Choose A, $p=.766$, or Avoid B, $p=.318$, behaviors. There was no effect of
feedback condition on subject’s percentage of Choosing A versus Avoiding B. This implies that there was no effect of positive versus negative feedback on learning.

**Discussion**

Different indices of learning strategy were examined in a reinforcement learning task in order to study the effects of social reward. During the learning phase, the type of feedback received had no significant effect on exploitative or explorative behavior. However, there was a significant effect of reinforcement probability on exploitative behavior. Subjects were more likely to exhibit exploitative behavior (choose the same stimulus after receiving positive feedback) in the highest reinforcement probability over the middle and lowest reinforcement probability. This same effect was not visible when examining explorative behavior. This implies that subjects were more likely to Win-Stay or exploit stimuli that had received positive feedback in high probability trials where stimuli were highly reinforced and correct/incorrect stimuli were easily discernable. In the testing phase, the type of feedback received had not effect on whether a subject chose A or avoided B. This implies that there was no difference in the effects of positive and negative feedback on learning behaviors.

It was interesting to find that type of feedback had no effect on reinforcement learning. While some difference was expected, it might be the case that the small sample size (N=17) made it difficult to discern any existing differences. Due to the within-subjects construct of this study, it was necessary for each participant to return three times for each separate session of the task. Often, subjects failed to come back for subsequent sessions, thereby resulting in an incomplete data set for certain subjects. These subjects were omitted from the data analyses. Furthermore, because high probability conditions were relatively simple, social feedback might not have provided any added benefits. Any real effect of feedback condition would be expected in the low reinforcement probability trails (EF). Again, small sample size might have precluded
any effect of feedback from appearing in the results. Moreover, time (number of days) in between sessions might have served as a confounding factor. While participants were asked to return at the same time and day after a minimum of 7 days from the original session for their subsequent sessions, not all did so. As such, some participants might have retained a greater familiarity with the task than others who were unable to come within one week of their previous session. Moreover, subjects had to reach criteria in order to move on to the testing phase of the task. This criteria consisted of choosing the correct stimulus 65%, 60%, and 45% of the time in AB, CD, and EF trials respectively. Subjects who did not meet criteria after 6 learning blocks were automatically pushed on to the testing phase. Because subjects differed in the amount of trials they underwent before moving on to the testing phase, it might be the case that subjects who met criteria in later learning blocks or who never met criteria skewed the results. Because only the first block of 60 trials was used in data analysis, it might be the case that subjects who took longer to learn the task only exhibited differences related to the type of feedback later on in the task. Lastly, while Frank’s original task describes “correct” and “incorrect” as feedback, it could be that this “feedback” actually had some social effect. Because the experimenter was present in the room during the learning phase, “correct” or “incorrect” visual feedback was visible to the experimenter. This could have provided a certain degree of embarrassment for subjects and served as a confound in subsequent analyses.

On the other hand, these results could be accurate. If so, it is possible that social feedback is ineffective at stimulating increased learning. Subjects might not have interpreted social statements such as “good job” or “you’re not really getting it” as rewarding. For example, in the Lin, et al., (2011) study, while neural substrates of monetary and social reward overlapped, monetary reward had a larger effect on reinforcement learning than did social reward. In the
current study, social reward statements might not have been interpreted as any more rewarding than seeing or hearing “correct” or “incorrect”. Following this train of thought, social feedback (in the reinforced feedback condition) was inconsistent compared to the text and neutral trials. While subjects repeatedly heard or saw “correct” or “incorrect” in these trials, the social statements meant to serve as feedback were variable in the reinforced feedback condition. This inconsistency and variability could have erased any expected benefit of social reward or feedback on learning.

An examination of the graphs seems to suggest that while not significant, the socially reinforced feedback condition might actually have had a negative effect. If this is the case, negative social feedback might have been embarrassing for subjects who subsequently could have become distracted by the negative statements. On the other hand, positive feedback might also have served as a distraction rather than an aid in reinforcement learning.

While feedback condition did not seem to have an effect on reinforcement learning, the study is ongoing to collect more subjects. Furthermore, the use of auditory social reward in the place of emotional faces or other more commonly used social rewards might have implications for reinforcement learning tasks conducted in patient populations. In these cases, smiling and frowning faces utilized as social reward can often be misleading when accounting for face processing difficulties that exist in certain studied populations such as ASD. If auditory social reward were to have an effect on reinforcement learning, this task could be repeated using control and patient populations and auditory social feedback could serve as an equally social but perhaps less distracting form of reward.
References


Figure Captions

*Figure 1.* Presentation of stimuli and feedback as seen by subject.

*Figure 2.* Schematic of novel stimulus pairs during testing phase.

*Figure 3.* Exploitative Strategy during early learning – Percentage of Win-Stay trials.

*Figure 4.* Exploratory Strategy during early learning – Percentage of Lose-Shift trials.

*Figure 5.* Testing phase – percent of Choose A vs. Avoid B.
まみ
CORRECT!
その
Test Phase

Choose A?

AC
AD
AE
AF

Avoid B?

BC
BD
BE
BF
AB (80/20)  CD (70/30)  EF (60/40)

- **Text**
- **Neutral**
- **Reinforced**
**Appendix**

List of Positive and Negative Rewards:

<table>
<thead>
<tr>
<th>Positive Social Reward</th>
<th>Negative Social Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fantastic!</td>
<td>Nope!</td>
</tr>
<tr>
<td>Good job!</td>
<td>Sorry, not this time!</td>
</tr>
<tr>
<td>Wonderful!</td>
<td>That’s wrong!</td>
</tr>
<tr>
<td>Yeah!</td>
<td>You’re not really getting it!</td>
</tr>
<tr>
<td>Way to go!</td>
<td>Thumbs-down!</td>
</tr>
<tr>
<td>You’re on a roll!</td>
<td>Wrong again!</td>
</tr>
<tr>
<td>You’ve really got it!</td>
<td>That’s not right!</td>
</tr>
<tr>
<td>Good work!</td>
<td>No!</td>
</tr>
<tr>
<td>Thumbs-up!</td>
<td>Way off!</td>
</tr>
</tbody>
</table>