Falska minnen som uppkommit från en beslutsblindhetsuppgift formger framtida politiska attityder

False Memories Resulting from a Choice Blindness Task Shapes Future Political Attitudes

David Bengtegård Book

handledare / supervisors

Christian Balkenius
Philip Pärnamets

KOGM20

2019-06-03

Masteruppsats i kognitionsvetenskap
Avdelningen för kognitionsvetenskap
Filosofiska institutionen
Lunds universitet

Master’s thesis (2 years) in Cognitive Science
Lund University Cognitive Science
Department of Philosophy
Lund University

False Memories Resulting from a Choice Blindness Task Shapes Future Political Attitudes

David Bengtegård Book

psy14dbe@student.lu.se

In many attitude theories, it is commonly assumed that what we believe in is partly based on our own past actions, and that these actions shape our present opinion towards an issue. This suggests that how one remembers and represents past decisions could have an instigating role in establishing future attitudes. However, the way attitudes change over time has generally been explained by either self-perception processes or from resolving internal motivational conflicts. The aim of this thesis is to go beyond this conception of attitude change and explore an alternative explanation: that attitudes are liable to the dynamics and processes of memory. To do this, participants stated their opinions on political issues, and the Choice Blindness Paradigm was used to manipulate some of their previous responses to indicate an opposite position. Participants were then asked to remember their previous responses together with their current opinion on the issue directly after the manipulation and one day later to investigate how memories of past attitudes are influenced when accepting the false feedback. Specifically, to test whether the choice blindness manipulation creates a false memory of a past attitude which participants' uses when generating their future response on a political statement. The result showed that participants' memory responses were strongly influenced by the manipulation and moved in direction of the false feedback, both directly following the manipulation as well as one day later. This effect was also found for attitude responses in which participants exhibited lasting shifts in attitudes. Additionally, the memory of past attitudes was a significant predictor for later attitude shifts and explained a large portion of variance in attitude change. These findings provide evidence that attitude change as well as choice blindness may result from memory mechanisms. And helps to understand how environmental forces and memory processes can interact in shaping future attitudes.

1 Introduction

A centerpiece in human socialization and development is learning - the ability to gather knowledge about how the world works through personal experience. Such learning can be seen functionally as allowing people to adjust their actions and attitudes to the shifting environment and illustrates the powerful role of the self in orchestrating perceptions, beliefs, and memories of reality. But learning is not only of instrumental value – it is a source of intrinsic psychological fulfillment which shapes the way we are and what we believe in. Many attitude theories hold that our current attitudes are, at least in part, based on our own past actions (Bem, 1967; Festinger, 1957; Lord & Lepper, 1999). These theories maintain that when reporting one's assessment of a given attitude or preference, an individual looks back at the actions she has taken in the past, and that these actions shape the current attitudes. This suggests that people understand their present and future in how they conceptualize and remember the past, as vividly put in the chilling novel Nineteen Eighty-four by George Orwell: "Who controls the past controls the future. Who controls the present controls the past". An intriguing question stemming from this is whether our conception of ourselves is shaped by the memories of the choices we make. For example, could a person's political attitude be altered by inducing false memories of past decisions, and if so, what might the mechanisms be?

Traditionally, research in economics and behavioral decision-making has generally overlooked the nature and role of memory mechanisms in preference construction, treating preference and choice in an "as-if" fashion formalized through axiomatic theories, grounded in psychophysical principles rather than on the underlying psychological mechanisms (Thaler, 2016). Psychologists have of course long known the fallibility of human memory, particularly people's tendency to misremember or fabricate memories of the past as consistent with the present selves (Ross, 1989). For example, in an experiment carried out by Goethals and Reckman (1973), high school students incorrectly recalled past political attitudes as highly similar to present ones, when, in fact, they only distorted their recall on their initial stand as to make it more consistent with their new attitude. Such congruency (or consistency) effects have often been explained by cognitive dissonance or self-perception theories. According to the phenomenological approach typified by the dissonance theory, a person holding two conflicting opinions will experience an aversive motivational state called cognitive dissonance which she will attempt to remove by altering the two "dissonant" opinions and this will result in a change in attitude (Festinger, 1957). Self-perception theory, which stems from the legacy of behaviorism, would instead argue that people tend to infer their intentions by observing their own behavior and from this attitude change will be induced without any form of direct introspection (Bem, 1967).

However, a new perspective emerging in preference and decision sciences that attempts to increase the psychological realism, are turning to memory for explanations of preference construction (for a review illustrating this point, see Weber and Johnson, 2006). For instance, Ariely and Norton (2008) challenge the neo-classical economics view that behavior is driven by and reflective of hedonic utility by providing empirical evidence that irrelevant situational factors (e.g. background music) and memory of past actions can shape and bias future attitudes and their perceived utility. A growing number of decision scientists have thus started to draw on the rich supply of prior memory research and stressed the effects of multiple memory representations and implicit memory processes during a choice situation (Reyna, 2012; Wimmer & Sohamy, 2012). One way to disentangle the question of whether memory processes affect attitude formation is by using the choice blindness paradigm (CBP) which the present thesis did. The underlying methodology of CBP is that it allows the experimenter to create a dissociation between decisions and outcomes in a simple choice situation (Johansson, 2006). In the original choice blindness study (Johansson et al., 2005), participants were shown two pictures of female faces and were instructed to point at the most attractive one, and then explain why they preferred that picture over the other. However, in some trials, the opposite outcome was presented to the participant as their actual choice. Interestingly, the results showed that participants often failed to detect the manipulation (no more than 26%) and instead accepted the false feedback as being their own intended choice (i.e. the choice blindness effect), and even gave verbal explanations for why they preferred the manipulated outcome (Johansson et al., 2005).

More recent studies using the CBP as an experimental format for studying attitude change have shown that false beliefs about past attitudes can lead to persistent changes in political attitudes (Strandberg et al., 2018). Other work on choice blindness has also found that when participants accept the false feedback about past choices, it leads to systematic distortions in source memory (Pärnamets et al., 2015). Beyond these studies, little work has sought to identify the effects choice blindness manipulation has on the memory of past attitudes. This thesis primary goal is therefore to investigate an underexplored aspect of the CBP: how individual memories of past attitudes are influenced when accepting the false feedback and which downstream effects this has on later attitude formation. More specifically, to test whether the choice blindness manipulation creates a false memory of a past response that, in turn, shapes future responses. The work in this thesis will consequently complement earlier studies of CBP by considering the possibility that attitudes are liable to the processes and dynamics of memory which might play an instigating role in driving attitude change. This will also lead to a better understanding of memory involvement in decision-making which is especially important because emerging evidence suggests that decisions are often guided by the memory of past experiences (Ariely & Norton, 2008; Carpenter & Schacter, 2017; Wimmer & Sohamy, 2012). For this reason, it is well motivated to examine the link between memory and attitudes in order to determine the cognitive mechanisms underlying choice blindness and attitude change.

The outline of the thesis is as follows: First, a theoretical discussion about the nature of memory is given and its implications for false memories and attitude change. Second, a memory-based model of attitude formation is applied to explain the effects of choice blindness. Thereafter, I briefly review some studies investigating memory involvement in decision-making as well as other related work. Following this, relevant studies regarding choice blindness, attitude change, and memory are summarized together with the memory mechanism explanation that may account for the choice blindness phenomena. The experiment's design and procedure are then described followed by an analysis of the results. Finally, a discussion of the results is presented and its significance in relation to previous literature along with a conclusion that will tie the thesis main findings together.

Constructive memory and reality

When looking at humans, memory underlies the ability to mentally relive past experiences from distinct episodes as well as to project oneself into possible futures (Tulving, 1972). Tulving (1983) further suggested that episodic memory is associated with autonoetic (self-knowing) consciousness: the feeling that past experiences belong to one's own history and enables one to re-experience and pre-experience episodes from a first-person perspective. In this sense, memory is much more than memorization, and cannot be viewed separately from the individual who encodes it (i.e. the memorizer). In other words, we don't *store* words or representations in our heads, because no single object or process can be isolated as a "sub-structure" within the person who retains external information (Milner, 1998). Instead of treating memory as an end to memorization, thus going against the traditional store-house metaphor of memory, this thesis assumes that memory is a fundamentally holistic, dynamic, and contextual system that represents information at multiple levels simultaneously. Correspondingly, some researcher argues that memory traces are better understood as "incomplete, partial, and context-sensitive, to be reconstructed rather than reproduced" (Sutton p. 229, 2009). Thinking about memory in this way is valuable because it emphasizes that remembering is a process of reconstruction, and more often captures the gist rather than the details of the original experience (Reyna, 1997).

Furthermore, considerable evidence has demonstrated that human memory is not an exact replica of past experiences but is instead characterized as an imperfect process that is prone to various kinds of errors and distortions (Bartlett, 1932; Schacter, 2011). The question whether these memory distortions reflect a fundamental deficiency in our cognitive processing is a topic that has raised different opinions, especially because inaccurate or false memories might tell us something about the nature and function of memory. Bernstein and Loftus (2009) clearly express that memory is flawed and should be regarded as an unreliable system: "In essence, all memory is false to some degree. Memory is inherently a reconstructive process, whereby we piece together the past to form a coherent narrative that becomes our autobiography. In the process of reconstructing the past, we color and shape our life's experiences based on what we know about the world" (Bernstein & Loftus p. 373, 2009). On the contrary, more recent accounts of memory distortions instead argue that this defect reflects an adaptive constructive process that contributes to the efficient functioning of several cognitive functions (Schacter, 2011), such as solving problems (Howe, Gardner, Charlesworth, & Knott, 2011), memory updating (Hardt et al., 2010), and extracting gist or meaning (Brainerd & Reyna, 1990). However, the same flexible retrieval processes that underpin these adaptive functions will also, as a consequence, produce various distortions in memory (Carpenter & Schacter, 2017). Thus, instead of thinking that false memories reflect a fundamental flaw in human memory, it is more instructive to view such errors as a by-product of adaptive processes that in effect improve memory performance in the long run.

Memories of past decisions, such as taking a position on a political statement, will unavoidably be affected by these flexible reconstructing processes regardless of whether they are adaptive or detrimental. This is the case for the current thesis, which looks at attitude change as a product of a well-functioning adaptive memory system. The key lesson from adopting this perspective on memory and attitudes is that it emphasizes the constructive nature of cognitive processes, how they develop and acquire meaning from personal life experiences and how this can provide us with a richer understanding of the mind. Taking on a constructivist epistemology, that is, an epistemology centered upon the active participation of the subject in construing reality, invites us to look at attitudes and decisions in a more dynamic way and analyze their development over time.

A memory-based account for choice blindness

Query theory (QT) is a psychological, memory-based model of preference (or attitude) formation which maintain that preferences, like all knowledge, are liable to the processes and dynamics of memory (such as encoding and retrieval) and has been used to explain choice biases in various decision-making contexts (Johnson et al., 2007; Weber et al., 2007). This theory thus proposes that preferences are constructed, rather than being pre-stored and fixed, by arguing (or interrogating) with ourselves internally. In other words, during a judgment or choice situation, people will consult their memory (or the external environment) by asking oneself questions, or queries, about the attributes of choice alternatives (particularly their benefits and downsides). This means that choice tasks often involve implicit generation of arguments for different courses of action (queries). For instance, if someone asked you to choose a favorite movie from a choice of three, you would probably consult with your memory about previous experiences with the same or similar movies in order to come up with an answer. These queries are typically grouped by valence, for example, memory is first queried about what you like about a certain movie, and only after no additional positive attributes are generated can a query about negative attributes arise.

QT makes four key assumptions regarding implicit memory-retrieval and argument-integration processes people use to evaluate choice options: (1) Decision makers query past experience for evidence supporting different choice options, (2) these queries are executed sequentially, and (3) the first query produces richer memory representations because of output interference. This occurs because when evidence for the first option is generated, evidence supporting other choice options is temporarily suppressed (or inhibited). Lastly, (4) choice follows from the resulting balance of evidence supporting the option. And because the order of option influences the balance of evidence, it is important to know which choice options that get queried first.

As mentioned in the Introduction, attitude change including the choice blindness phenomena has usually been interpreted by the self-perception theory (Johansson et al 2005; Johansson et al., 2014; Strandberg et al., 2018). From this perspective, attitude formation is supported by a process of self-perception, that is, inferences about one's own attitudes stems from observing the outcomes of past behavior. In other words, we infer our own attitudes the same way we infer other peoples' attitudes, simply by observing and interpreting our own overt behavior (Bem, 1967). Consequently, when participants accept the choice blindness manipulation as their original attitude, they will truly believe that the manipulated outcome is their self-generated response. Previous studies using the CBP, therefore, argues that the experimental format genuinely creates a self-inferential situation and that attitude calculation often is supported by self-perception processes (Strandberg et al., 2018). Another compatible explanation that differs from the self-perception account is that the choice blindness manipulation creates a false memory of a past response, which is used by the participants when asked to generate a future response or when evaluating the choice options a second time. According to this attitude-as-memory perspective, the choice blindness phenomena is supported by implicit memory processes and argument-integration processes.

Although QT generally has been applied as a model to describe different choice biases, this thesis uses it to explain the effects of choice blindness as a product of memory processes. The reason for this is that QT postulates process-level specification of implicit memory retrieval effects, and thus makes explicit testable predictions. Based on the four key assumptions mentioned earlier, the following is predicted to happen if the participants accept the false feedback portion of choice blindness: When the participants are asked to state their memory of a previous response (e.g. the attitude on a political issue), the false (or competing) memory will temporarily suppress the original memory representation (because queries are executed sequentially), thereby biasing the participant's perception about the choice. This means that, if one accepts the choice blindness manipulation, it will become the first query because of output interferences and thus produce a richer memory representation compared to when detecting the manipulation. Furthermore, the evidence (or arguments) supporting the manipulated option will temporarily inhibit arguments for other choice options, thus forcing the participants to give arguments against their initially favoured option or opinion. As a result, if the participants believe that the manipulated outcome to be their own choice, arguments for their initial opinion will be suppressed. This thesis aims to test whether the choice blindness phenomena is supported by these implicit memory retrieval processes as well as argument-integration processes in arriving at a decision.

Memory involvement in decision-making

One important component which generally has been neglected in decision sciences is how memory processes might influence and guide different aspects of value construction during a choice situation. Although it is known that memory can support decisions by retrieving relevant information at the time of choice (Johnson & Redish, 2007; Preston et al, 2004), relatively little work have investigated alternative memory mechanisms that could explain why someone would pick one alternative over the other, especially between novel options that have never been directly experienced. Wimmer and Sohamy

(2012) sought to fill in this gap by considering that the hippocampus encodes associative relationships between items and events that appear together, allowing for the integration of old memories with new ones. The central idea was therefore that the hippocampus, traditionally known for its role in consolidating and contextualizing long-term memories, will enable spread of reward to associated items stored in memory and thereby bias decisions between new choice options.

To test this prediction, they used functional magnetic resonance imaging (fMRI) to measure brain signals during a simple learning and decisions task. In the experiment, participants were first exposed to a set of different pairs of pictures in order to build new associative memories. Following this, participants learned that some items were associated with a monetary reward through the procedure of classical conditioning. Finally, participants were asked to decide between two stimuli for a possible monetary win. As expected, the findings indicated that reward can spread to bias the value of the option that, in fact, were themselves never directly rewarded. Moreover, this decision bias was predicted by activity in the hippocampus as well as reactivation of preestablished networks of associated memories in the brain. Importantly, these findings suggest that memory involvement in value assignment might not be driven by conscious awareness, rather it seems more likely that the transfer of value by the hippocampus is automatic and implicitly guided (Wimmer & Sohamy, 2012). Additionally, this internal valuation bias may reflect why we are not always aware of the explicit reasons for holding a particular preference toward an object and hence not really knowing why we chose option B over option A.

Another possible explanation for not knowing the explicit reasons when deciding may derive from the fact that people are quite poor at calibrating (or estimating) their personal certainty of a statement and therefore often wrongly believe that they know the answer to a question (Fischoff et al., 1977). This means that people tend to be overconfident, that is, they exaggerate the extent to which what they know is correct. People's poor calibration may be, in part, just a question of faulty inferences from memory. Johnson-Abercrombie (1960) studies on inferential processes in perception shows that the validity of inference is usually not inquired into but will nevertheless be accompanied by a feeling of certainty of being right. If people believe that memories are an exact copy of their original experiences, thus being unaware of the fact that one often makes inferences and reconstructions during a choice situation, then it comes as no surprise that they do not evaluate their inferred knowledge (Fischoff et al., 1977). Additionally, studies have shown that many errors in memory for verbal material occur because people remember what was pragmatically implied (something neither explicitly stated nor necessarily implied) instead of what was directly stated (Harris & Monaco, 1978). In other words, people find it difficult to distinguish between assertions and inferences, which may lead to the overconfidence seen in the studies by Fischoff et al. (1977).

Choice blindness, attitude change, and false memory

Choice blindness (CB) refers to the finding that people can be blind to mismatches between their actions and outcomes during a simple choice situation and, in addition, support the opposite of the chosen alternative. Many studies have provided evidence that this experimental effect is robust and has been replicated across a variety of domains, such as financial decisions (McLaughlin & Somerville, 2013), speech intention (Lind et al., 2014), taste and smell preferences (Hall et al., 2010), moral attitudes (Hall et al., 2012), and voting intentions (Hall et al., 2013). Recent work on CB has further shown that when participants accept the false feedback as being one's own response it can lead to lasting changes in one's political attitudes (Strandberg et al., 2018). Interestingly, this suggests that not only do the manipulation induce short-term effects on the participant's ratings, it can also lead to more persistent changes so that participants are more likely to choose an option they previously received false feedback on. For example, a study by Johansson et al. (2013) who asked participants to rate the attractiveness of two female faces found that participants ratings of initially rejected faces, following the manipulation, are chosen more frequently when asked to decide again a second time. This means that participants were more likely to prefer the face they had been manipulated to believe they had originally chosen rather than what they had actually chosen. Furthermore, these findings indicate that choice blindness itself can serve as an influential environmental feedback mechanism that changes (or reverses) preferences through the act of choice.

However, the above results have on later accounts been linked to how *beliefs* about past choices shape our memories of past options which will mediate the change in future attitudes (Pärnamets et al., 2015; Strandberg et al., 2018). Pärnamets et al. (2015) tested this idea specifically by investigating participants memories for stimuli in a choice blindness task involving preferential choices between pairs of faces. In the first phase, participants were presented with two faces on a computer screen and were instructed to choose the one they preferred as well as which facial features that contributed most to their decision. However, in manipulated trials, the non-chosen face was presented to the participants as their original choice. Following an unrelated filler task, participants were shown either the originally chosen or non-chosen face together with a foil and asked to select which of these faces they had seen during the choice phase. After this recognition memory task, participants were instructed to indicate if the selected face was the face they had originally chosen (i.e. the source memory task).

The results showed that participants correctly recognized the target face in 89.4% of all trials and there was no difference in recognition memory accuracy between manipulated (91.2 %) compared to nonmanipulated trials (88.9%) as well as for detected (92.5%) and non-detected trials (89.1%). This indicates that the CB effect is not due to a failure to encode the choice options before accepting the manipulation, because recognition rates for non-detected versus detected trials do not differ. Furthermore, overall source memory accuracy was 74.5% for all trial types and lower in manipulated trials (61.8%) compared to nonmanipulated trials (78.1%). Importantly, source memory accuracy was much lower for non-detected trials (43.6%) compared to detected trials (72.6%). These result patterns clearly indicate that if one accepts the false feedback, it does not only affect later choices, but also explicit memories of those choices together with the beliefs about those choices (Pärnamets et al., 2015). In line with other work showing how beliefs at the time of retrieval can lead to selective memory biases as well as source memory distortions (Mather et al., 2000; Henkel & Mather, 2007), this supports the idea that beliefs about past decisions might play an important role in driving attitude change, instead of preferential adjustment through the act of choice.

Moreover, these studies provide evidence for what Lind et al. (2017) call choice-supportive false memories; when attributes or facts that are not part of the original decision are remembered as presented. In this sense, false memories can arise when making a choice or judgment and could be understood as a type of self-serving memory distortion. Previous research on false memories has shown that falsely remembered events can influence our future attitudes and our future decisions. For example, implanting false memories about a negative experience with a particular food (e.g. hard-boiled eggs) or a particular alcoholic beverage can result in a diminished liking of it in adulthood (Bernstein et al., 2005; Clifasefi et al., 2013). In the context of the choice blindness paradigm, there are some studies that have looked at the effects of accepting the false feedback on later memories. One study by Sagana et al. (2014a) tested whether memory impairment is a plausible candidate to explain choice blindness by asking participants' to numerically rate the sympathy of female faces, and then after a short interval to recall their original ratings. If memory impairment is sufficient to explain choice blindness, then the manipulated outcome should hinder the recollection of the original rating, because the presentation of new information will interfere with (or decrease) the accessibility of the original memory trace. However, the authors found that later recall accuracy of the sympathy ratings in both manipulated and nonmanipulated trials was unaffected by the false feedback, thus indicating that memory impairment (or memory decay) does not fully account for the choice blindness phenomena (Sagana et al., 2014a).

Choice blindness has also been examined for eyewitness identification decisions (Sagana et al., 2014b). In a series of experiments, participants first witnessed a number of mock-crimes videos and were afterward asked to identify the people involved in the crimes from a photo line-up. In some trials, the participants choice was manipulated so that they were confronted with an non-chosen face from the line-up. The results from the second experiment showed that choice blindness rates were low for concurrent detection (30%) and almost completely absent for retrospective detection (0-6%). To investigate whether the manipulation would affect future choices, the participants came back in a second session 48 hours later to make the same choices again. The manipulation did not, however, affect these future decisions, and none of the originally chosen faces from the line-up were consistent with the false feedback (Sagana et al., 2014b). On the other hand, more recent experiments on eyewitness identification have shown that choice blindness manipulation can have long-lasting effects. Cochran et al. (2016) found that eyewitnesses who fail to detect the manipulation are more likely to change their identification decision as well as their memories of episodic details in the direction of the false feedback when asked to perform the rating task a second time. In other words, if people fail to detect the manipulation about their own decisions as well as their own memory reports, their subsequent decisions and memories can become biased by the false feedback they receive.

To summarize, the aforementioned studies suggest that choice blindness is not due to a failure to encode the choice options properly (Pärnamets et al., 2015), nor from memory impairment *per se* (Sagan et al., 2014a), and there appears to be converging evidence pointing to its long-term effects on future decisions, attitudes, and memories (Cochran et al., 2016; Johansson et al., 2013; Strandberg et al., 2018). Another plausible explanation that therefore may account for the effects of choice blindness is that when accepting the false feedback, a false memory is created of what was previously answered, which is used by the participants when generating their future response on a statement. This means that false feedback does not only affect the probability of choosing a manipulated outcome more frequently or the beliefs about previous choices, but also explicit memories of those choices. Pärnamets et al. (2015) study provides some preliminary empirical support for this memory mechanism explanation when showing similar false memories for manipulated choices about faces but has not been directly tested before. The present thesis sought to test this memory mechanism explanation by looking at the relationship between memory and political attitudes in the context of a choice blindness experiment.

Hypotheses and purpose

The purpose behind this thesis is to replicate and expand previous findings on choice blindness by exploring how the memory of past attitudes is affected by receiving false feedback, but with new stimulus material (digital questionnaires) and modality (political attitudes). The most important extension in the experimental design is thus that participants are instructed to explicitly state their memory for previous responses, which has not been done in past studies (e.g. in Strandberg et al., 2018). This is significant because none of the reviewed studies on false memories assessed memory using cued recall. Furthermore, the experiment also includes non-target control items to determine how accurate memory and attitude rating is when no feedback is received, which will work as a baseline measure.

Building on the results of previous research has led to the following expected results and hypotheses. The first and main hypothesis (H1) is that during a choice blindness manipulation a false memory is created of what was previously answered, which, in turn, competes against the alternative option, thereby biasing individuals' perception about the choice. This will, as a result, induce a shift in later attitudes because participants will use the false memory of a past response when generating their future response on a political statement. From this, it is predicted that in accepted manipulated trials, the memory changes in the direction of the false feedback, and participants should therefore also exhibit high memory accuracy in detected trials. Furthermore, this memory effect should be persistent one day later after the initial manipulation, because previous studies have shown that false beliefs about past choices can induce lasting changes in political attitudes (Strandberg et al., 2018). The second hypothesis (H2) is that attitude change should be evident in accepted manipulated trials and move towards the false feedback, and participants should therefore also restate their original attitude correctly in detected trials. The third hypothesis (H3) is that attitude change, following from accepting the false feedback, will be positively correlated with memory change, because it is the (false) memory of past attitudes that drives the change in attitude. Moreover, the detection rates of manipulated responses are expected to fall near 50% as in previous studies (Johansson et al., 2005; Strandberg et al., 2018).

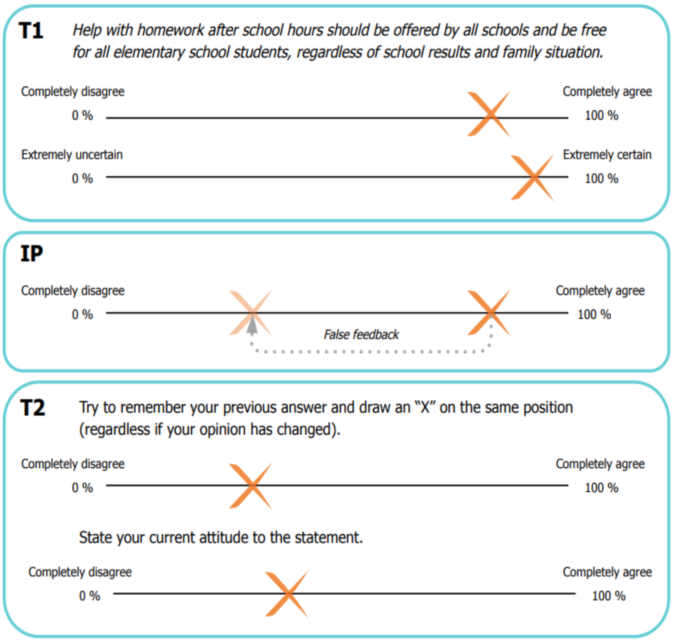
2 Method

Participants

A total of 163 participants were recruited from the student population at Lund University (92 females; *M* age = 22.8, *SD* = 3.2; age ranged from 19 to 39 years). Twenty-four participants were excluded from the analysis due to errors with the recorded data, technical malfunctions when interacting with the tablet application, or because of not showing up in the second session. In the final analysis, 139 participants were included. All participants received one cinema voucher as well as a gift card at Café LUX worth 40 SEK in exchange for their participation in two experimental sessions. The second follow-up session proceeded approximately one day after the initial session (*M* = 22.5 hours, *SD* = 1.9). Before starting the experiment, all participants received information about the outline of the experiment but were naïve about the actual purpose. Participation took place on the basis of volunteerism and that the participants had full right to cancel the experiment at any time for whatever reason. Anonymity was ensured through anonymized data collection. After completing the second-follow up of the experiment, all participants were thoroughly debriefed about the real purpose before signing informed consent.

Materials and design

Throughout the experiment, three digital questionnaires running on a tablet application were administered to the participants in order to register their political attitudes as well as their memory of past attitudes. The tablet application was designed specifically to give participants false feedback about their own rating on a political statement which was randomly displayed on the tablet one at a time (see Figure 1). All questionnaires consisted of twelve political statements in total and were divided into three main issues: education, environment, and healthcare. Of these twelve statements, eight were used across all three questionnaires, in which four were classified as target statements and the other four as non-target statements. Four new statements were also introduced in the two questionnaires administered after the first, which the participant had not previously taken a position on. All four target statements were randomly assigned as either manipulated (i.e. the false feedback) or non-manipulated, while non-target statements were used as a baseline measure to see how accurate the memory of past attitudes was when no feedback was received. An example of a target statement (translated from Swedish to English) which the participant had to answer concerning a proposed education policy was: “Help with homework after school hours should be offered by all schools and be free for all elementary school students, regardless of school results and family situation.”



T1 *Help with homework after school hours should be offered by all schools and be free for all elementary school students, regardless of school results and family situation.*

IP

T2 Try to remember your previous answer and draw an "X" on the same position (regardless if your opinion has changed).

State your current attitude to the statement.

Figure 1. An illustration of the tablet application used in the experiment. Participants first rate the extent to which they agree or disagree with a political statement as well as their level of confidence on a scale ranging from 0% to 100% (T1). After answering all 12 statements, participants are informed to go over four of the responses with the experimenter. At this interaction phase, the application moves two of the participants’ answers across the midline to the opposite side of the scale and randomly place them between 15% and 35%, or 65% and 85% (IP). Thereafter, participants are asked to state their memory of their previous response in T1 along with their current attitude on the issue (T2).

Furthermore, in order to capture the verbal reports and arguments given by the participant for holding a specific political opinion an audio recorder was used. The participants also performed a distraction task consisting of a Swedish translated pen and paper version of the Cognitive Reflection Test (CRT) used in Frederick (2005). CRT consisted of the three following questions: “(1) A bat and a ball cost $1.10 in total. The bat cost $1.00 more than the ball. How much does the ball cost? ___ cents”, “(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ___ minutes”, “(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ___ days”. After completing the CRT, the second survey (T2) was conducted in which participants had to try to remember what they last answered, including their current opinion on the issue (see Figure 1). These memory questions were investigated under the condition of cued recall, that is, participants were instructed to explicitly remember what they previously answered on a political statement. To be confident that the experiment genuinely taps into memory processes, and because one might argue that a relatively short retention interval may be suboptimal for studying memory distortions, the experiment included both cued recall directly following the manipulation (T2) as well as a delayed recall one day later in the final survey (T3).

Procedure

The experiment consisted of three sessions in total (see Figure 2); initial rating of the political statements and interaction with the manipulated answers (T1); A second rating session where participants stated their memory of past attitudes as well as their current attitude after completing the CRT (T2); and a third rating session one day later where participants once again recalled their memory of past attitudes together with their current opinion on the issue (T3).

The experiment proceeded as follows: participants were first recruited from the common areas of Lund University and asked if they were willing to participate in a survey study investigating political attitudes. If they accepted, participants were informed of the general outline of the experiment as well as the compensation they would receive for participating. The participants were then brought to a separate room where they were seated in front of a tablet. Before starting the tablet application, the experimenter explained how the scale should be interpreted as well as how to mark an “X” on the scale, and how to change their response if they wanted, but without mentioning the false feedback. The digital questionnaire contained 12 political statements which were randomly presented one at a time, and the participants were instructed to rate the extent to which they disagreed or agreed on the question as well as their level of confidence (see Figure 1). After making sure that all instructions were fully understood, the experimenter left the room and the participants were left to answer all questions in their own pace. This was the experiment’s first measure, and the attitude ratings obtained during this initial part are referred to as T1 ratings that will serve as a baseline against which later memories and attitudes are compared.

Thereafter, the experimenter returned to the room and explained to the participants that the tablet application will randomly select and display four of the twelve questions one by one along with their own ratings (but without the confidence rating). And before going over the questions with the experimenter, the participants were asked if it was acceptable to record their verbal reports. The participants were then instructed to read the question aloud, tell where on the scale their rating was (e.g. at 25% or about 80%), and explain whether this implied that they agreed or disagreed to the statement as well as the underlying arguments for holding the opinion. However, in two of the four displayed statements, the application moved the participants’ response across the midline to the opposite side of the scale and randomly placed between 15% and 35%, or between 65% and 85%, depending on the direction of the false feedback (see Figure 1). This meant that if the participants responded that they initially agreed to a statement, the attitude rating shifted to disagreeing (or vice versa). In addition, the participants were reminded that they could change their response by clicking the change button if they behaved like something were wrong, or for some reason, the attitude did not reflect their true opinion on the issue. As such, a manipulation was automatically registered as detected if the participant changed their answer and marked a new rating on the scale.

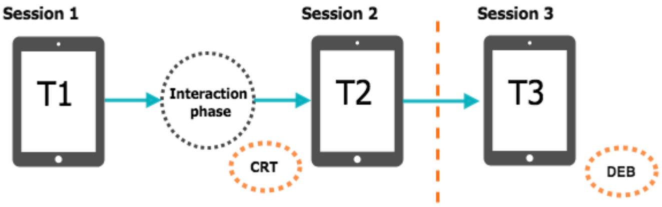


Figure 2. A flowchart illustrating the procedure of the experiment. In the first session (T1), participants make the initial ratings followed by the interaction phase in which they are asked to go over four of the responses together with the experimenter. After completing the CRT, the second rating session (T2) begins where participants state their memory of past attitudes as well as their current opinion on the issue. Approximately a day later, participants are asked again to state their memory of past attitudes together with their current opinion (T3). Finally, participants get a full debriefing (DEB) and receive compensation.

Directly after the interaction with the four target statements, participants had to perform the CRT with the aim to minimize the risk of using any form of mnemonic memory strategy for the upcoming memory questions in the first-follow up session (T2). After completing the CRT, a second questionnaire was administered to the participants in which they were instructed to remember what their previous answers were together with their current opinion on the issue. These ratings are referred to as T2 ratings and are the first memory measure following the manipulation. The questionnaire contained 12 statements; eight from the first questionnaire, four target statements and four non-target statements, as well as four new statements. Furthermore, participants were told that some of the statements they had previously responded on might reappear, because all statements were randomly drawn from the same bank of statements. During this session, the participant also had to tell the experimenter when they found one of the four new statements, and then mark the rating on the scale as close to 0% as possible, simply to see if they knew which of the statements they had not answered before.

Following this, participants were scheduled to come back one day later, on average 22.5 hours (*SD* = 1.9), in order to fill in the ending questionnaire (T3). In this follow-up session, the participants completed another questionnaire identical to the second one (T2) and were asked once again to remember their previous responses in T1 together with their current opinion on the issue. These ratings are referred to as T3 ratings and is the last memory and attitude measure. Finally, the participants were thoroughly debriefed about the real purpose of the experiment, signed informed consent, and received compensation for taking part in the experiment.

Analysis and measures

Ratings for both the attitude and memory questions can be understood on a 0-100 mm scale. All ratings were further realigned on the 0-100 mm scale in order to make them comparable and insensitive of whether the participants agreed or disagreed on a political statement. This means that ratings in T1 which were under the midline of the scale (i.e. below 50 mm) were flipped over to the opposite side (i.e. above 50 mm) and T1 ratings which were over the midline was left unchanged. For example, if the participant answered 30 in T1, 65 in T2, and 40 in T3, these ratings were then recoded to 80 in T1, 45 in T2, and 80 in T3. These realigned ratings were then used to measure attitude change between the original attitude compared to the later attitude ratings in T2 and T3. A negative number represents a weakened/depolarized attitude (towards the false feedback for manipulated trials) and a positive number represents a strengthened/polarized attitude. Memory change (or accuracy) were converted in a similar fashion as for attitude change. Because the main hypothesis is concerned with how the memory of past attitudes is affected by the false feedback, both T2 and T3 memory ratings were analyzed as differences compared with the original T1 attitude rating (i.e. memory change in T2 = memory rating in T2 – attitude rating in T1). A negative number represents a memory of a response that is weaker than the original attitude (towards the false feedback for manipulated trials), while positive numbers represent a memory of a response that is stronger than the original attitude. For example, if the participant's attitude rating in T1 was 60 and the memory rating was 20 in T2, this means a memory change value of -40 (i.e. a weakening of the memory).

R (R Core Team, 2018) was used for data analyzing and lme4 (Bates, Maechler & Bolker, 2015) to perform generalized linear mixed effects analysis of the relationship between the memory of past attitudes and attitude change as well as the effect of the false feedback on future attitude and memory responses. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. As fixed effects, the effects in question (e.g. the effect of manipulated trials on memory change) was entered into the model and random effects were always modeled as per participant intercepts and slopes. Unstandardized beta coefficients and their standard errors are reported for the fixed effects which can be understood in terms of the 0 – 100 mm scale. Significance for fixed effects for the generalized linear mixed models (GLMMs) was obtained by using Wald chi-square tests implemented in the *car* package. Model comparison was assessed with likelihood ratio tests of the full model with the effect in question against the model without the effect in question. The fitted model's marginal model R^2 was computed using the *MuMIn* package which describes the proportion of variance explained by the fixed factors.

3 Results

Detection rates in manipulated trials

Out of the 278 manipulated (M) trials, 166 (60.1%) were detected by the participants, resulting in 112 (39.9 %) accepted M-trials. Average by participant detection rate was 1.20 trials (*SD* = 0.75). Twenty-eight participants detected none of the M-trials (20.1%), 56 detected one M-trials (40.3%), and 55 detected both M-trials (39.6%).

Predictors of detecting the manipulation

Multiple linear regression analysis was used to develop a model for possible predictors on the probability of detecting the manipulation. The model included four predictors: CRT score, initial confidence, memory accuracy, and attitude change. All variables were entered as fixed effects into the model and random effects were modeled as per participant intercepts and slopes. Participants' memory accuracy of past responses was a significant predictor of detection, $\chi 2$ (1) = 30.2 *p* = <.001, as well as participants initial confidence in a political statement, $\chi 2$ (1) = 30.8 *p* = <.001, but CRT score and attitude change did not significantly predict detection, with model marginal R^2 = 0.19. Looking at the coefficients, both memory accuracy (*B* = 0.07, *SE* = 0.03) and initial confidence (*B* = 0.23, *SE* = 0.04) positively predicted increasing probabilities of detecting the false feedback.

Memory change

Participants were asked to state their memory of their previous response on a political statement, both for manipulated (M) and nonmanipulated (NM) trials in the two follow-up surveys (T2 and T3). Negative numbers indicate memory of a response that is weaker than the original attitude (towards the false feedback for M-trials), while positive numbers indicate memory of a response that is stronger than the original attitude. At T2 there appears to be large memory shifts for accepted M-trials (*M* = -31.58, *SD* = 22.24) as well as in T3 (see Figure 3 and 4) one day later (*M* = -30.57, *SD* = 23.35). Memory ratings for detected M-trials in T2 (*M* = -1.76, *SD* = 10.5) and T3 (*M* = -0.8, *SD* = 8.5) show high memory accuracy of previous attitudes. This pattern is also found for the NM-trials in T2 (*M* = -3.58, *SD* = 12.15) and T3 (*M* = -3.52, *SD* = 11.43).

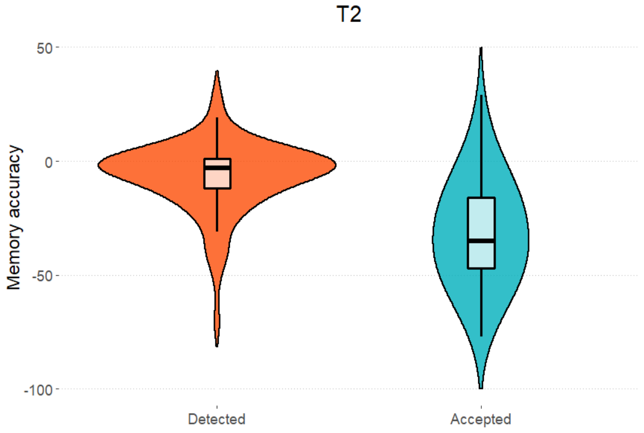


Figure 3. Memory accuracy in T2 compared with original (T1) attitude ratings for manipulated trials. A negative difference indicates that the memory of past responses is weaker than the original attitude (in the direction of false feedback). Boxplots illustrate median (straight lines), 25th and 75th quantile (box edges), as well as the interquartile range (vertical lines).

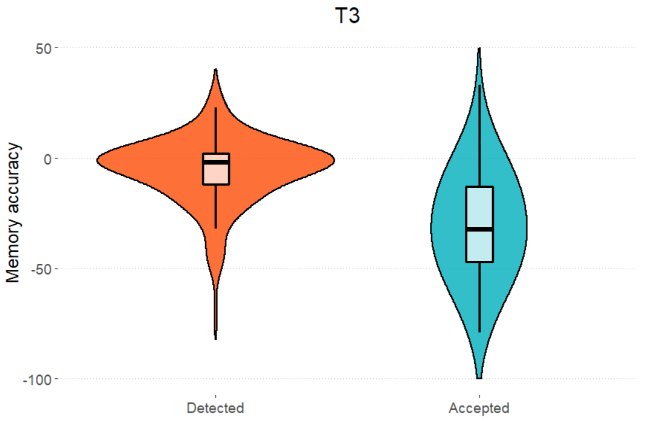


Figure 4. Memory accuracy in T3 compared with original (T1) attitude ratings for manipulated trials. A negative difference indicates that the memory of past responses is weaker than the original attitude (in the direction of false feedback). Boxplots illustrate median (straight lines), 25th and 75th quantile (box edges), as well as the interquartile range (vertical lines).

Effect of manipulation on memory

Generalized linear mixed models were used to analyze the effect of the false feedback on later memory ratings of past attitudes in T2 and T3. The first regression model tested whether memory change differed between manipulated (M) and non-manipulated trials (NM) and was entered as fixed effects into the model with time as an interaction term, while random effects were modeled as per participant intercept and slope. There was a significant main effect of manipulation, χ2 (1) = 49.53, *p* = <.001, but no main effect of time, χ2 (1) = 0.403, *p* = .525, nor between the interaction of manipulation and time, χ2 (1) = 3.07, *p* = .079, with model marginal R^2 = 0.07. Looking at the model coefficients, there appears to be a weakening of the memory for manipulated trials in T2 (*B* = -10.67 mm, *SE* = 1.2) which is slightly reduced one day later in T3 (*B* = 2.76 mm, *SE* = 1.7). In nonmanipulated trials, participants displayed high accuracy when stating their memory of past responses in T2 (*B* = 2.3 mm, *SE* = 1.2) as well as in T3 (*B* = 1.5 mm, *SE* = 1.7).

The second regression model (see Figure 5) tested the effect of accepting the false feedback on later memory shifts, subsetting the data to only accepted and detected manipulated trials which were entered as fixed effects into the model together with time as an interaction term. Random effects were again modeled as per participant intercepts and slopes. There was a significant main effect of accepted false feedback, χ2 (1) = 116.1, *p* = <.001, but no main effect of time, χ2 (1) = 1.02, *p* = .31, nor between the interaction of accepted false feedback and time, χ2 (1) = 0.02, *p* = .88, with model marginal R^2 = 0.32. Looking at the model coefficients, there appear to be large memory shifts for accepted manipulated trials in T2 (*B* = -25.3 mm, *SE* = 2.11) which is further depolarized in T3 (*B* = -0.3 mm, *SE* = 2.6). In detected trials, participants showed fairly accurate memories of past attitudes in T2 (*B* = -6.77 mm, *SE* = 1.5) which were highly accurate at T3 (*B* = 1.3 mm., *SE* = 1.6).

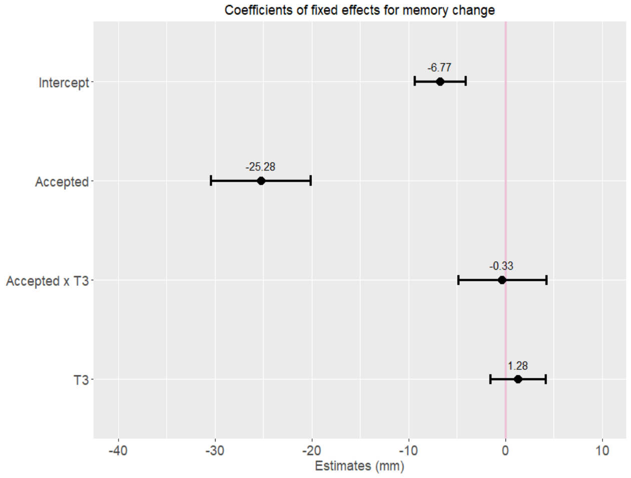


Figure 5. Coefficient plot for the contribution of fixed effects on memory change measured as a difference from to the original (T1) attitude rating. Estimates with a negative number imply a weakening of the original memory (towards the false feedback in accepted manipulated trials). The intercept represents the reference level of memory change for detected trials in T2. Points represent the estimates of the fixed effects while hinges represent the 95% confidence interval.

The third regression model (see Figure 6) tested whether the degree of confidence in a political statement affected the size of the memory change, by including an additional standardized confidence term into the model above with all interactions of accepted false feedback and time. There was a significant main effect of confidence, χ2 (1) = 13.52, *p* = <.001, a significant main effect of accepted false feedback, χ2 (1) = 117, *p* = <.001, as well as a significant interaction between confidence and accepted false feedback, χ2 (1) = 10.04, *p* = .002, with model marginal R^2 = 0.35. Looking at the fixed coefficient of confidence, it shows that higher initial confidence increases the size of the memory change in accepted manipulated trials (*B* = -6.88 mm, *SE* = 2.1).

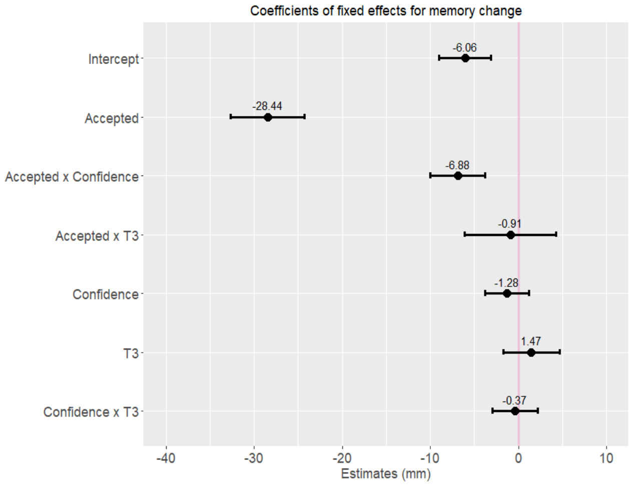


Figure 6. Coefficient plot for the contribution of fixed effects on memory change measured as a difference from to the original (T1) attitude rating. Estimates with a negative number imply a weakening of the original memory (towards the false feedback in accepted manipulated trials). The intercept represents the reference level of memory change for detected trials in T2. Points represent the estimates of the fixed effects while hinges represent the 95% confidence interval.

Attitude change

Following each memory question, both for manipulated trials and non-manipulated trials, participants were also asked to state their current attitude on the political statement. Negative numbers imply a weakened/depolarized attitude and positive numbers imply a strengthened/polarized attitude. Participants displayed large attitude change (see Figure 7 and 8) for accepted M-trials in T2 (M = -24.43, SD = 22.61) and there is small regress in the attitude at T3 (M = -23.2, SD = 23.07). For nonmanipulated trials it appears to exhibit attitude change at T2 (M = -7.98, SD = 20.95) which is attenuated at T3 (M = -4.87, SD = 15.46). In detected trials, participants were highly accurate in restating their original attitude in both T2 (M = -3.16, SD = 18.81) and T3 (M = -0.63, SD = 14.14). A Welch two sample t-test between the difference in ratings for memory change and attitude change for accepted M-trials shows a significant difference in T2 (M_{diff} = -7.07, 95% CI [-10.52, -3.62], t_{109} = -4.06, p <. 001) as well as in T3 (M_{diff} = -7.73, 95% CI [-10.90, -4.55], t_{109} = -4.824, p <. 001).

Effect of manipulation and memory on attitude change

Generalized linear mixed models were used to analyze the effect of false feedback on later attitude ratings in the two follow-up surveys. The first regression model tested whether attitude change differed between manipulated (M) compared to non-manipulated (NM) trials and were entered as fixed effects into the model together with time as an interaction term. There was a significant main effect of manipulation, χ2 (1) = 12.82, p = <.001, but no significant main effect of time, χ2 (1) = 3.08, p = .054, nor any effects of the interaction between time and manipulation, χ2 (1) = 1.30, p = .26, with model marginal R^2 = 0.02. When interpreting the model coefficients, it shows a small weakening in the attitudes for manipulated trials at T2 (B = -5.6 mm, SE = 1.3) which was slightly reduced at T3 (B = 2.1 mm, SE = 1.8). In nonmanipulated trials, participants showed high attitude accuracy both at T2 (B = 0.99 mm, SE = 1.3) and T3 (B = 2.8 mm, SE = 1.8).

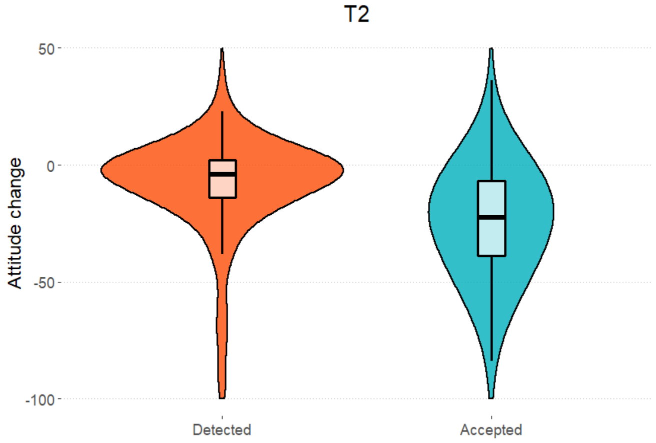


Figure 7. Attitude change in T2 compared with original (T1) attitude ratings for manipulated trials. A negative difference implies a weakening of the original attitude (in the direction of false feedback). Boxplots illustrate median (straight lines), 25th and 75th quantile (box edges), as well as the interquartile range (vertical lines).

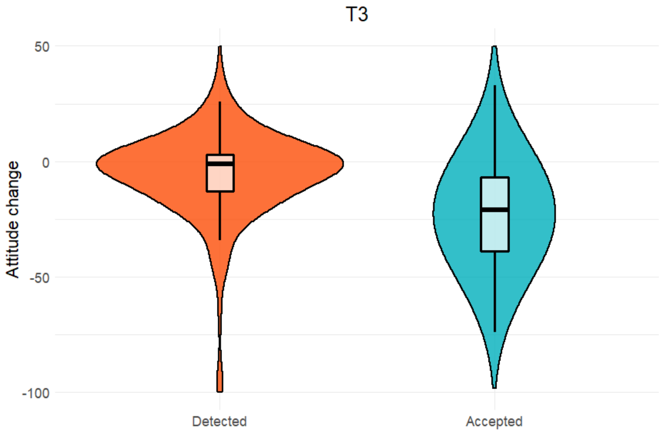


Figure 8. Attitude change in T3 compared with original (T1) attitude ratings for manipulated trials. A negative difference implies a weakening of the original attitude (in the direction of false feedback). Boxplots illustrate median (straight lines), 25th and 75th quantile (box edges), as well as the interquartile range (vertical lines).

The second regression model (see Figure 9) tested whether attitude change was larger when participants accepted the false feedback in manipulated trials, subsetting the data to only include accepted and detected manipulated trials which were entered as fixed effects into the model. As before, the model included time as an interaction term while random effects were modeled as per participant intercepts and slopes. There was a significant main effect of accepted false feedback, χ2 (1) = 58.14, p = <.001, but no main effect of time, χ2 (1) = 2.62, p = .10, nor for the interaction between accepted false feedback and time, χ2 (1) = 0.3214, p = .57, with model marginal R^2 = 0.19. Looking at the model coefficients, there appears to be a large weakening of the attitudes for accepted manipulated trials at T2 (B = -19.7 mm, SE = 2.3) which were further depolarized at T3 (B = -1.6 mm, SE = 2.8). In detected

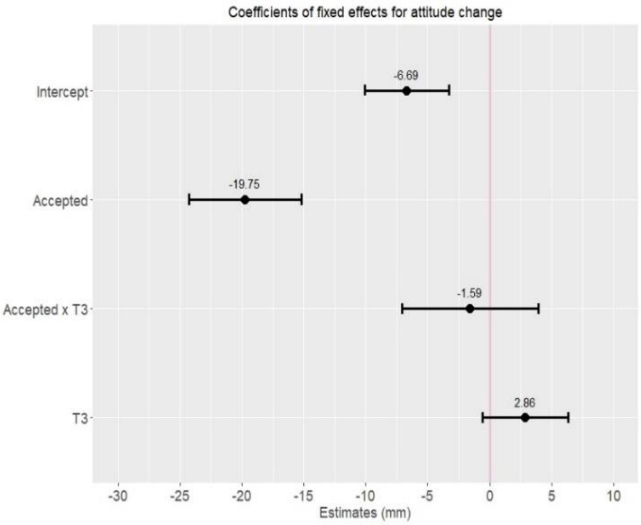


Figure 9. Coefficient plot for the contribution of fixed effects on attitude change measured as a difference from to the original (T1) attitude rating. Estimates with a negative number imply a weakening of the original attitude (towards the false feedback in accepted manipulated trials). The intercept represents the reference level of attitude change for detected trials in T2. Points represent the estimates of the fixed effects while hinges represent the 95% confidence interval.

trials, participants restated their original attitude approximately accurate at T2 (*B* = -6.7 mm, *SE* = 1.7) which nearly matches the original attitude rating in T3 (*B* = 2.8 mm, *SE* = 1.7).

The third regression model tested whether the memory of past attitudes was a good predictor for later attitude shifts, by comparing the goodness of fit with the second regression model. As fixed effects, memory change was additionally included in the model together with time as an interaction term. The data were again subsetted to only manipulated trials because the magnitude of attitude change was highest for accepted manipulated trials. There was a significant main effect of memory, $\chi2$ (1) = 186.18, *p* = <.001, as well as a significant interaction of memory and time, $\chi2$ (1) = 5.63, *p* = .017, but no significant main effect of time, $\chi2$ (1) = 3.42, *p* = .064, with model marginal R^2 = 0.514. As hypothesized, the memory of past attitudes seem to explain a large proportion of variance in later attitude shifts for manipulated trials, and significantly improved the models fit compared to the second fitted regression model, $\chi2$ (2) = 307, *p* = <.001.

Memory accuracy and attitude change in non-target trials

A control condition consisting of non-target statements was used to assess how accurate the memory of past attitudes was when no feedback was given to the participants during the experiment. Memory accuracy was analyzed as differences between the original T1 attitude rating compared to the memory rating in T2 or T3. As before, negative numbers imply a memory of a response that is weaker than the original attitude, while positive numbers imply a memory of a response that is stronger than the original. Memory accuracy was slightly higher for non-target trials in T2 (*M* = -5.91, *SD* = 14.99) compared to the second follow-up survey in T3 (*M* = -7.35, *SD* = 16.35). Participants were also asked about their attitude for non-target statements in order to look if attitude ratings fluctuate when no feedback is received between T1 and T3 (i.e. a form of baseline measure). There appears to be a moderate attitude change for non-target trials at T2 (*M* = -8.98, *SD* = 20.29) which is also evident one day later in T3 (*M* = -8.75, *SD* = 19.1).

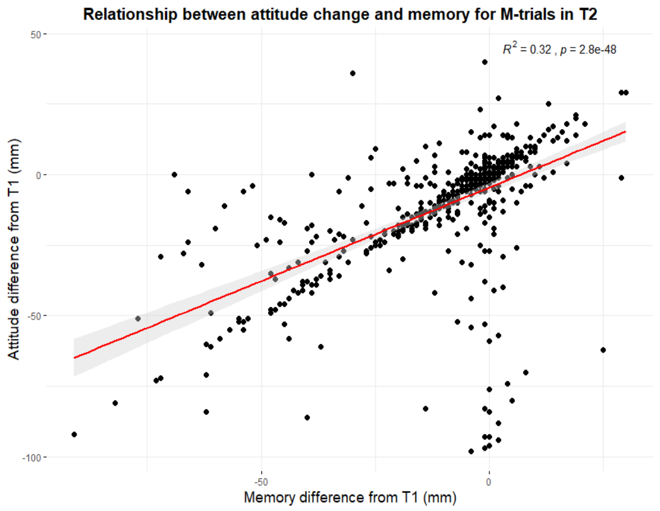


Figure 10. Scatterplot with line of best fit (95% confidence interval) showing a moderately strong, positive relationship between memory of previous responses and attitude change for accepted and detected M-trials in T2 with some potential outliers. Differences for both memory and attitude rating is compared to the original attitude rating (T1) in millimeters. R-squared (R^2) and significance level are also shown.

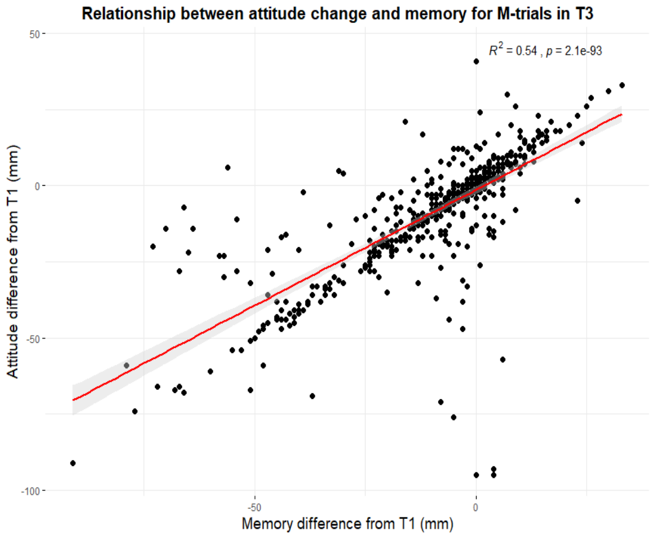


Figure 11. Scatterplot with line of best fit (95% confidence interval) showing a strong, positive relationship between memory of previous responses and attitude change for accepted and detected M-trials in T3 with a few potential outliers. Differences for both memory and attitude rating is compared to the original attitude rating (T1) in millimeters. R-squared (R^2) and significance level are also shown.

Relationship between attitude change and memory

The first correlational model tested whether memory of past responses was associated with attitude change for accepted and detected manipulated trials in T2 (see Figure 10). There was a significant positive correlation between attitude change and memory in T2, *r*(548) = 0.56, *p* <.001, 95% CI [0.50, 0.62]. The second correlational model tested whether this positive relationship also is evident in the second follow-up one day later (T3) for manipulated trials (see Figure 11). There was a significant, strong, positive correlation between attitude change and memory in T3, *r*(543) = 0.73, *p* <.001, 95% CI [0.69, 0.77]. The memory of past attitudes was also able to account for 32% of the variance in attitude change for manipulated trials in T2 as well as 54% of the variance in attitude change for manipulated trials at T3.

4 Discussion

The purpose of the experiment was to investigate whether false feedback on specific responses concerning political statements on a survey would influence later memories of those past responses as well as which downstream effects this have on later attitudes towards these issues. The main hypothesis predicted that the choice blindness manipulation creates a false memory of a past response that, in turn, shapes future responses. This was supported by the results because participants' memory of past responses was strongly influenced by the false feedback if they accepted it as being their own, both directly following the manipulation and one day later. The second hypothesis predicted that attitude change would occur in accepted manipulated trials and move towards the false feedback. This was also supported by the results as participants' attitude responses shifted towards the false feedback, both in the first session following the manipulation as well as one day later. The third hypothesis predicted that attitude change would be positively correlated with memory change. Looking at the findings, it showed that participants' memory of past attitudes in both follow-up sessions was positively correlated with attitude change and explained a large portion of variance for later attitude shifts. In summary, these findings suggests that the false feedback, when accepted, creates a false memory of a past response, which is used by the participants when generating their future response on a political statement, thus showing that explicit memories of past attitudes are strongly affected by the choice blindness manipulation and may play an important role in driving attitude change. Understanding whether memory mechanisms can explain choice blindness is valuable because emerging evidence suggests that decisions oftentimes is guided by the memory of past experience. This will also deepen the knowledge of the cognitive mechanisms underlying choice blindness.

Predictors of detecting the manipulation

One of the most crucial components in a choice blindness experiment is to what degree participants detects the false feedback. This experiment found that approximately half of the manipulated trials were accepted (39.9%) as being their own response, which is slightly higher detection rate than in previous studies regarding political attitudes (Strandberg et al., 2018; Hall et al., 2012). To better understand why some participants are more likely to detect the manipulation, the experiment tested several possible predictors that could explain the differences between participants. The results showed that memory accuracy of past responses, as well as initial confidence on a statement, positively predicted increasing probabilities of detecting the false feedback. Interestingly, if participants had higher memory accuracy regarding their past attitudes, it increases the likelihood to detect the manipulation. This makes sense because if one accurately remembers having taken a position on a specific political issue, then one should also be more resistant to accept the false feedback as reflecting the original memory trace. Previous studies using the CBP have not found this predictor of detection before (Sagana et al., 2014a; Sagana et al., 2014b; Strandberg et al., 2018), which is noteworthy because it suggests that acts of recollection influences participants' ability to discern the false feedback.

In line with other work showing that low confidence facilities the acceptance of the false feedback (Sagana et al., 2014b), this experiment also found that higher initial confidence regarding the attitude increases the likelihood to detect the manipulation. In other words, participants were more willing to accept a manipulated outcome if they had a lower degree of confidence regarding their attitude. The simplest explanation for this is that participants with low confidence ratings do not have a strong opinion toward the issue and are therefore more prone to accept false information about their past decisions. However, it is important to note that participants, in general, displayed stable attitudes (60.1% of the participants corrected the manipulation), thus showing that both stable and flexible attitudes can exist side-by-side.

Furthermore, it has recently been found that higher scores on CRT increase the likelihood to detect false feedback during a choice blindness experiment (Strandberg et al., 2018). CRT is designed to measure peoples' tendency to inhibit an incorrect "gut" response and engage in further reflection to find a correct answer, thus capturing the ability for critical reasoning (Frederick, 2005). Considering this, it has been hypothesized that it might affect the ability to detect false feedback. But the results in this experiment did not find a significant relationship between higher CRT scoring and the ability to detect the manipulation. In total, the most interesting finding mentioned above is that participants' memory of prior answers seems to affect the ability to detect the false feedback and more research is needed to find out which memory mechanisms that might be in play to give rise to this effect.

Influence of false feedback on future memories and attitudes

Considering previous studies regarding choice blindness and its effect on memory, this experiment aimed to investigate an underexplored aspect of the CBP: how the memory of past attitudes is affected when accepting the false feedback. In the first follow-up survey (T2), when participants are asked to try to remember what they answered last, the memory shifts are largest, and these shifts were still persistent when asked again one day later (T3). In other words, if the participant accepts the manipulated rating as being their own, memory changes in the direction of the false feedback, both directly following the manipulation and one day later. This is consistent with Pärnamets et al. (2015) who found that later choices as well as source memory is selectively affected when accepting the false feedback. However, this experiment also demonstrates that explicit memories of past choices are influenced, which has not been found before (Sagana et al., 2014a; Sagana et al., 2014b). Although previous experiments have shown that if participants fail to detect the manipulation, their subsequent decisions and memories can become biased by the false feedback (Cochran et al., 2016), they have only used indirect memory measures. This is also the case for false memory research, which does not typically assess false memories using cued recall which the present thesis did (Bernstain et al., 2005; Clifasefi et al., 2013). Consequently, this experiment demonstrates that not only can false memories about past choices implicitly influence our future attitudes, but also explicit memories of those choices.

Furthermore, in nonmanipulated trials, participants displayed high accuracy when stating their memory of past responses, both in the first session following the manipulation as well as one day later. Generally, this was also the case for those trials where the participants detected the false feedback. This is important because the main hypotheses predicted that if the memory changes when accepting the false feedback, then participants should also exhibit high memory accuracy in detected and nonmanipulated trials. Additionally, there seemed to be a greater spread and less memory accuracy for those trials that received no feedback at all. However, these changes were quite small and could be due to natural variation in the ability of recollection, or because the lack of feedback makes it harder for the participants to remember their attitude on the specific issue.

Looking at how participants' attitudes is affected by the manipulation, the results showed that attitudes moved towards the false feedback in accepted manipulated trials and these changes were still evident one day later as in previous studies (Strandberg et al., 2018). In those trials where the participants detected the manipulation, the attitude moved back to the original attitude. However, if the participants did not receive any feedback during the experiment, their attitude rating seemed to fluctuate a bit. This could be due to the fact that participants are asked to restate their current attitude directly after being asked to recall their original rating (in both T2 and T3) and thus may simply reevaluate the political statement within the same interval that the attitude was earlier. Nevertheless, the results above replicate earlier studies by demonstrating that choice blindness itself can serve as an influential environmental feedback mechanism that causes attitude change over time (Cochran et al., 2016; Johansson et al., 2013; Strandberg et al., 2018).

False memories influence future attitudes

To investigate whether memory mechanisms could account for the choice blindness phenomena, rather than self-perception processes as in previous studies (Johansson et al., 2005; Strandberg et al., 2018), analysis was performed to look at the relationship between memory and attitude change. First of all, the results showed that the memory of past attitudes was a significant predictor for later attitude shifts in manipulated trials with a large effect size. Furthermore, there was a strong positive correlation between attitude change and memory in the first session following the manipulation (T2) in which memory of past attitudes were able to account for 32% of the variance in attitude change. And this effect was even larger in the second session (T3) where the memory of past attitudes accounted for 54% of the variance in attitude change for manipulated trials. Additionally, when comparing the mean differences between memory change and attitude change in accepted manipulated trials, it showed that memory change was significantly more extreme in contrast to attitude change at both T2 and T3. Participants thus seem to be more affected by the false feedback when trying to remember their original attitude compared to when restating their current attitude on the political issue. These findings thus suggest that the memory of past decisions in some way influences which attitudes one holds in the future (Ariely & Norton, 2008; Wimmer & Sohamy, 2012), and suits well with the attitude-as-memory program that attempts to conceptualize attitude construction according to well-known effects of memory processes (Reyna, 2012; Weber & Johnson, 2006; Wimmer & Sohamy, 2012).

One possible explanation for the relationship between memory and attitude change noted above is therefore that the false feedback, when accepted, creates a false memory of a past response, which is used by the participant when generating their future response on a political statement. Accordingly, when participants are asked to state their memory, the false (or competing) memory will temporarily suppress the original memory representation, thereby biasing the participants' perception about the choice. This means that, arguments (or evidence) supporting the manipulated option will temporarily inhibit arguments for other choice options, thus forcing the participants to give arguments against their initially favored option. However, this memory-based interpretation, borrowed from the predictions in query theory, needs further scrutiny as well as to be replicated in future studies. Nevertheless, the current thesis is consistent with the study by Pärnamets et al. (2015), which showed similar false memories for manipulated choices about faces, thus providing another piece of evidence for the memory mechanism explanation and opens up a new way of understanding the long-term effects of choice blindness.

Implications for the nature of memory and attitude

An important theoretical note is that even though participants demonstrates "false" memories after accepting the manipulation, it should not give way to the sceptical narrative of Loftus false memory research which argues that "In essence, all memory is false to some degree" (Bernstein & Loftus p. 373, 2009). This is a somewhat problematic view because there is no criterion that effectively differentiates populations of so-called "true" and "false" memories in real-world situations, thus making a clear dichotomizing of them difficult to sustain (Brown & Reavey, 2017). However, this does not mean that it rules out the possibility for establishing whether a given recollection is sufficiently accurate or not (e.g. source memory errors). Rather, the point is to overcome the unhelpful connotations that "truth" and "falsity" give rise to and instead try to understand the complex situated nature of memory. From this perspective, the act of recollection is often complemented with situational (social, emotional, and environmental) aspects that should be seen as integral rather than external to the cognitive process of remembering. For example, Nelson and Fivush (2004) review research showing how the ways in which parents structure conversations with their children about past events strongly influence how children construct their own narrative history, suggesting that memory is both culturally mediated and contingent to situations. As such, one needs to be humbler when categorizing memory traces as either true or false in general, because we all have a fraught relationship to the past we strive to remember.

The following words of Bartlett appear to argue for this more ecological and relational characterization of memory: "An organism which possesses so many avenues of sensory response as man's, and which lives in intimate social relationship with numberless other organism of the same kind, must find some way in which it can break up this chronological order and rove more or less at will in any order over the events which have built up its present momentary 'schemata'. It must find a way of being dominantly determined, not by the immediately preceding reaction, or experience, but by some reaction or experience more remote... We must, then, consider what does actually happen more often than not when we say what we remember. The first notion to get rid of is that memory is primarily or literally reduplicative, or reproductive. In a world of constantly changing environment, literal recall is extraordinarily unimportant" (Bartlett, 1932, pp. 203-204).

What can be said about the process of remembering is therefore that its primary function is most likely not to provide exact details of past experiences, but rather to cover the gist of the original event and to update your memory belief when a new source of information is given about an event. This means that cognitive scientists or psychologists that use the empirical literature on misinformation effects as an argument for verifying how unreliable our memory system is misleading (e.g. Bernstein & Loftus, 2009). One reason is that in misinformation experiments, participants are often provided with misleading information about an event, and then tested to see whether their subsequent judgments or choices reflect the misinformation. The advantage of this experimental design is that it enables one to identify those memory beliefs that are affected by the information presented after the event: those memory beliefs will reflect the false information. However, the observation that people's memories are easily manipulated does not provide good reason for thinking that in real-word situations people will use inaccurate information to update their memory beliefs. Instead, it might be preferable to view such false memories (or memory errors) as a function of adaptive constructive processes which in effect improve memory performance in the long run (Carpenter & Schacter, 2017). On the one hand, these flexible reconstructive processes can lead to cognitive benefits because information about past experiences will be integrated with new information derived from memory, allowing one to add missed or misrepresented details that could improve future judgments or decision-making. On the other hand, the same flexible recombination processes that underpin these adaptive processes can also lead to memory errors in which the representation of the past tends to be biased, glorified and reflecting of one's own interests (Ross, 1989).

One way of understanding why participants' memory changes in T2, during accepted manipulated trials, is therefore that they update their memory of past attitudes. This occurs because when accepting the false feedback new information about past actions is added to the original memory trace, thus reflecting the most up-to-date information available to the participant. The ability to flexibly combine information about past decisions with newly gathered information allows participants to change how they recall their previously responses and therefore also which political attitude one holds. What this means is that participants' behaviour might not be as irrational as one initially imagines, especially if they do not have a strong opinion towards the political issue to begin with. Rather, the findings in this experiment highlight the flexible and dynamic way attitudes develop over time and its sensitivity to situational factors. It is therefore not surprising that memories of past actions can influence and bias future attitudes (Ariely & Norton, 2008) and possibly drive the change in participants' attitudes toward a specific political issue.

Limitations and future studies

One limitation in the experiments design is that it is not sufficiently strong to answer the causal questions regarding how memory involvement affects later attitude change. The current experiment show that memories of past choices and attitude change covary and the results suggest that false memories (or memory errors) resulting from the choice blindness manipulation changed attitudes. It might seem more likely that memory change drives the attitude change when accepting the false feedback as being one's own response (i.e. memories change attitudes). But it is possible that the memory report adjusts to fit the attitude report (i.e. attitudes change memories) and that attitude change causes source confusions in memory. In other words, participants changed their attitudes during the interaction phase when accepting the false feedback, and these false attitudes led them to recall having taken a positive position to a political statement, when they, in fact, had chosen a negative position from the beginning. According to this mechanism, attitude change causes memory errors or a memory bias in recalling past choices (i.e. recollection of past memory beliefs will become biased towards the current memory belief). The purpose of doing the second experiment is to test these competing explanations for the results in the first experiment – do memories change attitudes or do attitudes change memories?

It would also be interesting to look at how episodic memory processes influence the change of attitudes and how competing memory representations are weighed against each other to construct new attitudes within that range. For example, one could possibly look at ways of introducing multiple competing memories or try to induce different degrees of competition (e.g. different distances to the supposed original memory). This would also lead to a better understanding of memory involvement in decision-making which is important because emerging evidence suggests that memory plays an essential role in at least some kinds of value-based decisions, particularly those that rely on the integration of information across distinct past events (Wimmer & Sohamy, 2012). Future work should also try to replicate the findings in this experiment to increase its generalizability as well as to explore other domains than politics, such as esthetic values or economical judgments.

Conclusions

In summary, this thesis provides complementary evidence supporting the view that the choice blindness phenomena might result from memorial processes. Specifically, the results demonstrate that if one accepts the false feedback about past choices it induces substantial changes in explicit memories of those past choices. Consequently, a feasible explanation that may account for the effects of choice blindness is that during a manipulation a false memory is created of what was previously answered, which is used by the participants when generating their future response. Additionally, these memory changes seemed to explain later attitude shifts towards a specific political issue. This suggests that someone can redefine who they are and, in fact, which political attitudes one holds by being able to change the perception of how the past is configured. Better understanding how memory-based decisions influence attitudes is an important step towards understanding the cognitive mechanism underlying attitude change as well as how situational forces and external feedback shape these processes.

5 References

Ariely, D., and Norton, M. I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences, 12*(1), 13–16.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. New York, NY: Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200.

Bernstein, D. M., Laney, C., Morris, E. K., and 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*(39), 13724–13731.

Bernstein, D.M. and Loftus, E.F. (2009) How to tell if a particular memory is true or false. Perspect. Psychol. Sci. 4, 370–374

Brainerd, C. J., and Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10,* 3–47.

Brown, S. D., & Reavey, P. (2017). False memories and real epistemic problems. *Culture & Psychology*, *23* (2), 171 – 185.

Carpenter, A. C., and Schacter, D. L. (2017). Flexible retrieval: When true inferences produce false memories. *Journal of experimental psychology. Learning, memory, and cognition*, *43*(3), 335–349.

Clifasefi, S. L., Bernstein, D. M., Mantonakis, A., and Loftus, E. F. (2013). "Queasy does it": false alcohol beliefs and memories may lead to diminished alcohol preferences. *Acta Psychol*, *143*, 14–19.

Cochran, K. J., Greenspan, R. L., Bogart, D. F., and Loftus, E. F. (2016). Memory blindness: altered memory reports lead to distortion in eyewitness memory. Mem. Cogn. 44, 717–726. doi: 10.3758/s13421-016-0594-y

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552-564.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19,* 25–42.

Goethals, G. R., and Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology, 9,* 491-501.

Harris, R. J. and G. E. Monaco. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology 107, No 1*: 1-22.

Hall, L., Johansson, P., Tärning, B., Sikström, S., and Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition, 117,* 54 – 61.

Hall, L., Johansson, P., and Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE, 7,* e45457.

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., and Johansson, P. (2013). How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions. *PLoS One*, *8*(4), e60554.

Hardt, O., Einarsson, E. O., and Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61*, 141–167.

Henkel, L. A., & Mather, M. (2007). Memory attributions for choices: How beliefs shape our memories. *Journal of Memory and Language*, *57*(2), 163-176.

Howe, M. L., Garner, S. R., Charlesworth, M., and Knott, L. (2011). A brighter side to memory illusions: False memories prime children's and adults' insight-based problem solving. *Journal of Experimental Child Psychology, 108,* 383–393.

Johnson-Abercrombie, M. L. (1960). *The anatomy of judgment*. New York: Basic Books.

Johansson, P., Hall, L., Sikström, S., and Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310,* 116–119.

Johansson, P., Hall, L., Sikström, S., Tärning, B., and Lind, A. (2006). How something can be said about telling more than we can know: On choice blindness and introspection. *Consciousness and Cognition*, *15,* 673–692.

Johansson, P., Hall, L., Tärning, B., Sikström, S., and Chater, N. (2014). Choice blindness and preference change. *Journal of Behavioral Decision Making, 27,* 281–289.

Johnson, A., and Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience, 27*(45), 12176–12189.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., and Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science, 25,* 1198–1205.

Lind, M., Visentini, M., Mäntylä, T., & Del Missier, F. (2017). Choice-Supportive Misremembering: A New Taxonomy and Review. *Frontiers in psychology*, *8*, 2062.

Lord, C. G., and Lepper, M. R. (1999). Attitude representation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 265-343). San Diego, CA: Academic Press.

Mather, M., Shafir, E., and Johnson, M. K. (2000). Misrememberance of options past: Source monitoring and choice. *Psychological Science, 11,* 132–138.

McLaughlin, O., & Somerville, J. (2013). Choice blindness in financial decision making. *Judgment and Decision Making, 8*(5), 561–572.

Milner, B., Squire, L.R., and Kandel, E.R. (1998). Cognitive neuroscience and the study of memory. *Neuron, 20,* 445–468.

Nelson K, Fivush R. The emergence of autobiographical memory: a social cultural developmental theory. *Psychol Rev 2004, 111*:486–511.

Orwell G. *Nineteen Eighty-four*. New York: Harcourt, Brace; 1949.

Preston, A. R., Shrager, Y. , Dudukovic, N. M. and Gabrieli, J. D. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*: 148-152.

Pärnamets, P., Hall, L., and Johansson, P. (2015). Memory distortions resulting from a choice blindness task. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1823–1828). Austin, TX: Cognitive Science Society.

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Reyna V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and decision making*, *7*(3), 332–359.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*, 341-357.

Sagana, A., Sauerland, M., and Merckelbach, H. (2014a). Memory impairment is not sufficient for choice blindness to occur. *Front. Psychol*. *5*:449.

Sagana, A., Sauerland, M., and Merckelbach, H. (2014b). 'This is the person you selected': eyewitnesses' blindness for their own facial recognition decisions. *Appl. Cogn. Psychol*. *28*, 753–764.

Schacter, D. L., Guerin, S. A., & St Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*, 467–474

Strandberg, T., Sivén, D., Hall, L., Johansson, P., and Pärnamets, P. (2018). False beliefs and confabulation can lead to lasting changes in political attitudes.*Journal of Experimental Psychology: General*, *147*(9), 1382-1399.

Sutton J. Remembering. In: Robbins P, Aydede M, eds. *The Cambridge Handbook of Situated Cognition*. New York: Cambridge University Press; 2009, 217–235

Thaler, Richard H. (2016). Behavioral Economics: Past, Present, and Future. *American Economic Review*, *106* (7): 1577-1600

Tulving, E. (1972). Episodic and semantic memory. In: Tulving E & Donaldson W (eds) *Organization of memory*. Academic Press, New York, pp 381-403.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press: Oxford.

Weber, E. U., and Johnson, E. J. (2006). Constructing preferences from memory. In: Lichtenstein, S. & Slovic, P., (Eds.), *The Construction of Preference* (pp. 397-410). New York NY: Cambridge University Press.

Weber, E. U., Johnson, E. J., Milch, K., Chang, H., Brodscholl, J., and Goldstein, D. (2007). Asymmetric discounting in intertemporal choice: A query theory account. *Psychological Science*, *18*, 516-523.

Wimmer G, E., and Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*, 270-273.