



City Research Online

City, University of London Institutional Repository

Citation: Wang, J., Otgaar, H., Smeets, T., Howe, M. L. & Zhou, C. (2019). Manipulating Memory Associations Changes Decision-making Preferences in a Preconditioning Task. *Consciousness and Cognition*, 69, pp. 103-112. doi: 10.1016/j.concog.2019.01.016

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/21367/>

Link to published version: <https://doi.org/10.1016/j.concog.2019.01.016>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Manipulating Memory Associations Changes Decision-making Preferences in a
Preconditioning Task

Jianqin Wang^{1,2}, Henry Otgaar^{1,2}, Tom Smeets^{2,3}, Mark L. Howe^{2,4}, and Chu
Zhou⁵.

¹Catholic University of Leuven, Belgium

²Maastricht University, The Netherlands

³Tilburg University, The Netherlands

⁴City, University of London, UK

⁵Fudan University, China

Correspondence to Jianqin Wang (jane.wang@maastrichtuniversity.nl), Clinical

Psychological Science, Faculty of Psychology and Neuroscience, Maastricht

University, PO Box 624, 6200 MD, Maastricht, the Netherlands

Or to Chu Zhou (zhouchu@fudan.edu.cn), Department of Psychology, Fudan

University, 220 Handan Rd., Shanghai, China

IN PRESS: *Consciousness and Cognition*

(Accepted January 28, 2019)

Highlights

- We have examined the role of memory associations in a reinforced decision-making task.
- False feedback can undermine participants' memory associations.
- False feedback on memory associations reduces decision bias.

Abstract

Memories of past experiences can guide our decisions. Thus, if memories are undermined or distorted, decision making should be affected. Nevertheless, little empirical research has been done to examine the role of memory in reinforcement decision-making. We hypothesized that if memories guide choices in a conditioning decision-making task, then manipulating these memories would result in a change of decision preferences to gain reward. We manipulated participants' memories by providing false feedback that their memory associations were wrong before they made decisions that could lead them to win money. Participants' memory ratings decreased significantly after receiving false feedback. More importantly, we found that false feedback led participants' decision bias to disappear after their memory associations were undermined. Our results suggest that reinforcement decision-making can be altered by false feedback on memories. The results are discussed using memory mechanisms such as spreading activation theories.

Keywords: memory, false feedback, decision making, reinforcement learning, sensory preconditioning

Manipulating Memory Associations Changes Decision-making Preferences in a Preconditioning Task

Remembering previous experiences enables organisms to make decisions to acquire reward or avoid harm. For example, people would pick a restaurant where they had a good memory of having enjoyed food there previously. Or people with memories of getting allergic reactions to peanuts would avoid eating peanuts again. Surprisingly, little empirical research has been conducted regarding how memory impacts reinforcement decision-making, and only recently has the role of episodic memory in reinforcement decision-making attracted scientific attention (see Gershman & Daw, 2017, for a recent review; see also Weber & Johnson, 2006). Classical reinforcement learning theories usually adopt the computational perspective such as a statistical summary of past experiences (Shohamy & Daw, 2015). For instance, by making a bet on number 8, a person sometimes wins money (e.g., 80 times out of 100 times) and sometimes loses money (e.g., 20 times out of 100 times). The computational approach would assume that the person will bet on number 8 again because the summarised winning probability of betting on number 8 is 80%. This approach, however, has ignored how the experiences might be encoded in people's minds. For instance, as in selective forgetting (Anderson & Hanslmayr, 2014; Lind, Visentini, Mäntylä, & Missier, 2017), the person might selectively remember only the times that he has won, which could inflate his confidence in winning; or he merely remembers the times that he has lost, which would prevent him from betting more.

Indeed, Ludvig, Madan, and Spetch (2015) have recently found that priming memories of past wins can induce risk-seeking. They showed that participants who received a reminder of a winning experience tended to choose a risky choice more often than those who received no reminders. More recently, Bornstein, Khaw,

Shohamy, and Daw (2017) were the first to illustrate how consulting individual memories can bias reinforced decision-making. In their task, participants chose between two slot machines to win money and learned by trial and error that each slot machine delivered winning or losing tickets at different probabilities. At a later stage, participants were presented with the tickets as memory reminders before they made choices to win money. The results showed that those reminders of past choices strongly influenced participants' choices afterwards in that a reminder of a past action with reward made participants repeat the action while a reminder of loss made them avoid the action.

Neurobiological research has suggested the importance of memory associations in reinforced decision-making as well. Wimmer and Shohamy (2012) have shown that reward decision-making was correlated with activation in the hippocampus, where memory associations are usually formed. In their task, one similar to Pavlovian sensory preconditioning task (Brogden, 1939; Kimmel, 1977), participants first learned that a picture (S1+) and a circle always appeared together. In a later reward phase, they learned that the circle led to monetary reward. As a result, participants displayed a preference to choose the S1+ picture to win money above other equally familiar stimuli, even though it had never been rewarded. Neural imaging data revealed that this decision preference was predicted by activity in the hippocampus, a brain area encoding memory associations. This research showed that reward can spread across memory associations (e.g., from the rewarded circle to associated S1+ picture) via memory circuits in the hippocampus, which then impacts decision making. Our research question is, if memories for associations can guide reward decision-making, is it possible to alter decision making by changing participants' memory associations?

Memory Malleability and False Feedback

Episodic memory refers to mentally re-experiencing one's own previous experiences and it binds different elements of "what", "where" and "when" into an integrative experience (Tulving, 2002). More than 80 years of false memory research has demonstrated that memory is a highly adaptive and re-constructive system where its elements can be selectively sampled, intentionally forgotten or even completely distorted (Howe & Otgaar, 2013; Loftus, 2005; Schacter, 2012). For instance, people watched a parade of eight people, but then misremembered there were six people after answering a suggestive question "was the leader of four people a male?" (Loftus, 1975). Or one can falsely recall that a blue-eyed person committed a crime after receiving suggestive false information, while the perpetrator actually had brown eyes (e.g., Zajac & Henderson, 2009).

A plethora of research has demonstrated that memory can be manipulated via providing false feedback. One type of manipulation is to create rich autobiographical false memories through false feedback when in fact there are no relevant experiences underpinning such a memory. After being falsely suggested that they had been sick after eating a particular food such as egg salad in childhood, some participants report remembering vivid details of being sick on egg salad as a child and later show reduced preference for, and eating of, that particular food (Bernstein, Laney, Morris, & Loftus, 2005; Geraerts, Berstein, Merckelbach, Linders, Raymaekers, & Loftus, 2008; Scoboria, Mazzoni, & Jarry, 2008; Wang, Otgaar, Bisback, Smeets, & Howe, 2017). In those studies, participants have never experienced the egg salad-sickness association, yet they exhibit preference changes when they come to believe and remember this suggested association.

A more recent research line has used false feedback to undermine or “weaken” genuine episodic memories. In Mazzoni, Clark, and Nash’s (2014) study, participants performed actions such as clapping their hands and rubbing the table in front of a video camera. A few days later, their memories of the performed actions were tested and false feedback was provided to some of the actions suggesting that those actions were never performed. Some participants no longer believed that they had performed these actions but still remembered them, creating what are known as nonbelieved memories. Importantly, some memory characteristics, such as spatial and temporal clarity, even became weaker after false feedback. Other studies have found that false feedback led people to no longer believe in their memories of studied words and later changed memory related priming effects and problem-solving performance (Otgaar, Moldoveanu, Wang, & Howe, 2017; Wang, Otgaar, Howe, Smeets, Merckelbach, & Nahouli, 2017). Here, participants have experienced performing actions/seeing words, but false feedback about their memories made them disbelieve these memories and changed their memory-related behavior.

As we have summarized above, false feedback to experiences can either create a completely false memory or make people retract a memory. As we have reviewed, recent research has shown that memories about previous experiences play an important role in reinforcement decision-making such as gambling. Thus, it is possible to manipulate people’s memories for past experiences which can then change people’s choices generated from conditioning or reinforcement. More specifically, re-establishing or eliminating a memory association may directly impact behavioral performance based on these altered memory associations. According to associative activation theories of memory (Anderson, 1983; Howe, Wimmer, Gagnon, & Plumpton, 2009; Otgaar, Howe, Muris, & Merckelbach, 2017; Roediger, Balota, &

Watson, 2001), memory is a network consisting of nodes and various associations between these nodes. Crucially, activation of one memory node will automatically spread to nearby memory nodes via associations in a memory network and thus impact people's performance on tasks related to these activated nodes. For example, if people view the word "nurse" before identifying a related word "doctor", their response speed will be faster, and their responses more accurate, to "doctor" compared to not viewing any related words (i.e., priming effect of memory; McNamara, 2005; Meyer & Schvaneveldt, 1971). The reason is that the "nurse" node has been activated in the memory network and the activation spreads to related nodes such as "doctor", which facilitates the speed/accuracy response to "doctor".

In studies in which it was falsely suggested to people that they had been sick after eating a food such as peach yoghurt, a (false) "peach yoghurt – sickness" memory association was likely created. According to the spreading activation principle, presenting them with peach yoghurt would activate the node of "sickness" through the "peach yoghurt– sickness" memory association. As a result, people who had established such an association showed a lower preference for peach yoghurt and an avoidance to eating peach yoghurt (e.g., Scoboria et al., 2008). Conversely, when a memory association has already been established, for instance an "egg salad – sickness" association, telling participants that their memory was false led to no avoidance of egg salad anymore (Wang et al., 2017). These results suggest that feedback to an established memory association may change related decisions. In the study of Wimmer and Shohamy (2012), participants were conditioned to form "picture – circle – money" associations. Thus, the activation of money can spread to the picture via the rewarded circle. If false feedback is provided to the "picture –

circle” pairing to weaken this memory association, then it is possible to shift people’s performance on reward decision making tasks.

Recent studies have revealed that people’s choices can be changed by reminders of relevant memories (i.e., selectively sampled memory) (Bornstein et al., 2017; Ludvig et al., 2015). However, to our knowledge, no study has used false feedback for true memories to alter reinforcement decision making. The current study aims to examine whether false feedback on memory associations would change participants’ choices to gain reward.

The Present Study

Combining findings from memory research with the reward decision-making paradigm (e.g., Wimmer & Shohamy, 2012), in the current study we used false feedback to change participants’ memory associations that particular items were paired together, and then examined whether such false feedback would affect participants’ choices among items to win money. To be more specific, participants first learned associations among pictures and circles (e.g., an S1+ picture always associated with a yellow circle) and then learned that certain circles (e.g., the yellow circle) led to winning of money. In a memory test later, participants were told that the S1+ picture did not appear together with the yellow circle, but was associated with another non-rewarded circle. Finally, they were asked to choose from pictures and circles to gain money.

A recent model of memory suggests that autobiographical memory can be dissociated into two components, namely *recollection* referring to mental re-experiencing of events with sensory details and *belief* that refers to the truth value subjectively attached to a memory (Rubin, 2006; Scoboria, Jackson, Talarico, Hanczakowski, Wysman, & Mazzoni, 2014). For instance, people believe that they

were born but do not have recollections of the birth event or someone remembers vivid images of Santa Claus but does not believe in his existence anymore. Plenty of studies have demonstrated that belief in a memory can be changed or undermined via feedback, in which participants are told that their memory is incorrect (e.g., Mazzoni, Clark, & Nash, 2014; Otgaar, Wang, Franken, & Howe, 2018). Such feedback can lead people to disbelieve that particular experiences happened though their sense of recollection is maintained (see Otgaar, Scoboria, & Mazzoni, 2014 for a review). After false feedback was provided in our study, recollection and belief of memory associations were measured in the memory test, and we predicted that participants' memory associations would be undermined in at least one of these memory aspects. Based on research on nonbelieved memories, we expected that beliefs in the memory associations should be greatly impacted.

One novelty with our procedure is that we did not change memory for the *stimuli*, but manipulated the memory *association* between two stimuli. As associative memory theories propose (Anderson, 1983; Howe et al., 2009; Roediger et al., 2001), memory associations between memory nodes are key structures in an associative memory network, which determines the direction of spreading memory activation. If participants' self-reported memory associations do contribute to reinforcement decision-making, by undermining memory associations between the S1+ and rewarded circle, we would expect participants' decision preferences for S1+ pictures to disappear after false feedback. On the other hand, if reinforcement decision making is driven by mechanisms other than memory, we would expect participants' decision-making preferences to be unchanged by memory manipulations.

Method

Participants

Before recruiting participants, we used G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) to calculate the required sample size. With an estimated power of .80, power analysis revealed that 41 participants were required to detect a medium effect size ($d = 0.45$). Forty-one students from Fudan University, Shanghai participated in our study for a financial reward of ¥30. The sample consisted of 7 males and 34 females, with age ranging from 18 to 27 years old ($M_{age} = 21.2$, $SD = 2.22$).

Design and Procedure

The current study was a within-subject design. In the False feedback condition, participants received false feedback on their memorized associations in order to break the established associations. In the True feedback (control) condition, participants received true feedback about their memory associations. Half of the associations were provided with false feedback and the other half with true feedback. We also included filler pictures (no associations formed) in the task, which served as a baseline of decision preferences. The procedure basically followed the same steps as in the sensory preconditioning paradigm by Wimmer and Shohamy (2012), except that we included a memory test with feedback before the decision phase. Throughout the experiment, the participant sat in front of a computer, with the distance around 45-50 centimeters from the screen. An experimenter (a research assistant) sat next to the participant giving instructions when needed.

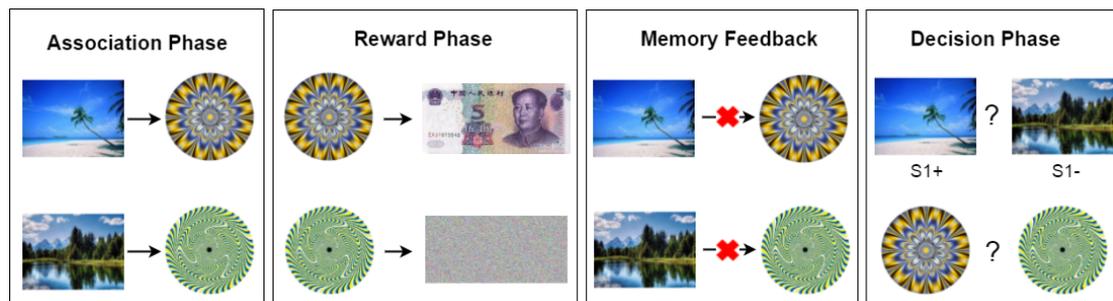


Figure 1. Demonstration of procedure in the False feedback condition. The True feedback condition had an identical procedure except that the memory feedback given was true. Figure 1 illustrates one category of pictures (scenes). In total there were four categories of pictures (see Appendix).

Preconditioning Phase 1: Association Phase. As Figure 1 shows, in the association phase, neutral pictures were paired with neutral patterned circles. Participants were merely asked to view some pictures in the instruction. However, without telling them to memorize associations, a picture (scene, furniture, body part or vehicle) always appeared before a particular patterned circle (see Walther, 2002). There were four categories of pictures and in each category there were two pictures: one picture (labeled as S1+ picture) paired with a later rewarded circle and the other picture (labeled as S1- picture) paired with a non-rewarded circle. Order of the pairs was randomized and each pair was presented for 10 times randomly. Four filler picture pairs (picture-picture) appeared randomly as well. Each picture or circle was presented for 1.5 s, with 1 s interval between them. The interval between each pair was 3.5 s to help participants learn the associations. All trials were completely randomized in one block. After all pairs were presented, participants rated their likings for all stimuli on 1-7 scales (1 = very disliked, 7 = very liked).

Preconditioning Phase 2: Reward Phase. Following the association phase, participants learned in the reward phase that half of the circles led to monetary reward while the other half led to no reward. They were instructed to respond with the “F”

key within 500 ms when they saw a picture of a money bill and respond with the “J” key when detecting a grey square. Once they responded in time and correctly, they won the money to their virtual money account. The money bill was always preceded by half of the circles while the grey square was preceded by the other half. Thus, there was a 100% contingency rate in the reward phase. Participants were explicitly told to use the circle cues to respond so that they could learn the reward associations.

Participants went through 128 trials in total.

Feedback Phase. Before the decision phase, participants completed an incidental memory test for the associations in the association phase. The memory test first asked participants to recognize which circle was paired with a particular picture (the picture was shown at the center of the screen, with two choices provided below it: a correct one and a wrong one). There were 12 recognition trials in total. After each recognition choice, verbal feedback was provided immediately via the computer with “correct :)” or “incorrect :(” shown on screen for 1.5 seconds and then the computer screen presented the following “The correct answer is...” with the acclaimed correct circle presented on screen. For half of the associations ($n = 4$; two S1+ associations and two S1- associations), the default correct answers were purposely exchanged for S1+ and S1- stimuli in the E-prime program. So for those associations, the program falsely told them that the correct answer of an S1+ stimulus was a non-rewarded circle, while it told participants that the correct answer of an S1- stimulus was a rewarded circle. At the same time, the experimenter told the participant: “Your memory is wrong. I saw clearly that the picture was paired with this circle.” For the other half of the associations ($n = 4$), the program provided the true status of associations while giving feedback. Four filler picture pairs were also included in the memory test, and they were always provided with correct feedback to make the

overall feedback credible. After feedback was given, participants rated their recollection (“Do you actually remember that the two items were paired together?”; 1 = no memory at all, 8 = complete memory) and belief (“Do you believe that the two items were paired together?”; 1 = do not believe at all, 8 = definitely believe) on 1-8 scales for the original association (i.e., the associations they recalled) with the two items presented on screen (see Scoboria et al., 2014).

Decision Phase. Finally, participants went through the decision phase. For each trial, two pictures or two circles appeared left and right on the screen. Participants were asked to choose the picture that they thought would win them money. Each trial consisted of two pictures from the same category (e.g., beach vs. lake or leg vs. arm), and one picture (S1+ picture) had been associated with a rewarded circle while the other (S1- picture) had been associated with a non-rewarded circle in the association phase. The rewarded circle and non-rewarded circle were presented in another trial to assess reward learning. In the decision phase, the same two stimuli were presented for four times, with each stimulus randomly assigned to the left or right side. There were 4 categories \times 2 pairs \times 4 times = 32 critical trials in total, which were presented in random order.

Two filler pictures that were never paired with any circles were also presented together with S1- pictures in the decision phase, to assess a baseline non-associated decision preference. Note that as decision choice might be impacted by the stimuli, materials were counterbalanced in a way that S1+ and S1- stimuli were exchanged in around half of the participants (53.6%; $n=22$). Pre-liking data also revealed no significant liking difference for S1+ ($M= 4.34$, 95%CI [4.05, 4.63]) and S1- ($M= 4.23$; 95%CI [3.94, 4.51]) stimuli, $t(41) = 1.01$, $p = .32$.

Results

We will first report analyses on data of the feedback phase on participants' memory ratings about their learned associations. We examined whether their recollection and belief ratings on the memory associations were undermined by the false feedback. Then we will discuss analyses concerning the final decision-making phase including the percentage of times participants chose rewarded circles and corresponding associated pictures (i.e., S1+) to win money as well as decision preference scores of choosing a certain picture in the decision phase. Our interest was whether the nature of the feedback (true vs. false) would impact participants' choosing rates and preference scores of S1+ pictures.

Feedback phase: Baseline memory data and manipulation check

In the memory test, participants first were asked to choose the patterned circle that they recalled was associated with the picture. Memory accuracy for associations pre-false feedback ($M= 0.68$, 95%CI [0.57, 0.79]) did not differ significantly from the memory accuracy for associations pre-true feedback ($M= 0.77$, 95%CI [0.65, 0.89]), $t(40) = -1.13$, $p=.27$, indicating equivalent levels of associative memories in the two conditions pre-feedback.

After feedback provided in the memory test, participants rated their recollections and beliefs for the S1+ associations. As Figure 2 shows, false feedback led to statistically lower recollection ($t(40) = -5.26$, $p < .001$, Cohen's $d = -0.84$) and belief ratings ($t(40) = -7.11$, $p < .001$, Cohen's $d = -1.11$) than true feedback. False feedback manipulation worked at both belief and recollection dimensions, suggesting that it substantially weakened memory associations. Using a cut point of 4.5 to dichotomously categorize recollection versus no recollection and belief versus no belief in 1-8 scales, Figure 2 shows that mean ratings for associations in the true

feedback condition fell into the upper range (4.5-8) while mean ratings for associations in the false feedback condition fell into the lower range (1-4.5).

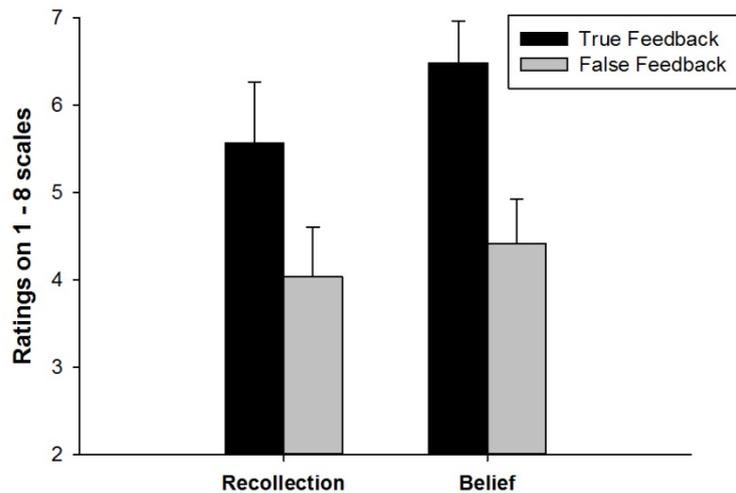


Figure 2. Recollection and Belief ratings after True and False feedback. Error bars represent 95%CI.

Decision-phase: Decision-making rates

Since preconditioning effect is limited by the conditioning effect (i.e., reward learning), data per comparison (S1+ vs. S1-) in the decision phase were included only when participants learned the corresponding reward association. For example, if a participant did not learn which circle would win them money in the first place, there was unlikely any spreading of preference from the circle to the picture. Five participants did not learn any reward association in the False feedback condition (not greater than chance level) and four participants did not learn any reward association in the True feedback condition. Thus, thirty-two participants' data were included in the final set of analyses. To be noted, the results did not differ with or without exclusion of participants.¹ Post-hoc power analysis calculated the power as 0.83 when sample

¹ Even if we included all 41 participants data without controlling for reward learning, similar results were achieved: False feedback condition had significantly lower rate of choosing S1+ than True feedback condition, $t(40) = -2.24, p = .03$, Cohen's $d = -0.55$. However, controlling reward learning is

size was 32. Those participants chose on average 97.66% (95%CI [0.96, 1.00]) of the rewarded circles in the False feedback condition and 96.48% (95%CI [0.94, 1.00]) of the rewarded circles in the True feedback condition. Paired samples *t*-test revealed no statistically significant difference, $t(31) = 0.83$, $p = .41$, implying similar (near ceiling) reward learning in the two conditions.

We were most interested in the decision making on S1 stimuli when true and false feedback were provided to participants' memory associations. We calculated the average rate of choosing S1+ over S1- in four rounds for each participant as per Wimmer and Shohamy's (2012). In the baseline decision-making condition (filler vs. S1-), participants chose the filler pictures at an approximately chance level ($M = 0.49$, 95%CI [0.36, 0.62]). In the True feedback condition, participants chose S1+ in 66.41% (95%CI [0.54, 0.79]) of the choices, which was statistically larger than 50% chance level, $t(31) = 2.62$, $p = .01$, and statistically larger than the baseline condition, $t(31) = 2.80$, $p = .009$, Cohen's $d = 0.50$. Thus we have replicated the sensory preconditioning effect in the true feedback condition. However, in the False feedback condition, participants chose S1+ only in 40.23% (95%CI [0.28, 0.53]) of the choices, which was statistically lower than the True feedback condition, $t(31) = -2.74$, $p = .01$, Cohen's $d = -0.74$, and not different from the baseline condition, $t(31) = -0.98$, $p = .33$.

Decision-making preference scores

To better illustrate decision preferences over S1+ or S1- in the True vs. False feedback conditions, we calculated participants' average decision preference scores.

Decision preference score for pictures was calculated by subtracting times of

cleaner, more honest way of analyzing the data (i.e., the effect might be due to different levels of reward learning), and thus we mainly report the results based on the conditionalized data.

choosing S1- stimuli from times of choosing S1+ stimuli over four rounds. A positive value indicates that participants preferred choosing S1+ stimuli over S1- stimuli; a negative value indicates participants preferred S1- stimuli; 0 value means no preference over S1+ or S1- stimuli (at 50% chance level). The preference score ranged from - 4 to 4. For example, a value of 4 means that the participant chose 4 times of S1+ and 0 time of S1- picture. Paired samples *t*-test showed that the mean preference score in the True feedback condition ($M= 1.44$, 95%CI [0.47, 2.40]) was significantly higher than that of the False feedback condition ($M= -0.78$, 95%CI [-1.80, 0.24]), $t(31) = -2.99$, $p = .005$, Cohen's $d = -0.53$. Figure 3a demonstrates that True feedback led participants to preferably choose S1+ stimuli (i.e., the preconditioning effect), while false feedback led them to show no preference over S1+ or S1- stimuli (i.e., the preconditioning effect disappeared).

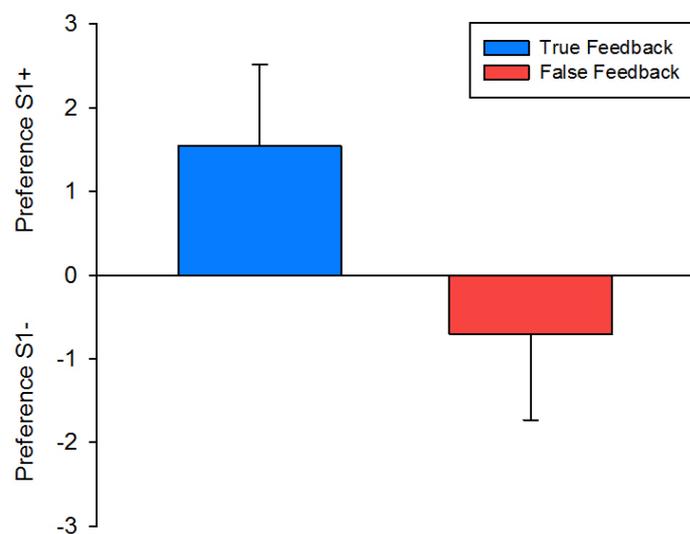


Figure 3a. Mean decision preferences in False and True feedback conditions. Decision preference score = (Times of choosing S1+) – (Times of choosing S1-). E.g., a value of 0 means no preference and a value of 2 means participants averagely chose 2 more

times of S1+ in four rounds (i.e., choosing S1+ 3 times and choosing S1- once). Error bars represent 95%CI.

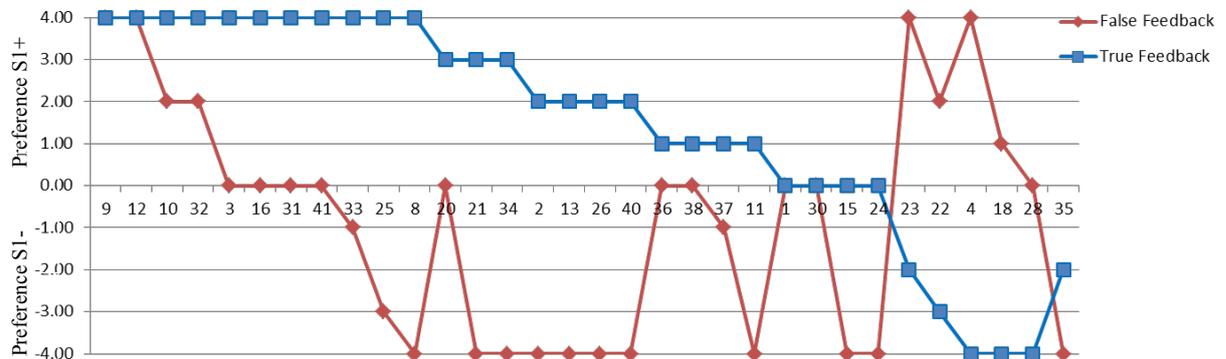


Figure 3b. Individual data of decision preferences in False and True feedback conditions. Each number on the X-axis represents one participant.

A recent study suggests that individual data should be presented with small sample sizes (Weissgerber, Milic, Winham, & Garovic, 2015). Figure 3b illustrates individual data of decision preference in true and false feedback conditions. It shows that decision preference was even reversed in the False feedback condition compared to the True feedback condition at an individual level. When a participant preferred S1+ in the True feedback condition, this participant preferred S1- (or not preferring S1+) in the False feedback condition; and when a participant preferred S1- in True feedback condition, she/he was more likely to choose S1+ stimuli in the False feedback condition. We separated participants into two groups, those who preferred S1+ stimuli in the True feedback condition (i.e., preference score > 0; $n = 22$) and those who preferred S1- stimuli in True feedback condition (i.e., preference score < 0; $n = 6$). Note that all these participants learned the reward conditioning very well (*Mean rate of choosing rewarded circles = 97%*; 95%CI [94%, 100%]). Paired samples *t*-test compared decision preferences between False versus True feedback in

each group: for the former group, $t(21) = -7.90, p < .001$, Cohen's $d = -1.96$; for the latter group, $t(5) = 3.20, p = .02$, Cohen's $d = 1.39$.

Decision-making latencies

Reaction times (RTs) for S1+ versus S1- choices were recorded during the decision phase. After RTs < 300 ms or > 10000 ms were excluded (2.02% of RTs data) based on general rules in reaction time tasks such as the implicit memory test (Greenwald, Nosek, & Banaji, 2003), RTs were averaged in each condition for each participant. A paired samples t -test was conducted to examine whether false feedback impacted decision-making speed. No statistically significant difference in reaction times was found between False feedback ($M = 1802$ ms, 95%CI [1539, 2066]) and True feedback ($M = 1782$ ms, 95%CI [1494, 2071]) conditions, $t(40) = 0.17, p = .86$.

Discussion

This is the first study to manipulate memory associations in a reinforcement decision-making task. The results showed that false feedback to participants' memories resulted in decreased recollection and belief ratings for their memory associations. That is, after being told that they misremembered two items paired together, participants indicated that they recollected the association to a lesser extent and believed less in their memory associations than when receiving true feedback. Thus, the false feedback manipulation successfully attenuated participants' memory associations, suggesting the strength of memory associations between stimuli can be impacted by false feedback.

Previous research on false feedback and memory has mostly focused on how feedback can change people's beliefs in a memory while recollections are retained (Otgaar, Scoboria, & Mazzoni, 2014). For example, developmental psychologist Jean

Piaget vividly remembered that he was kidnapped as a child but later he was told by his nurse that she fabricated the story. He no longer believed in his memories but he could recall vivid recollections of the kidnap event (Mazzoni, Scoboria, & Harvey, 2011).

In our study, participants not only reported lower beliefs in their memory associations but also reported lower recollections of these associations after receiving false feedback. The reason might be that memory associations formed in the current study were temporal in nature. That is, the presence of an S1 picture always preceded the presence of a circle to form an association. Mazzoni et al. (2014) found that, unlike sensory details, time characteristics of a memory can be significantly impacted with false feedback, which explains why temporal associations in our study can be easily deconstructed. Another reason why participants' memory associations were greatly impacted by false feedback may be that participants established relatively weak memory associations in the association phase (e.g., memory accuracy was around 0.68 - 0.77 pre-feedback), which may make memories easy to manipulate. Although we have followed a standard sensory preconditioning procedure as in previous research (Wimmer & Shohamy, 2012), future research may want to study the impact of feedback on strong memory associations by increasing the stimuli exposure or telling participants explicitly to memorize the associations.

This is also the first study showing that feedback on how memory associations are formed can impact reinforcement decision-making. Intriguingly, the most novel finding is that false feedback changed participants' decision preferences compared to true feedback. Participants exhibited the classical sensory preconditioning effect after receiving true feedback in that they chose S1+ stimuli more often over S1- stimuli to gain money. However, after they received false feedback telling them that certain

associations were unfounded, they no longer preferred S1+ stimuli and showed no sensory preconditioning effect anymore. Individual data showed that some participants even preferred S1- to gain monetary reward in the false feedback condition. Perhaps because we switched the paired circles for S1+ and S1- stimuli in the false feedback, some participants might have formed a false memory that the S1- picture was associated with a rewarded circle and thus demonstrated preferences over S1- stimuli instead of showing no preference during decision making.

Recent research has shown that reminders of particular memories can greatly guide people's decision making (Ludvig et al., 2015; Bornstein et al., 2017). Our study showed that feedback on memories about past experiences can change reward-based decision-making. All these studies suggest that how organisms remember their experiences (at the time of making a decision) is a key mechanism underlying decision making or general reward learning. More specifically, our results can be readily explained by memory mechanisms such as spreading activation theories (Anderson, 1983; Howe et al., 2009; Roediger et al., 2001). From a memory network perspective, memory consists of mental representations of stimuli (i.e., nodes) that can be associated with each other. In the true feedback condition that exhibited a typical sensory preconditioning effect, an S1+ stimulus was encoded to associate with a circle that was encoded to associate with a reward. Thus, associations were established between an S1+ picture and a circle and also between the circle and reward in a participant's memory network. According to the spreading activation principle, activation of a node can spread along its associations to nearby nodes automatically (Anderson, 1983; Howe et al., 2009). Although the S1+ stimulus was never rewarded, the activation of the reward can be automatically spread via the rewarded circle to the S1+ stimulus during decision making.

However, by providing false feedback to participants' memories, their memories were re-assessed and might be deconstructed, consistent with numerous studies that have shown the flexibility and malleability of our memory system (e.g., Loftus, 2005; Schacter, 2012). Indeed, our results showed that participants' memories for the associations between S1+ stimuli and rewarded circles were significantly undermined. Thus, when an S1+ picture was presented in the decision phase, the value of reward might not be activated since the association between S1+ and the rewarded circle was weakened, as demonstrated by the memory ratings data. As a result, participants did not show any preference to S1+ stimuli.

Previous research has also considered the role of memory structure and spreading activation in decision making (e.g., Morewedge & Kahneman, 2010; Weber & Johnson, 2006). Our explanation shares a lot of similarity with Morewedge and Kahneman's (2010) proposition that principles of associative memory activation should apply to judgment processes. They propose that strongly activated information is likely to gain more weight than information that is not activated in associative memory network during decision making, which can explain why reminders of a specific memory (i.e., stronger activation) determine later decision making. In the current experiment, we found that false feedback led to change in memory ratings and later led to change in their decision preferences, which aligns with the principle of spreading activation along memory associations. Of note, what we measured were participants' subjective ratings of their memory associations instead of objective measurements. We did not ask participants to take a recognition test again following the feedback because there might be a confirmatory effect that participants just confirm the feedback. Although most human memory research so far has focused on subjective or self-reported measurements, objective measurements of memory such as

physiological measures might be helpful in uncovering the exact mechanisms underlying memory and decision making in future research.

There are also some alternative mechanisms that may account for the current results. One possibility is the episodic recall of the false feedback during the decision phase. When participants encountered pictures that received false feedback from the memory test, they might recall a memory of the false feedback and thus select the alternative picture based on the feedback that is memorized. This mechanism suggests that the originally learned memory associations were not changed (although they were attenuated as demonstrated by memory ratings data) but co-existed with memories of the feedback phase. Another potential mechanism is that the sensory preconditioning effect may be driven by some metacognitive processes such as belief-based judgement and discounting. The information or beliefs generated from false feedback may have overridden the learned memory association by the time of making a decision and thus false feedback became a dominant factor influencing decisions. According to a metacognitive model of autobiographical memory and belief by Mazzoni and Kirsch (2002), belief is a metacognitive judgment of the occurrence of an event and it has been found to be a primary factor in shaping behavior. For example, a mega-analysis of eight studies has found that belief in a food aversive event (e.g., being sick after eating strawberry ice cream) but not recollection of the event predicted eating behavior (Bernstein et al., 2015). In our study, we observed a decreased belief in stimuli associations after false feedback was provided, which may explain the disappearance of the sensory preconditioning effect later. Notably, it remains difficult to ascertain the exact component of memory (recollection or belief) that led to the results since both memory components were undermined by false

feedback. Nevertheless, our study illustrated that memory in general impacted decision making.

Taken together, the current study has implications for theories of reinforcement decision-making in general. Research on reinforcement learning usually focuses on algorithms such as the model-free and model-based approaches, where organisms compute possible values for different choices and choose the option leading to greatest expected value (O'Doherty, Cockburn, & Pauli, 2017). Gershman and Daw (2017) pointed out that the computational perspective is limited in mimicking situations relevant for real life (e.g., facing a situation that has never been encountered), and they have proposed an episodic reinforcement learning model where episodic memory serves as a nonparametric estimation of the value in the algorithms. The PAM theory also suggests that decisions are made by retrieving information from memory in order to determine the best option (Weber & Johnson, 2006). Although the current study examined the role of memory from a different approach by manipulating memories, the results support the recent theoretical view that episodic memory should be integrated within the reinforcement decision-making system. That is, principles of memory (e.g., memory activation, false memory effect) may account for decision making. For example, participants who were implanted with a false memory that they used to get sick with peach yogurt showed lower preferences for peach yogurt (Scoboria et al., 2008). Further research is needed to investigate the role of memory in decision making.

Some might argue that the inclusion of a memory test would change the nature of the sensory preconditioning task. However, results in the true feedback condition still revealed a sensory preconditioning effect while the baseline condition revealed no effect, which suggests that the inclusion of a memory test did not impact the

sensory preconditioning effect. Some might wonder whether the experimental effect was due to demand characteristics that participants purposively selected S1+ stimuli in the true feedback condition and purposively showed no preference in the false feedback and baseline conditions. There are several reasons why this is unlikely to be the case. Demand effects refer to participants' change of behavior by discovering the experiment's purpose. However, it would be difficult for participants to find out in the decision phase which trials referred to the true feedback condition, which referred to the false feedback condition, and which referred to filler pictures since there were 112 trials in total mixed across experimental conditions and filler pictures. We have also switched the S1+ and S1- stimuli in half of the participants, so during the decision-making phase, both the experimenter and the participant did not know exactly which stimuli were S1+ or S1- and which stimuli had received true or false feedback (one can only know by checking the program). What's more, Wimmer and Shohamy's (2012) study suggests that the transference of value from a rewarded circle to the S1+ picture is out of participants' awareness. Participants in our study made a decision relatively fast (in around 1.8 s) and RTs in the true feedback condition did not differ from RTs in the false feedback condition. Nevertheless, it is unknown from the current study whether participants had explicitly referred to their memories or if the spreading of preference via associations was automatic. Further research is needed to closely examine participants' decision-making process in the preconditioning task.

To conclude, our research has presented evidence that false feedback provided to participants regarding their memories led to participants no longer showing decision preferences to gain reward. The results suggest that episodic memories may play an important role in reinforcement decision-making. It is time now to investigate how principles of memories may impact reinforcement decision-making.

Acknowledgement

We want to thank Sofie Dings and Yiru Shao for their help in data collection.

Supplementary materials

All materials and data can be accessed at <https://osf.io/y8hbp/>.

Funding

This paper has been supported by a fellowship from the China Scholarship Council (CSC) to Jianqin Wang [No. 201406100057], by a grant from the Ministry of Education in China (MOE) Project of Humanities and Social Sciences [No. 11YJA190026] and by a grant from the Shanghai Pujiang Program [No. 15PJC015] to Chu Zhou.

References

- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22, 261-295.
- Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, 18, 279-292.
- Bernstein, D. M., Laney, C., Morris, E. K., & Loftus, E. F. (2005). False beliefs about fattening foods can have healthy consequences. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 13724-13731.
- Bernstein, D. M., Laney, C., Morris, E. K., & Loftus, E. F. (2005). False memories about food can lead to food avoidance. *Social Cognition*, 23, 11-34.
- Bernstein, D. M., Scoboria, A., & Arnold, R. (2015). The consequences of suggesting false childhood food events. *Acta Psychologica*, 156, 1-7.
- Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, 8, 15958.
- Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, 25, 323-332.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Geraerts, E., Bernstein, D. M., Merckelbach, H., Linders, C., Raymaekers, L., & Loftus, E. F. (2008). Lasting false beliefs and their behavioral consequences. *Psychological Science*, 19, 749-753.

- Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology, 68*, 101-128.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.
- Howe, M. L., & Otgaar, H. (2013). Proximate mechanisms and the development of adaptive memory. *Current Directions in Psychological Science, 22*, 16-22.
- Howe, M. L., Wimmer, M. C., Gagnon, N. & Plumpton, S. (2009). An associative activation theory of children's and adults' memory illusions. *Journal of Memory and Language, 60*, 229-251.
- Kimmel, H. D. (1977). Notes from "Pavlov's Wednesdays": sensory preconditioning. *American Journal of Psychology, 90*, 319-321.
- Lind, M., Visentini, M., Mäntylä, T., & Del Missier, F. (2017). Choice-supportive misremembering: A new taxonomy and review. *Frontiers in Psychology, 8*, 2062.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*, 560-572.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361-366.
- Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2015). Priming memories of past wins induces risk seeking. *Journal of Experimental Psychology: General, 144*, 24.

- Mazzoni, G.A.L., Clark, A., & Nash, R.A. (2014). Disowned recollections: Denying true experiences undermines belief in occurrence but not judgments of remembering. *Acta Psychologica, 145*, 139–146.
- Mazzoni, G., & Kirsch, I. (2002). Autobiographical memories and beliefs: A preliminary metacognitive model. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 121–145). Cambridge, UK: Cambridge University Press.
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Hove, England: Psychology Press.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.
- Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences, 14*, 435-440.
- O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual Review of Psychology, 68*, 73-100.
- Otgaar, H., Muris, P., Howe, M. L., & Merckelbach, H. (2017). What drives false memories in psychopathology? A case for associative activation. *Clinical Psychological Science, 5*, 1048-1069.
- Otgaar, H., Moldoveanu, G., Wang, J., & Howe, M. L. (2017). Exploring the consequences of nonbelieved memories in the DRM paradigm. *Memory, 25*, 922-933.
- Otgaar, H., Scoboria, A., & Mazzoni, G. (2014). On the existence and implications of nonbelieved memories. *Current Directions in Psychological Science, 23*, 349–354.

- Otgaar, H., Scoboria, A., & Smeets, T. (2013). Experimentally evoking nonbelieved memories for childhood events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 717–730.
- Otgaar, H., Wang, J., Fränken, F., & Howe, M. L. (2018). Believing does not equal remembering: The effects of social feedback and objective false evidence on belief in occurrence, belief in accuracy and recollection. *Acta Psychologica*, *191*, 271-280.
- Roediger, H. L., Balota, D., & Watson, J. (2001). Spreading activation and arousal of false memories. In H. Roediger, J. Nairne, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder. Science conference series* (pp. 95–115). Washington, DC: American Psychological Association.
- Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on Psychological Science*, *1*, 277-311.
- Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*, 603-613.
- Scoboria, A., Jackson, D., Talarico, J., Hanczakowski, M., Wysman, L., & Mazzoni, G. (2014). The role of belief in occurrence within autobiographical memory. *Journal of Experimental Psychology: General*, *143*, 1242-1258.
- Scoboria, A., Mazzoni, G., & Jarry, J. L. (2008). Suggesting childhood food illness results in reduced eating behavior. *Acta Psychologica*, *128*, 304-309.
- Shohamy, D., & Daw, N. D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences*, *5*, 85-90.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*, 1-25.

- Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919-934.
- Wang, J., Otgaar, H., Bisback, A., Smeets, T., & Howe, M. L. (2017). The consequences of false recollections and beliefs on food preferences. Annual conference of European Association of Psychology and Law, Mechelen, Belgium, 2017.
- Wang, J., Otgaar, H., Howe, M. L., Smeets, T., Merckelbach, H., & Nahouli, Z. (2017). Undermining belief in false memories leads to less efficient problem-solving behaviour. *Memory*, *25*, 910-921.
- Weber, E. U., & Johnson, E. J. (2006). Constructing preferences from memory. In Lichtenstein, S. & Slovic, P. (Eds.), *The Construction of Preference* (pp. 397–410). Cambridge University Press.
- Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS biology*, *13*, e1002128.
- Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*, 270-273.
- Zajac, R., & Henderson, N. (2009). Don't it make my brown eyes blue: Co-witness misinformation about a target's appearance can impair target-absent line-up performance. *Memory*, *17*, 266-278.

Appendix

