Disowned recollections: Denying true experiences undermines belief in occurrence but not judgments of remembering

Giuliana Mazzoni[a*], Andrew Clark[a], Robert A. Nash[b]

[a]Department of Psychology, University of Hull, UK

[b]School of Psychology, University of Surrey, UK

Paper published in *Acta Psychologica* (2014). http://dx.doi.org/10.1016/j.actpsy.2013.11.007

Word count: 6482 words (main body)

*Corresponding author:

Prof. Giuliana Mazzoni
Department of Psychology
University of Hull
Cottingham Road
Hull, HU6 7RX
United Kingdom
Email: g.mazzoni@hull.ac.uk

Abstract

Recent research findings have illustrated that false memories induced in the laboratory can be dissociated from the beliefs that the events had in fact occurred. In this study we assessed whether this dissociability is a quality peculiar to false memory, or whether it represents a general characteristic of autobiographical memory. To this end we examined whether people can be induced to stop believing in memories for true experiences. Participants observed and performed simple actions, and were later falsely informed that they had not performed some of them—that false memories for these actions had been implanted through the use of fabricated evidence. Before and after receiving this misinformation, participants rated their belief in and memory of performing those actions, other actions that they had also performed, and actions that they had not performed. Whereas the misinformation substantially undermined participants' beliefs in the specific performed actions about which they had been misinformed, it had little effect on their endorsement of remembering those actions. The misinformation thus boosted the proportion of occasions in which participants rated their memories as stronger than their beliefs, and it weakened the correlation between belief and memory ratings. Thus, this study provides the first experimental demonstration of non-believed memories of true experiences. We discuss our findings with reference to the small literature concerning the use of socially-communicated misinformation to undermine event memories, and with reference to the structure of autobiographical memory.

Key words: Non-believed memories; misinformation; autobiographical belief; autobiographical memory

Disowned recollections: Denying true experiences undermines belief in occurrence but not judgments of remembering

1. Introduction

When we state that we remember an event, we generally mean two things. One is that we can recall it happening; the other is that we believe it did in fact occur. The first is an autobiographical memory; the second is an autobiographical belief (Mazzoni & Kirsch, 2002). Although these two constructs are often conflated in memory research, recent studies have shown that they can vary independently (Clark, Nash, Fincham, & Mazzoni, 2012; Mazzoni, Scoboria, & Harvey, 2010; Otgaar, Scoboria, & Smeets, 2013; Scoboria, Jackson, Talarico, Hanczakowski, Wysman, & Mazzoni, in press; Scoboria, Mazzoni, Kirsch, & Relyea, 2004).

As others have already demonstrated, what people report in episodic memory tasks is not only the result of the access and retrieval of memory traces. Rather, these reports are the result also of subjective metacognitive judgments, crucial for deciding whether a mental event is indeed a memory for experiences from one's past (Jacoby, Kelley, & Dywan, 1989; Johnson, Hashtroudi, & Lindsay, 1993; Koriat & Goldsmith, 1996; Mazzoni & Kirsch, 2002; Nelson & Narens, 1990). For example, judgments are at the core of the Source Monitoring Framework—which has shown that the subjective characteristics of mental experiences influence metacognitive decision-making (Johnson et al., 1993; Johnson, Foley, Suengas, & Raye, 1988; Johnson, Kahan, & Raye, 1984)—and also of the remember/know distinction used in much memory research (e.g., Lövdén, Rönnlund, & Nilsson, 2002; Tulving, 1985). The dissociation between what people judge as a memory of an event (hereafter called 'memory' for brevity) versus the belief that the event had occurred (hereafter called 'belief'

for brevity) has most often been studied in the arena of false memory research and, until recently, has been conceived as a single dissociation— that is, experimental manipulations would increase either belief judgments alone, or both belief and memory judgments, but not memory judgments alone. So, in various false memory procedures individuals end up falsely believing that suggested or imagined events had happened (e.g., Garry, Manning, Loftus, & Sherman, 1996), and only in more rare cases these believed events are also claimed to be remembered (e.g., Mazzoni & Memon, 2003; Wade, Garry, Nash, & Harper, 2010). This single dissociation between believing and remembering also holds in everyday autobiographical memory. It is widely accepted that there are many autobiographical events that we believe happened to us, despite having no memory of them (e.g., being born in a particular city on a specific date).

In contrast to the wide acceptance that autobiographical beliefs can exist without any memory, vivid memories for events that the person "knows" did not happen (memories without beliefs, also called non-believed memories) have until recently been considered only as the object of rare—albeit intriguing—anecdotal evidence, such as the widely cited example of Jean Piaget (1951). Piaget's detailed memory of an attempted kidnap at the age of 2 turned out 13 years later to be false. Although on the basis of the evidence Piaget stopped believing in the attempted kidnap, he was unable to stop vividly remembering the event. Non-believed memories are memories that were once believed but no longer are, and recent research has found that they are more common than previously assumed. Mazzoni et al. (2010) examined more than 1500 adults and found that more than one-fifth reported having a vivid memory for an event that they did not believe happened to them. These non-believed memories were held with a level of vividness that made them indistinguishable from believed memories and, even more remarkably, were characterized by the same recollective qualities

as in true memories (D'Argembeau & Van der Linden, 2006; Johnson et al., 1988; Rubin, Schrauf, & Greenberg, 2003). Participants claimed to be able to mentally travel back in time and re-experience these events as if they were true, even though they "knew" these events had not happened. Mazzoni et al. (2010) found that participants' main reason for dropping their belief was social information, that is, being told that the memory was false and the event did not happen (see also Sheen, Kemp, & Rubin, 2001 for a similar example on personal memories in twins).

It is important to note that non-believed memories are different from most typically-studied types of false memories. The false memory literature is concerned with memories of events that did not happen, but that the rememberer believes did happen. Non-believed memories concern events that may or may not have factually happened, but that are not believed to have happened despite vivid recollections of their occurrence. Mazzoni et al. (2010) used the term 'refuted memories' to designate non-believed memories for events that had not occurred. At the same time they hypothesised the existence also of non-believed memories for events that had in fact occurred, which they called *disowned memories*. This study examines the possibility of creating in the laboratory *'disowned memories'* for recent events that are vividly remembered.

Whereas Mazzoni et al. (2010) studied naturally-occurring non-believed memories, a recent false-memory study provided the first empirical demonstration of the creation of non-believed memories for recent events in the laboratory (Clark et al., 2012). After developing very vivid false memories for performing particular actions, participants in that study were debriefed and told that their false memories were indeed false; that is, that the participants had never performed those actions. This debriefing (i.e., information provided by a reliable source), substantially decreased participants' belief in having performed the actions, but had

only a minor effect on their memory ratings. What the debriefing therefore effectively produced was frequent reports of non-believed memories that, as in Mazzoni et al. (2010), had the same recollective qualities and the same level of vividness as true memories for performed actions. Subsequent experiments have shown that this effect can also be obtained for false memories of childhood events, with memory-like recollective qualities persisting for at least several weeks after they are undermined (Otgaar et al., 2013), and that judgments of autobiographical belief and recollection are distinct and based in distinct sources of information and processes (Scoboria et al, in press).

Although the Clark et al. (2012) and Otgaar et al. (2013) studies provide compelling demonstrations that experimentally created beliefs in *false* events can be easily undermined, they do not show that beliefs can be dissociated from *true* memories. Indeed, findings from many domains of psychology show subtle ways in which true and false memories can differ. Neuroimaging evidence has shown, for example, that although true and false memories involve largely comparable patterns of brain activation, some small differential patterns can be identified. True memories have been more strongly associated with the medial temporal lobe regions (e.g., Cabeza, Rao, Wagner, Mayer, & Schacter, 2001; Dennis, Bowman, & Vandekar, 2012; Okado & Stark, 2003; Slotnick & Schacter, 2004), whereas false memories have been more strongly associated with the pre-frontal cortex and parietal regions (e.g., Cabeza et al., 2001; Garoff-Eaton, Kensinger, & Schacter, 2007; Schacter, Buckner, Koutstaal, Dale, & Rosen, 1997; Slotnick & Schacter, 2004). In addition, two recent studies (Beltrami & Mazzoni, 2011; Marini, Agosta, Mazzoni, Dalla Barba, & Sartori, 2012) demonstrated that true and false memories can be discriminated using an implicit behavioral task (aIAT; Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008). Thus, memory and

belief might be independent and dissociable in false memories but not when memories represent an event that has, in fact, occurred.

A further reason why we should corroborate prior findings on non-believed memories by focusing on true events, is that prior studies involving what we might call *negative misinformation* manipulations have had very mixed success. By 'negative misinformation,' we mean information that falsely disputes the occurrence of genuine events, as contrasted with 'positive misinformation', which falsely proposes the occurrence of fabricated events (e.g., suggesting that a person experienced a fictional hot-air balloon ride). Whereas several studies have shown that leading people to actively deny experiencing genuine events is either extremely difficult (Pezdek & Roe, 1997) or completely ineffective (e.g., Roos af Hjelmsäter, Granhag, Strömwall, & Memon, 2008), other studies have shown that negative misinformation can lower people's likelihood of reporting details of an event (e.g. Gabbert, Memon, & Allan, 2003; Merckelbach, van Roermund, & Candel, 2007; Roos af Hjelmsäter, Granhag, & Strömwall, 2012). Yet a common feature of all of these studies is that they did not separately measure both belief and memory; the present study thus stands to add to the research literature on the effects of negative misinformation, by exploring how these two distinct components underlying memory reports are separately influenced.

In sum, whereas Clark et al. (2012) and Otgaar et al. (2013) demonstrate the importance of distinguishing between remembering and believing when studying autobiographical memory (Mazzoni & Kirsch, 2002), the final demonstration of the independence between the belief in the occurrence of an event and the memory for that same event requires showing that it is possible to substantially decrease individuals' belief in the occurrence of vividly remembered events that genuinely occurred. The present study aimed at producing such a demonstration by providing participants with negative misinformation

about the occurrence of true events; that is, by falsely claiming that those events never occurred.

1.1. Overview of the present study

To address our research questions, we used the same general methodology as Clark et al. (2012), which involves a variant of Nash and colleagues' doctored-video procedure (Nash, Wade, & Brewer, 2009; Nash, Wade, & Lindsay, 2009). Clark et al.'s (2012) procedure was designed to create false memories, for which participants' belief could then be undermined. Participants in that study saw doctored video clips that purported to 'prove' (through the inclusion of fabricated clips in the video-evidence) they had performed actions that they did not truly perform. Each fabricated clip was created by digitally combining two genuine clips: one that showed the researcher performing a critical action after the participant had left the room, and one that showed the participant passively observing a different action. The images from these clips were combined to produce composites that seemed to prove that the participant had in fact observed the fake actions. A few hours later, participants were debriefed and told which actions had not been performed.

The methods of the present study differed from those of Clark et al. (2012) in that all clips in the current study showed actions that the participants had in fact performed. The initial "debriefing" therefore involved falsely telling participants that specific genuine video-clips had actually been doctored. Showing participants non-doctored videos served two functions. First, using this paradigm gave us a concrete and compelling explanation to offer participants as to why they would remember things that (allegedly) did not happen. We predicted that telling participants their memories were implanted through the use of doctored video-clips would be highly effective in undermining belief judgments. Second, Nash, Wade, and Brewer (2009) found that showing participants non-doctored (filler) video-clips that

'proved' they performed specific actions increased their reported beliefs and memories of performing those actions. Thus, the use of non-doctored videos ensured that the beliefs and memories we attempted to undermine would be held with high confidence.

We should note that like in Clark et al. (2012), the 'events' in the present study were simple actions, a methodology that itself follows many prior studies exploring the subjective characteristics of recent memories (e.g., Lampinen, Odegard, & Bullington, 2003) and the comparative effects of performing versus simply observing (e.g., Hornstein & Mulligan, 2004; Lindner, Echterhoff, Davidson, & Brand, 2010; Lövdén et al., 2002). Whereas some have questioned whether action memories might behave differently from other types of memories (Roediger & Zaromb, 2010), this consideration was prevailed over by the ease of producing perceptually-rich memories for actions in the lab, and by the frequency with which the simple-actions methodology has been used successfully in applied memory research.

Based on the prior research findings we have outlined above, we predicted that negative misinformation would have little effect on participants' endorsement of remembering their performed actions (e.g., Pezdek & Roe, 1997; Roos af Hjelmsäter et al., 2008), nor upon the phenomenological characteristics of their mental experiences of those actions such as their clarity of visualization or sense of re-living. However, based on prior data regarding non-believed memories, we also predicted that the same misinformation would significantly reduce participants' willingness to endorse believing that those actions occurred (Clark et al., 2012; Otgaar et al., 2013).

2. Method

2.1 Participants

Twenty-four participants (16 females) completed the study. Their ages ranged from 18-34 (M = 21.04, SD = 4.9). Participants who were psychology students were compensated with course credit, non-psychology students' participation was not compensated. The study had a within-subjects design with critical action-type (performed, performed-undermined, new) and session (Session 2 pre-undermining, Session 3 post-undermining) as the independent variables. The Ethics Committee at the Department of Psychology, University of Hull approved this study.

2.2 Materials and Procedure

We used 42 simple actions from Nash, Wade, and Brewer (2009), as also used in Clark et al. (2012). Six of these were critical actions (*clap your hands, click your fingers, rub the table, salute, cover your face with your hands*, and *flex your arm*), selected because pilot participants in Nash, Wade, and Brewer (2009) rated these specific actions as neither highly memorable nor highly unmemorable. The six critical actions were assigned in pairs to be the *performed*, *performed-undermined* or *new* critical actions. Assignment of the pairs was counterbalanced across participants. Performed critical actions were actions that participants genuinely performed in Session 1. Performed-undermined critical actions were also performed in Session 1, but participants were falsely told in Session 3 that they did not perform them. New critical actions were not performed by participants.

2.2.1 Session 1. Participants volunteered to take part in a study purportedly investigating mimicry of actions. They were told that the research involved being filmed observing and copying simple actions. After participants consented to participate, the researcher and participant sat opposite each other at a table and the researcher began video-recording. The researcher then performed the first action for 12 secs, after which the

participant copied this action for 12 secs. This 'observe-copy' process continued until all 28 actions (24 non-critical, 2 performed critical, 2 performed-undermined critical) had been performed by both the researcher and participant. For all participants, the performed critical actions were performed in positions 10 and 18, and the performed-undermined critical actions were performed in positions 6 and 23 of the sequence (the orders of each action-pair were randomly counterbalanced). Note that the 'performed-undermined' actions would not actually be 'undermined' until Session 3; nonetheless for consistency we use this label throughout when referring to these actions.

2.2.2 Session 2. Participants returned 2 days after Session 1. During Session 2, participants were shown a 3-minute video comprising 12 clips, each of which showed the researcher performing an action whilst the participant observed. The actions seen in the video included the two performed critical actions (positions 3 and 7), two performed-undermined critical actions (positions 5 and 10) and eight non-critical actions. Participants watched this video twice. To ensure participants paid attention to the video, on the first viewing they estimated how many times they performed each action per week, and on the second viewing they wrote down a name for each action. Following this task, participants completed a 5-minute anagram-solving filler task. Next, participants completed two questionnaires which examined their memory (questionnaire 1) and belief (questionnaire 2) for performing various actions in Session 1. The rationale for testing belief and memory separately, rather than together like in other studies (e.g., Scoboria et al., 2004), was to reduce the likelihood that participants would use one rating to inform the other. For instance, when choosing their belief ratings participants might expect that these should logically be higher than their memory ratings, and so respond accordingly. Administering the two questions separately reduced the risk of this kind of relative judgment.

For both questionnaires, 28 actions were rated which comprised the 6 critical actions, plus 10 performed non-critical actions and 12 new non-critical actions. Both questionnaires began with the instruction that participants would see a list of actions, some of which they performed in Session 1 and some of which they didn't perform in Session 1. Participants were instructed that they should indicate how strongly they believed and how strongly they remembered performing each action in Session 1, using an 8-point scale. They were asked to think carefully about each of their responses before moving on.

As reported in Clark et al. (2012), our piloting work suggested that participants best understood the distinction between belief and memory when they were asked about their memories first, therefore all participants completed the memory questionnaire first. They were asked, for each action, "How strongly do you *remember* performing this action in Session 1? (1 = No memory at all; 8 = Clear and detailed memory). Following the memory questionnaire participants completed the belief questionnaire, which contained the same 28 actions in a different order. For each action participants responded to the question: "How strongly do you *believe* you performed this action in Session 1? (1 = Definitely did not do this; 8 = Definitely did do this).

2.2.3 Session 3. Session 3 was scheduled between 4-6 hours after Session 2. In this session we aimed to undermine participants' belief in performing the two performed-undermined critical actions. To this end, participants were told that after Session 1 the researcher had doctored the video-recording and created two fake clips of actions, which they saw in the video in Session 2. They were told which two actions had supposedly been faked (i.e., the performed-undermined critical actions). Following this information, participants were asked to rate their belief and memory for each of the six critical actions, using the same scales used in Session 2. Finally, for each of these six actions, participants completed a

questionnaire in which they rated 25 phenomenological characteristics of their memories on a 7-point scale (various anchors were used depending on the item, e.g. dim – sharp; vague – clear, etc.). These included items assessing recollective qualities (e.g., the ability to relive the event); sensory/perceptual qualities (i.e., vision, sound, etc.); valence and intensity of emotions; etc. (for the complete questionnaire and the different scale anchors used for each item, see online supporting materials in Clark et al., 2012).

3. Results

Participants' mean Belief and Memory ratings are displayed in Figure 1. To begin, we ran two separate two-way ANOVAs, one on Belief ratings and one on Memory ratings, with action type (performed, performed-undermined, new) and session (Session 2 pre-undermining, Session 3 post-undermining) as within-subjects variables.[1] The ANOVA on Memory ratings (see right-half of Figure 1) produced a significant main effect of action type, $F(2, 46) = 116.69$, $p < .001$, $\eta^2_p = .84$, and a significant two-way interaction, $F(2, 46) = 65.75$, $p < .001$, $\eta^2_p = .74$, but no significant main effect of session, $F(2, 46) = 1.53$, $p = .23$, $\eta^2_p = .06$. There was no significant change in memory ratings after undermining, and overall memory ratings were higher for performed actions and performed-undermined actions than for new actions. The crucial result of this ANOVA is the interaction. To interpret the interaction, post-hoc analyses were conducted, first examining memory ratings in Session 2

[1] A three-way ANOVA including Measure (Belief vs. Memory) as a within-subjects measure revealed a significant three-way interaction $F(2, 46) = 5.05$, $p = .01$, $\eta^2_p = .18$. For ease of exposition we have not reported this analysis in full.

(pre-undermining), and then in Session 3 (post-undermining), which are reported in the following section.

For Belief ratings (see left-half of Figure 1), the analysis produced significant main effects of action type, $F(2, 46) = 138.86$, $p < .001$, $\eta^2_p = .86$, and session, $F(2, 46) = 4.65$, $p = .04$, $\eta^2_p = .17$, qualified by a significant two-way interaction, $F(2, 46) = 35.33$, $p < .001$, $\eta^2_p = .61$. Belief ratings were significantly higher in Session 2 (pre-undermining) than Session 3 (post-undermining), and they were higher for performed and performed-undermined actions than for new actions. In this case too the interaction was the most important result, and it will be further explored in the next section.

3.1 Pre-undermining ratings (Session 2)

For Memory, a one-way repeated measures ANOVA revealed significant differences across critical action-types, $F(2, 46) = 130.04$, $p < .001$, $\eta^2_p = .85$. Paired sample *t*-tests revealed no significant difference between performed critical actions and performed-undermined critical actions, $t(23) = 0.29$, $p = .78$, $d_z = 0.06$ – unsurprising given that no undermining had yet occurred. However, new critical actions were rated significantly lower for Memory than both performed critical actions, $t(23) = 13.51$, $p < .001$, $d_z = 2.76$, and performed-undermined critical actions, $t(23) = 13.92$, $p < .001$, $d_z = 3.33$.

A similar pattern of results was found for Belief. A one-way repeated measures ANOVA revealed significant differences between the three critical action-types, $F(2, 46) = 119.96$, $p < .001$, $\eta^2_p = .84$. Paired sample *t*-tests revealed no significant difference between performed and performed-undermined critical actions, $t(23) = 0.53$, $p = .60$, $d_z = 0.11$. However, Belief ratings for new critical actions were rated significantly lower than performed critical items, $t(23) = 11.96$, $p < .001$, $d_z = 2.44$, and performed-undermined

critical actions, $t(23) = 16.30$, $p < .001$, $d_z = 2.84$. Together, these findings confirm that there were no initial differences in either type of rating for performed critical actions and performed-undermined critical actions.

3.1.1 Session 2 non-believed memories. In Clark et al. (2012), participants occasionally reported non-believed memories even before debriefing. To examine whether this was the case in the present study, we counted the number of occasions on which participants rated a critical action at least 2 scale-points higher on the Memory scale than on the Belief scale. Prior research (e.g., Scoboria et al., 2004) has conceptualized non-believed memories as being evidenced when Memory is rated at least 1 scale-point greater than Belief. However, because our participants completed the Belief and Memory scales separately as in Clark et al., we used here their more conservative criterion of 2 scale-points to reduce the effect of error variance on the results. We found that 2.78% of critical actions were rated at least two points higher for Memory than for Belief (4.17% of performed actions; 2.08% of performed-undermined critical actions; 2.08% of new actions). When we used an even more conservative criterion for classifying non-believed memories—a difference of 3 or more scale-points between Memory and Belief—2.08% of actions met this criterion (4.17% performed, 2.08% performed-undermined, 0% new). These results replicate those of Clark et al. (2012) in showing that non-believed memories can occur spontaneously in a small minority of cases.

3.2 Post-undermining ratings (Session 3)

At the beginning of Session 3, participants were invited to guess the aim of the study. No participants guessed our hypothesis, or that the research involved examining the dissociation between belief and memory ratings. Furthermore, participants appeared to have believed the cover story.

In Session 3 we were less interested in the ratings of new critical actions, which did not change significantly between sessions [Belief, $t(23) = 0.22$, $p = .83$; Memory, $t(23) = 0.76$, $p = .45$; see Figure 1]; we therefore focus here only on the ratings of performed and performed-undermined actions. We conducted a 2 (Measure: Belief vs. Memory) x 2 (Action type: performed vs. performed-undermined) ANOVA to investigate the effect of the debriefing upon participants' ratings. This analysis revealed a significant interaction effect, $F(1, 23) = 5.26$, $p = .03$, $\eta^2_p = .19$. Specifically, follow-up *t*-tests showed that Memory ratings of performed and performed-undermined critical actions did not significantly differ in this session, $t(23) = 1.33$, $p = .20$, $d_z = 0.27$, but that Belief ratings of performed-undermined critical actions were now significantly lower than those of performed actions $t(23) = 2.54$, $p = .02$, $d_z = 0.52$. In sum, the results from Session 3 show that the negative misinformation lowered Belief ratings, while it did not affect Memory ratings.

3.2.1 Session 3 non-believed memories. Negative misinformation successfully undermined participants' beliefs about the occurrence of the actions without affecting their judgments of remembering those actions. This finding indicates that additional reports of non-believed memories should be observed in Session 3. To assess this possibility, we examined participants' Session 3 ratings to see how often their Memory ratings exceeded their Belief ratings by at least 2 scale-points. Overall, this occurred for 13.89% of critical actions (10.42% performed, 29.17% performed-undermined, 2.08% new). Compared to the Session 2 data, following undermining there were significantly more reports of non-believed memories for performed-undermined actions, McNemar's exact $p = .001$. The same was not true for performed actions, McNemar's exact $p = .38$, or new actions, McNemar's exact $p = 1.00$. This pattern of results held when we used the more conservative criterion of 3 or more scale-points difference between Belief and Memory: 8.33% of critical actions were classified

as non-believed memories, (2.08% performed, 20.83% performed-undermined, 2.08% new); this represented a significant increase from Session 2 only for performed-undermined actions, McNemar's exact $p = .01$, and not for performed actions, McNemar's exact $p = 1.00$, or new actions, McNemar's exact $p = 1.00$.

It is therefore clear that our undermining procedure substantially increased the incidence of reports of non-believed memories for performed-undermined actions, even when a highly stringent classification criterion was used. Indeed, as Figure 1 illustrates, in Session 3 after undermining, the Memory ratings were significantly *greater* than the Belief ratings only for performed-undermined actions, $t(23) = 3.29$, $p < .01$; for other critical actions the Belief and Memory ratings did not significantly differ (for both contrasts, $t < 1.44$, $p > .16$).

3.3 Correlation analysis

As an alternative way of assessing the association between Belief and Memory ratings, we calculated the correlations between these ratings of the critical actions. Each participant rated two of each critical action-type per session; we therefore calculated the correlations by action (2 actions x 24 participants = 48 actions of each type) rather than by participant. Our results showed that in all cases the Belief—Memory correlations were strong. However, for performed-undermined actions the correlation dropped significantly between Session 2 and Session 3 ($r = .90$ and $r = .73$, respectively, $z = 2.48$, $p = .01$). The same was not true for performed actions ($r = .87$ and $r = .91$, respectively, $z = 0.87$, $p = .38$) or new actions ($r = .72$ and $r = .72$, respectively, $z = 0.05$, $p = .96$). These analyses support our above results in showing that the provision of negative misinformation substantially weakened the association between participants' Belief ratings and their Memory ratings.

3.4 Phenomenological data

The final element of our analysis was to look at the phenomenological characteristics of participants' beliefs and memories. To assess whether non-believed memories differ subjectively from other types of belief and memory phenomena, we collapsed the Session 3 data across critical action-types, and categorized all 144 critical actions (6 actions x 24 participants) as either a believed memory (n = 48), a non-believed memory (n = 16), a believed non-memory (n = 3), or a non-believed non-memory (n = 77). Responses to the rating scales were classified as 'beliefs' whenever participants gave Belief ratings of 7 or 8, and as 'memories' whenever they gave Memory ratings of 7 or 8. Thus, instead of defining non-believed memories as before, i.e. in terms of the difference between Memory and Belief ratings, here a non-believed memory is defined specifically as a memory rated 7 or 8, accompanied by a belief rated 6 or below.

We compared the characteristics of the different response types; however, we excluded believed non-memories (events believed to have happened for which no memory was available) from this analysis due to their low frequency. Our approach was thus a 25 (Phenomenological characteristic) x 3 (Response type: non-believed memory vs. believed memory vs. non-believed non-memory) mixed-factor ANOVA. We were particularly interested in whether non-believed memories differed from non-believed non-memories (i.e., comparing non-believed events with vs. without an accompanying memory), and from believed memories (i.e., comparing memories with vs. without an accompanying belief). Our test revealed a significant main effect of response type, $F(2, 138) = 45.72$, $p < .001$, $\eta^2_p = .40$. Following this main effect up with Bonferroni post-hoc tests showed that non-believed non-memories were rated lower across the various memory characteristics than both believed memories and non-believed memories (both ps < .001). In contrast, believed memories and non-believed memories did not significantly differ overall, $p = .18$ (see Figure 2).

It is important to also note, though, that alongside the main effect of response type, our omnibus test also revealed a significant interaction between phenomenological characteristic and response type, $F(48, 3312) = 5.98$, $p < .001$, $\eta^2_p = .08$. To test whether this interaction might be masking significant differences for individual memory characteristics, we conducted a series of one-way ANOVAs (Bonferroni corrected $\alpha = .05/25 = .002$). These revealed significant differences between response types on 18 of the 25 characteristics (i.e., those listed in Figure 2; those that did not differ are listed in the figure caption). Even using this extremely conservative test, several differences between non-believed memories and non-believed non-memories reached significance: clarity of thinking, $t(39.32) = 9.03$, $p < .001$, $d = 1.60$, detail of thinking, $t(91) = 5.31$, $p < .001$, $d = 1.43$, visual details, $t(30.47) = 3.91$, $p < .001$, $d = 0.93$, movements, $t(30.89) = 4.93$, $p < .001$, $d = 1.17$, feeling of mental time-travel, $t(91) = 3.99$, $p < .001$, $d = 1.07$, and coherence of the story, $t(91) = 3.21$, $p < .001$, $d = 0.82$. However, none of the 25 memory characteristics distinguished non-believed memories from believed memories even before the conservative Bonferroni correction was applied (largest $t = 1.93$, smallest $p = .06$). These findings broadly show that participants' non-believed memories for true events were subjectively experienced as being much like their believed memories.

4. Discussion

These findings represent the first systematic creation in an experimental study of non-believed memories for true events, what Mazzoni et al. (2010) called 'disowned memories'. These are mental contents referring to events that truly happened. Individuals continue to experience them as vivid recollections, even though they cease to believe in their veracity. Here we show that reports of disowned memories occur and are relatively easy to induce via

negative misinformation (i.e., socially-conveyed misinformation aimed at denying the occurrence of a real event). After receiving negative misinformation implying they had not performed specific actions that in fact they had performed, participants' beliefs that they had performed those actions dropped substantially. This result indicates how easily social influence affects what people believe happened to them. Work on social influence on memory, and in particular in eyewitness memory, has already shown that in memory tasks people tend to conform to the memory reports of others (e.g., Gabbert et al., 2003; Wright, Self, & Justice, 2000). However most of these studies aimed to make people add elements to a memory report, rather than take away elements from it. Also, they have not separately assessed whether social conformity changes what people believe and whether it affects their sense of recollection of the event. Here we found that the negative misinformation provided by a reliable source (the experimenter) decreased the belief while no significant drop was observed in the reported sense of recollection. As in Clark et al. (2012), the recollective qualities of the non-believed memories for true actions were not substantively different from those of believed memories of true actions. This further confirms that the negative misinformation influenced only the belief and not the perceived quality of the memories.

Clark et al. (2012) showed that false memories are easily undermined, and remarked that it appears to be relatively easy to undermine a belief, yet relatively difficult to undermine what people perceive as a memory. The present findings extend those of Clark et al. (2012) and Otgaar et al. (2013), by demonstrating that also in the case of *true* events, beliefs can be easily undermined and decreased. This extension is important, because it demonstrates that the ability to dissociate belief from memory judgments is not restricted to false memories, which some might consider unusual cases with unique properties. Because the belief can be stripped away from correct judgments of remembering, the current finding provides a firmer

and final demonstration of the independence of subjective evaluations of autobiographical belief and memory (Mazzoni & Kirsch, 2002; Mazzoni et al., 2010). Indeed, it is interesting to note that the proportion of non-believed true memories elicited in this study was remarkably similar to the proportion of non-believed false memories elicited in Clark et al.'s (2012) study. Using the 2 scale-points criterion described above, we obtained a 29% non-believed memory rate, whereas Clark et al. obtained 30%. Using the 3 scale-points criterion, our rate was 21%, compared to Clark et al.'s 25%. The similarity of induced memory endorsement rates for true and false events offers tentative evidence that the malleability of belief—as well as the presence of belief—can be independent of the perceived truth of the memory.

We propose that our present findings are also of methodological importance: they show that experimentally inducing non-believed memories in participants does not require first implanting false memories, which can be both difficult and time-consuming to achieve. In fact, it is noteworthy to observe that around 2% of the initial memories in this study were spontaneously disowned. Although in this study we used this result as the base-rate against which to interpret the effect of the negative misinformation, we need to remind that a similar result was obtained also in Clark et al. (2012), where a small minority of spontaneous non-believed memories were found before the misinformation was given. If we suppose that these cases are more than just extreme error variance, then these spontaneous non-believed memories are interesting in their own right and warrant further investigation.

Our data add also to the ongoing debate regarding the capacity of socially-communicated negative misinformation to undermine event memories. As described in the introduction, some of the studies involving negative misinformation have either found little effect on the outcomes of a memory test (Pezdek & Roe, 1997; Roos af Hjelmsäter et al.,

2008) or have shown that this misinformation can only lower the rate of reporting details of an event (e.g. Gabbert et al., 2003; Merckelbach et al., 2007; Roos af Hjelmsäter et al., 2012). In contrast, ours appears to be the first study to show that negative misinformation can result in people actually denying belief in an event that they experienced. Researchers investigating this topic in the future would benefit from assessing belief and memory ratings separately, to reflect the compelling notion that endorsing an event as 'remembered' does not necessarily mean that the rememberer believes the event occurred. Of course, whereas our experimental paradigm provides excellent control over and knowledge of what really happened, this meant that the experimenter was participants' only source of external information on which to base their metacognitive decisions – a source that also appeared to wield unequivocally correct knowledge of what happened. We know that in real situations where a memory is disputed, people turn to a variety of information sources to verify whether or not their remembered experience really did happen, and these sources vary in terms of how reliable and informative they are perceived to be (Mazzoni et al., 2010; Nash & Takarangi, 2011; Wade & Garry, 2005). Therefore there is an interesting balance for researchers to strike between studying non-believed memories in a rigorous, controlled manner, and in a way that mirrors how these memories typically evolve in naturalistic settings.

Of course, the actions we used in this study were not highly memorable and this methodological feature would undoubtedly be partly responsible for the relatively high proportion of non-believed memories we observed. Nonetheless we expect that the difference between these simple actions and other more salient experiences—such as actions involving objects, or witnessed events—would be one of degree rather than kind. That is to say, whereas it might be more difficult to undermine beliefs in certain kinds of memories, we have no reason to predict that certain types of memory would be immune to stronger and

more compelling forms of negative misinformation. Otgaar et al.'s (2013) findings of non-believed memories of childhood experiences lend support to this reasoning. A further issue of note is that we asked participants to rate their beliefs and memories separately so as to avoid one rating being guided by the other, and memories were always rated first. It would be worthwhile considering in future studies whether evaluating memories first, beliefs first, or both simultaneously best emulates real-life metacognitive monitoring, and how taking these alternative methodological approaches might have influenced the data obtained. Most times, judgments about recollecting a cued event and of believing in its occurrence are made simultaneously. Typically, autobiographical memories are accompanied by the belief of having personally experienced the event (Brewer, 1996), and, as Tulving (1983) notes, the experience of remembering includes both "the feeling that the present recollective experience refers to a past event, and the feeling that the experience is veridical" (p. 187; see also Scoboria et al, in press). We therefore predict that the simultaneous elicitation of the two ratings would maintain the confound between memory and belief in occurrence.

A challenge for this and future studies is to convincingly rule out a demand-effects account of the data - that participants only changed their belief *ratings*, not their actual beliefs (i.e., "I know I did this, but the experimenter says I didn't, so I had better pretend to agree"). We do not think, however, that this account can adequately explain our data. One reason is that a number of participants who substantially lowered their initial belief rating for one performed-undermined action frequently maintained or only slightly reduced their belief rating for the other action – a finding at odds with the idea that they were responding compliantly. Furthermore, because remembering events almost always implies belief in the events' occurrence, we might argue that a compliant participant—who feigns trust in the experimenter's suggestion—should shift their memory ratings downwards as well as their

belief ratings. That is, we would not expect compliant participants to continue saying that they vividly remember, because doing so seems contrary to feigning trust in the suggestion. In line with research showing that compliance can be dissociated from interrogative suggestibility (e.g., Hansen, Smeets, & Jelicic, 2010), this reasoning would suggest that our participants' behaviour was more attributable to interrogative suggestibility, defined as "The extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as a result of which their subsequent behavioural response is affected" (Gudjonsson & Clark, 1986, p. 84). Moreover, our participants' reactions when they were 'debriefed' in Session 3 frequently indicated that they were impressed and surprised about how easily they were tricked by the experimenter. Again, these reactions offered us anecdotal evidence that participants were compelled by the misinformation.

Nonetheless, future research might better address the issue of demand effects by adapting procedures such as Kassin and Kiechel's (1996) 'confederate in the waiting room' technique, or Laney et al.'s (2008) 'Red Herring' technique, both of which are designed to subtly probe for participant suspicion.

We speculate that our results on the ease with which belief in the occurrence of an event can be changed without changing the perceived recollective quality of the memory, in combination with those of many prior studies in this vein (Clark et al., 2012; Mazzoni et al., 2010; Otgaar et al., 2013; Scoboria et al., in press; Scoboria et al., 2004) point toward a distinction between two components in the autobiographical system, which mirror Conway's concepts of coherence and correspondence in his Self-Memory System framework (SMS; Conway, 2005; Conway & Pleydell-Pearce, 2000). The function of the first component (the belief component) is to provide ongoing adjustment to major and minor changes in situation,

context and information, which might be important for quickly updating self-referring information. This interpretation is compatible with Conway's notion of coherence and might explain some rather surprising results showing that memory-related beliefs can be manipulated with extreme ease (e.g., Bernstein, Whittlesea, & Loftus, 2002; Mazzoni, Loftus, Seitz, & Lynn, 1999). Here we have shown that with similar ease beliefs can be decreased when people are provided with negative misinformation, while endorsements of remembering remain unaltered. The belief can be conceived as the attribution of truth to mental contents that might or might not correspond to actual events (see also Scoboria, et al, in press), and fits with Garry, Loftus, and Brown's (1994) description of memory as the "justification of belief" (p. 445). The function of a relatively stable second component (the perceived recollective quality of the memory component) is to provide individuals—changes in belief notwithstanding—with stability and continuity of past experiences, and ultimately, a stable identity. The individual past, which represents the source of identity, is necessarily much slower to change, an idea compatible with Conway's notion of correspondence. As the SMS framework also posits, the ability to negotiate life successfully then depends on the interaction and integration of these two components, one that provides constant adjustment, and one that provides stability.

References

Beltrami, D., & Mazzoni, G. (2011). *Can aIAT detect true and false remote autobiographical memories?* Poster presented at ICOM, York, UK.

Bernstein, D. M., Whittlesea, B. W. A., & Loftus, E. F. (2002). Increasing confidence in remote autobiographical memory and general knowledge: Extensions of the revelation effect. *Memory & Cognition, 30*, 432-438.

Brewer, W. F. (1996). What is recollective memory? In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 19-66). New York: Cambridge University Press.

Blank, H. (2009). Remembering: A theoretical interface between memory and social psychology. *Social Psychology, 40*, 164-175.

Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schachter, D. L. (2001). Can medial temporal lobe regions distinguish true from false" An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences, 98*, 4805-4810.

Clark, A., Nash, R. A., Fincham, G., & Mazzoni, G. (2012). Creating non-believed memories for recent autobiographical events. *PLoS ONE,* e32998. doi:10.1371/journal.pone.0032998.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53,* 594-628.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self memory system. *Psychological Review, 107*, 261-288.

D'Argembeau, A., & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and Cognition, 15*, 342-350.

Dennis, N. A., Bowman, C. R., & Vandekar, S. N. (2012). True and phantom recollection: An fMRI investigation of similar and distinct neural correlates and connectivity. *NeuroImage, 59*, 2982-2993.

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17*, 533-543.

Garoff-Eaton, R. J., Kensinger, E. A., & Schacter, D. L. (2007). The neural correlates of conceptual and perceptual false recognition. *Learning & Memory, 14*, 684-692.

Garry, M., Loftus, E. F., & Brown, S. W. (1994). Memory: A river runs through it. *Consciousness and Cognition, 3*, 438-451.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*, 208-214.

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1*, 83-104.

Hansen, I., Smeets, T., & Jelicic, M. (2010). Further data on interrogative suggestibility and compliance scores following instructed malingering. *Legal & Criminological Psychology, 15*, 221-228.

Hornstein, S. L., & Mulligan, N. W. (2004). Memory for actions: Enactment and source memory. *Psychonomic Bulletin & Review, 11*, 367-372.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp.391-422). Hillsdale, NJ: Erlbaum.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General, 117*, 371-376.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Johnson, M. K., Kahan, T. L., & Raye, C. L. (1984). Dreams and reality monitoring. *Journal of Experimental Psychology: General, 113*, 329-344.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science, 7*, 125-128.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517.

Lampinen, J. M., Odegard, T. N., & Bullington, J. L. (2003). Qualities of memories for performed and imagined actions. *Applied Cognitive Psychology, 17*, 881-893.

Laney, C., Kassa, S. O., Morris, E. K., Berkowitz, S. R., Bernstein, D. M., & Loftus, E. F. (2008). The Red Herring technique: A methodological response to the problem of demand characteristics. *Psychological Research, 72*, 362-375.

Lindner, I., Echterhoff, G., Davidson, P. S., & Brand, M. (2010). Observation inflation: Your actions become mine. *Psychological Science, 21*, 1291-1299.

Lövdén, M., Rönnlund, M., & Nilsson, L. G. (2002). Remembering and knowing in adulthood: Effects of enacted encoding and relations to processing speed. *Aging, Neuropsychology, and Cognition, 9*, 184-200.

Marini, M., Agosta, S., Mazzoni, G., Dalla Barba, G., & Sartori, G. (2012). True and false DRM memories: Differences detected with an implicit task. *Frontiers in Psychology, 3*, 310.

Mazzoni, G. A. L., Loftus, E. F., Seitz, A., & Lynn, S. J. (1999). Changing beliefs and memories through dream interpretation. *Applied Cognitive Psychology, 13*, 125-144.

Mazzoni G., & Kirsch, I. (2002). Autobiographical memories and beliefs: A preliminary metacognitive model. In T. J. Perfect & B. L. Schwartz (Eds.). *Applied metacognition* (pp. 121-146). Cambridge, UK: Cambridge University Press.

Mazzoni, G., & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science, 14*, 186-188.

Mazzoni, G., Scoboria, A., & Harvey, L. (2010). Nonbelieved memories. *Psychological Science, 21,* 1334-1340.

Merckelbach, H., van Roermund, H., & Candel, I. (2007) Effects of collaborative recall: Denying true information is as powerful as suggesting misinformation. *Psychology, Crime & Law, 13*, 575-581.

Nash, R. A., & Takarangi, M. K. T. (2011). Reconstructing alcohol-induced memory blackouts. *Memory, 19*, 566-573.

Nash, R. A., Wade, K. A., & Brewer, R. J. (2009). Why do doctored images distort memory? *Consciousness and Cognition, 18*, 773-780.

Nash, R. A., Wade, K. A., & Lindsay, D. S. (2009). Digitally manipulating memory: Effects of doctored videos and imagination in distorting beliefs and memories. *Memory & Cognition, 37*, 414-424.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125-174). London: Academic Press.

Okado, Y., & Stark, C. (2003). Neural processing associated with true and false memory retrieval. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 323-334.

Otgaar, H., Scoboria, A., & Smeets, T. (2013). Experimentally evoking nonbelieved memories for childhood events. *Journal of Experimental Psychology: Learning, Memory & Cognition, 39*, 717-730.

Pezdek, K., & Roe, C. (1997). The suggestibility of children's memory for being touched: Planting, erasing, and changing memories. *Law & Human Behavior, 21*, 95-106.

Piaget, J. (1951). *Play, dreams, and imitation in childhood* (L. Gattegno & F. M. Hodsen, Trans.). New York: Norton.

Roediger, H. L., III, & Zaromb, F. M. (2010). Memory for actions: How different? In L. Bäckman & L. Nyberg (Eds.), *Memory, aging and the brain: A Festschrift in honour of Lars-Göran Nilsson* (pp. 24-52). New York: Psychology Press.

Roos af Hjelmsäter, E., Granhag, P. A., & Strömwall, L. A. (2012). Co-witness influence on children's memory reports: The difference is in the details. *Journal of Applied Social Psychology, 42*, 320-339.

Roos af Hjelmsäter, E., Granhag, P. A., Strömwall, L. A., & Memon, A. (2008). The effects of social influence on children's memory reports: The omission and commission error asymmetry. *Scandinavian Journal of Psychology, 49*, 507-513.

Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition, 31*, 887-901.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science, 19*, 772-780.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., & Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *NeuroImage, 6*, 259-269.

Scoboria, A., Jackson, D. L., Talarico, J., Hanczakowski, M., Wysman, L., & Mazzoni, G. (in press). The role of belief in occurrence within autobiographical memory. *Journal of Experimental Psychology: General.*

Scoboria, A., Mazzoni, G., Kirsch, I., & Relyea, M. (2004). Plausibility and belief in autobiographical memory. *Applied Cognitive Psychology, 18*, 791-807.

Sheen, M., Kemp, S., & Rubin, D. (2001). Twins dispute memory ownership: A new false memory phenomenon. *Memory & Cognition, 29*, 779-788.

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience, 7*, 664-672.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1-12.

Wade, K. A., & Garry, M. (2005). Strategies for verifying false autobiographical memories. *American Journal of Psychology, 118*, 587-602.

Wade, K. A., Garry, M., Nash, R. A., & Harper, D. N. (2010). Anchoring effects in the development of false childhood memories. *Psychonomic Bulletin & Review, 17*, 66-72.

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology, 91*, 189-202.

Acknowledgement

The research reported in this paper has been partly supported by a British Academy grant to the first author.

Figure Captions

Figure 1. Mean Belief and Memory ratings ascribed to critical actions before (Session 2) and after (Session 3) the undermining procedure. Error bars indicate standard error of the mean.

Figure 2. Phenomenological characteristics that differed between non-believed memories, believed memories and non-believed non-memories. Error bars indicate standard error of the mean. The seven characteristics that did not differ were: smell, taste, location, intensity of current feelings, negative emotions, connections, and talked about before (See Clark et al., 2012, for full item wordings and scale anchors).

Figure 1

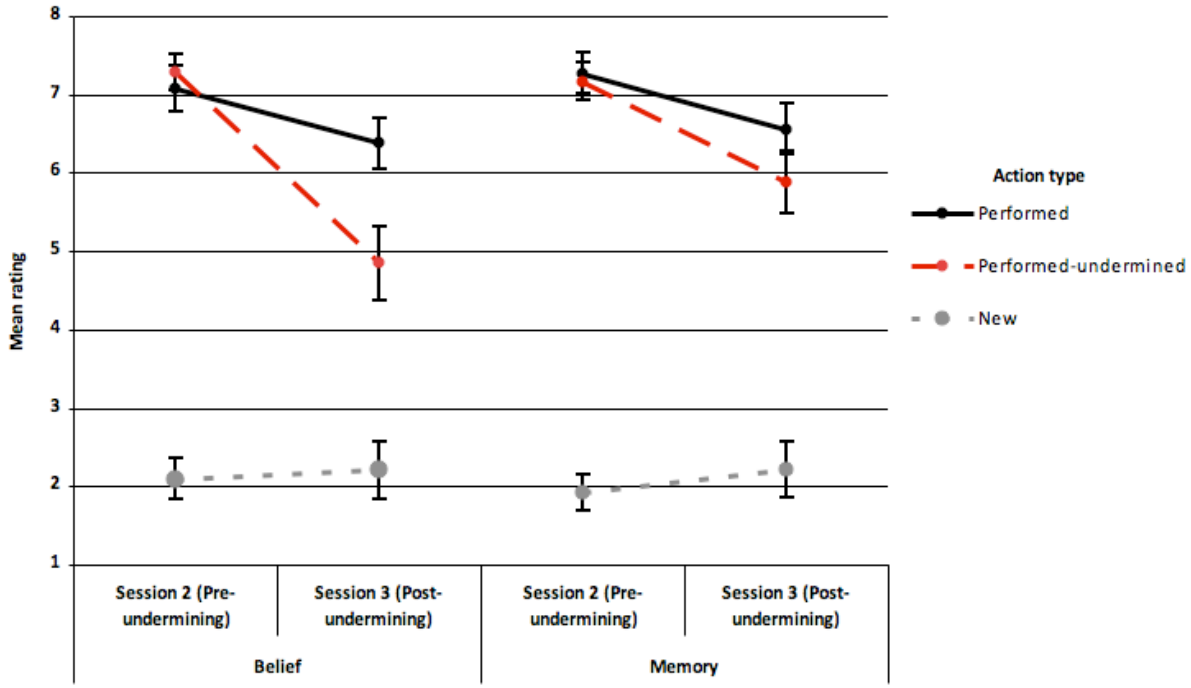

Figure 2