Acta Psychologica 151 (2014) 164–173

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Confidence–accuracy resolution in the misinformation paradigm is influenced by the availability of source cues

Ruth Horry, Lisa-Marie Colton, Paul Williamson *

School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia, 5001, Australia

ARTICLE INFO

Article history:
Received 8 October 2013
Received in revised form 28 May 2014
Accepted 7 June 2014
Available online 28 June 2014

PsycINFO codes:
2343 Learning & Memory
2340 Cognitive Processes

Keywords:
Misinformation effect
Confidence
Resolution
Metacognition
Source monitoring

ABSTRACT

After witnessing an event, people often report having seen details that were merely suggested to them. Evidence is mixed regarding how well participants can use confidence judgments to discriminate between their correct and misled memory reports. We tested the prediction that the confidence–accuracy relationship for misled details depends upon the availability of source cues at retrieval. In Experiment 1, participants ($N = 77$) viewed a videotaped staged crime before reading a misleading narrative. After seven minutes or one week, the participants completed a cued recall test for the details of the original event. Prior to completing the test, all participants were warned that the narrative contained misleading details to encourage source monitoring. The results showed that the strength of the confidence–accuracy relationship declined significantly over the delay. We interpret our results in the source monitoring framework. After an extended delay, fewer diagnostic source details were available to participants, increasing reliance on retrieval fluency as a basis for memory and metamemory decisions. We tested this interpretation in a second experiment, in which participants ($N = 42$) completed a source monitoring test instead of a cued recall test. We observed a large effect of retention interval on source monitoring, and no significant effect on item memory. This research emphasizes the importance of securing eyewitness statements as soon as possible after an event, when witnesses are most able to discriminate between information that was personally seen and information obtained from secondary sources.

1. Introduction

In a seminal study, Loftus, Miller, and Burns (1978) showed that people often incorporate misleading information encountered after a witnessed event into their memory reports of that event. Dozens of studies have replicated this *misinformation effect*, repeatedly showing that participants often report having seen details that were merely suggested (e.g., Belli, Lindsay, Gales, & McCarthy, 1994; Chambers & Zaragoza, 2001; Lindsay, 1990). The metacognitive experiences associated with these errant memory reports have recently come under the empirical spotlight. A key question is whether participants are able to discriminate between their real and errant memory reports after exposure to misinformation. To date, evidence is mixed, with some researchers reporting very poor discrimination (Bonham & González-Vallejo, 2009; Cann & Katz, 2005; Tomes & Katz, 2000), and others reporting reasonably high discrimination (Higham, Luna, & Bloomfield, 2011). In this paper, we focus on the availability of source cues as a moderator of metacognitive discrimination in the misinformation paradigm. We argue that when source cues are relatively accessible, participants are better able to discriminate between their correct and suggested memories, but when source cues are relatively inaccessible, discrimination worsens.

1.1. Source monitoring and misinformation

The basic misinformation paradigm includes three stages. First, participants witness an event. The event is often depicted in slides (e.g., Frost, Ingraham, & Wilson, 2002; Higham et al., 2011; McCloskey & Zaragoza, 1985) or by video (e.g., Bonham & González-Vallejo, 2009; Cann & Katz, 2005), though live events have also been used, including those in which the participant was actively involved (e.g., Eisen, Gomes, Lorber, Perez, & Uchishiba, 2013; Holmes & Weaver, 2010; Sondhi & Gupta, 2007). Second, the participant is exposed to misinformation, which may be embedded in a narrative (e.g., Belli, Windschitl, McCarthy, & Winfrey, 1992; Lindsay, 1990), incorporated into a series of post-event questions (e.g., Chambers & Zaragoza, 2001; Hekkanen & McEvoy, 2002), or presented by another witness (e.g., Meade & Roediger, 2002; Wright, Self, & Justice, 2000). Finally, the participant's memory is tested. Test formats have varied between studies and have included recognition (e.g., Loftus et al., 1978; McCloskey & Zaragoza, 1985), cued recall (e.g., Belli et al., 1994; Thomas, Bulevich, & Chan, 2010), free recall (e.g., Gabbert, Memon, & Wright, 2006), and source memory tests (e.g., Lindsay & Johnson, 1989; Zaragoza & Koshmider,

* Corresponding author. Tel.: +61 8 8201 3644.
E-mail addresses: Ruth.Horry@flinders.edu.au (R. Horry), Paul.Williamson@flinders.edu.au (P. Williamson).

http://dx.doi.org/10.1016/j.actpsy.2014.06.006
0001-6918/

1989). A consistent finding in these studies is that people often report having seen details that were merely suggested to them.

Although it was initially suggested that the original memory trace was irrevocably altered by the misinformation (Loftus et al., 1978; see also Greene, Flynn, & Loftus, 1982; Loftus, 1979), there is now considerable evidence that the original memory trace can co-exist, unaltered, alongside the memory trace for the suggested detail (e.g., Christiaansen & Ochalek, 1983; Lindsay & Johnson, 1989). The upshot of this is that, under the right conditions, the original detail can be retrieved and the harmful influence of the misinformation can be undone (e.g., Gordon & Shapiro, 2012; Wright, 1993).

How might a participant resolve the discrepancy of having two conflicting memory traces available? According to the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993), when information is stored in memory, it is stored alongside various cues that can be used to infer the source of the information. These cues include perceptual details (e.g., visual and auditory details), spatial and temporal information, records of cognitive operations (e.g., elaboration, retrieval), and affective information. Consider a misinformation study in which a participant sees a hammer but later reads that the tool was a wrench. The memory for the hammer may be accompanied by perceptual details concerning its colour, shape, size, location within the scene, and so on. The memory for the wrench, however, may include perceptual details about the typeface in which the word was written, and the sound of the word as it was read. In a memory test, a participant could use these cues to discount the wrench and to correctly respond that the item was a hammer.

From the source monitoring perspective, performing accurately on a memory test following exposure to misinformation depends upon two factors. First, the participant must actively engage in source monitoring at retrieval. Second, diagnostic source cues (i.e., those that reliably differentiate between the event and post-event sources) must be available and accessible. Neither of these conditions, alone, will be sufficient for accurate performance. If a participant has ready access to diagnostic cues yet does not attempt to retrieve them, instead relying on retrieval fluency, then the participant will likely report some misinformation. Conversely, if a participant attempts to source monitor, but there are no (or very few) diagnostic cues available, source monitoring will be unsuccessful, potentially leading to reporting of misinformation. Below we present evidence that 1) participants do not automatically engage in source monitoring in misinformation tasks; and 2) even if participants are attempting to source monitor, the availability of source cues determines the likelihood of misinformation being reported.

First, there is considerable evidence that participants do not automatically engage in source monitoring. For example, test formats that encourage source monitoring typically produce smaller misinformation effects than testing conditions that promote responding based on retrieval fluency. Lindsay and Johnson (1989) showed participants a scene of a cluttered office before presenting them with a narrative containing several incorrect details. The participants were then presented with a list of items including event details and suggested details. Half of the participants made yes/no recognition decisions about whether each item had appeared in the picture, while the remaining participants made source judgments for each item. The proportion of items incorrectly attributed to the scene was .66 for the recognition test participants and only .32 for the source test participants (which was not significantly different from the control participants' error rate of .30). The authors argued that the recognition test participants had responded based on retrieval fluency, and had not engaged in source monitoring. The source test participants, however, could not rely on retrieval fluency, and so had to actively engage in source monitoring.

Further evidence that participants do not automatically engage in source monitoring comes from studies that have warned participants prior to the test that the post-event information contained incorrect information (these are often called *postwarnings*, as they are presented subsequent to the misinformation). Several studies have found that postwarnings reduce the size of the misinformation effect, at least under some conditions (e.g., Chambers & Zaragoza, 2001; Christiaansen & Ochalek, 1983; Echterhoff, Hirst, & Hussy, 2005; Thomas et al., 2010). Because postwarnings are, by definition, presented after the misinformation has been encoded, their effectiveness cannot be due to differential encoding of the misleading details. Rather, the effects must be due to differences in retrieval processes. Specifically, it has been argued that warnings alert participants to the need to monitor the sources of their recollections. Given a postwarning, a participant who retrieves the suggested detail and who also recovers source cues that link the item to the post-event information may continue to search their memory for an alternative response. Without a postwarning, the participant may accept the suggested detail on the basis of its familiarity, thus terminating their memory search before the original detail is retrieved.

Even if a participant attempts to monitor the source of their memories, source misattribution errors will still occur. These errors should be relatively infrequent when diagnostic source cues are readily available, but common when source cues are unavailable. One factor that should have a large impact upon the availability of source cues is retention interval. Several studies have reported that the magnitude of the misinformation effect increases with longer retention intervals (e.g., Frost, 2000; Frost et al., 2002; Holmes & Weaver, 2010; Underwood & Pezdek, 1998). Frost et al. (2002), for example, showed participants a slide sequence in a first session, which was followed by a narrative containing some misleading details. Half of the participants completed a memory test 10 min later, while the remaining participants completed the memory test in a second session one week later. Participants in the one week delay condition were around 30%–40% more likely to report misinformation than participants in the 10 min delay condition. The authors concluded that there were fewer source cues available to participants after a longer delay, increasing the source similarity between the original and suggested details in memory.

In summary, source monitoring appears to play a central role in the production of misinformation errors. Participants will make fewer misinformation errors if they are encouraged to engage in source monitoring, but only if there are diagnostic source cues available at test.

1.2. Metacognitive monitoring and misinformation

Although the mechanisms that underlie the misinformation effect are still under debate, it is clear that participants often report misinformation. An important question, both theoretically and practically, is to what extent participants are able to discriminate between their correct and incorrect memories. A useful statistic for answering this question is *resolution*. Assessing resolution requires that participants respond to a reasonable number of items, assigning a confidence rating to each response. High resolution would be demonstrated if participants consistently assigned higher confidence ratings to their correct responses than to their incorrect responses; lower resolution would be demonstrated if there was considerable overlap in the confidence ratings for correct and incorrect responses.

A handful of studies have examined resolution after exposure to misinformation. Three studies have found reasonably strong resolution for control items but very poor resolution for misled items (Bonham & González-Vallejo, 2009; Cann & Katz, 2005; Tomes & Katz, 2000). In fact, in each of these studies, resolution for misled items was not significantly different from zero, indicating that the participants were unable to discriminate between their correct and misled responses. Tomes and Katz (2000) concluded that after the presentation of misinformation, "confidence becomes useless as an indicator of veracity" (p. 279). However, this conclusion may have been premature. In two experiments, Higham et al. (2011) reported similarly high resolution for misled items as for control items.

What could account for the discrepancy between the results of Higham et al. (2011) and those of prior studies? Higham et al. designed their procedure to encourage source monitoring by providing an