# Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences

Hayden C. Bottoms , Andrea N. Eslick & Elizabeth J. Marsh

Psychology Press
Taylor & Francis Group

# Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences

**Hayden C. Bottoms**
*The University of Nebraska — Lincoln, NE, USA*

**Andrea N. Eslick and Elizabeth J. Marsh**
*Duke University, Durham, NC, USA*

Although contradictions with stored knowledge are common in daily life, people often fail to notice them. For example, in the Moses illusion, participants fail to notice errors in questions such as "How many animals of each kind did Moses take on the Ark?" despite later showing knowledge that the Biblical reference is to Noah, not Moses. We examined whether error prevalence affected participants' ability to detect distortions in questions, and whether this in turn had memorial consequences. Many of the errors were overlooked, but participants were better able to catch them when they were more common. More generally, the failure to detect errors had negative memorial consequences, increasing the likelihood that the errors were used to answer later general knowledge questions. Methodological implications of this finding are discussed, as it suggests that typical analyses likely underestimate the size of the Moses illusion. Overall, answering distorted questions can yield errors in the knowledge base; most importantly, prior knowledge does not protect against these negative memorial consequences.

*Keywords:* Moses illusion; False memory.

People often encounter factual errors in everyday life, such as when a friend mistakenly refers to Anchorage as the capital of Alaska, or a period piece takes liberties with historical fact. These errors are frequently overlooked, even when people are warned about their presence. For example, in one experiment, participants were instructed to press a key whenever they noticed an error embedded in a fictional story. Critically, sometimes the story contradicted well-known facts (e.g., by stating that St. Petersburg is the capital of Russia). Readers only identified a third of the sentences containing errors, although normative data (from Nelson & Narens, 1980) suggested error detection should have been much greater (Marsh & Fazio, 2006). People's relative inability to catch semantic errors can be described as a form of *knowledge neglect* and raises questions about the conditions that improve versus impair error detection.

The Moses illusion is an example of knowledge neglect that is easily obtained in the laboratory.

Participants are asked to detect incorrect pre-suppositions in general knowledge questions such as "*How many animals of each kind did Moses take on the Ark?*" They often fail to notice any problem with this question and respond "*two*". Critically, participants' knowledge is later assessed with questions such as "*Who took two animals of each kind on the Ark: Moses, Noah, or don't know?*" Participants miss many of the incorrect presuppositions during the error detection phase, even though they later demonstrate knowledge of the critical facts.

The Moses illusion is very robust and persists under circumstances intended to facilitate error detection. Participants fall prey to the illusion after seeing a sample error and even though there is no time pressure (Erikson & Mattson, 1981). Error detection is impaired if the incorrect presupposition is semantically (van Oostendorp & de Mul, 1990) or phonologically related to the correct presupposition (Shafto & MacKay, 2000). In contrast, error detection improves when the error appears in the cleft phrase, or main focus, of the sentence (Bredart & Modolo, 1988). Error detection also improves when questions appear in difficult-to-read font, as this reduces processing fluency, which in turn makes material seem less familiar and less true (Song & Schwarz, 2008).

Our focus is on another factor expected to affect ability to detect errors: error prevalence. Across 26 published Moses illusion experiments, the proportion of questions containing errors ranged from 10% (e.g., Barton & Sanford, 1993) to over 50% (e.g., Hannon & Daneman, 2001).[1] As shown in Figure 1, error detection tended to be better in studies in which errors were relatively more common ($r = .45$, $p = .02$; asterisks in the reference section mark papers contributing data points). When a reader encounters more errors, it likely reminds her/him of the need to monitor questions for their accuracy and may also lead the reader to evaluate each individual question more
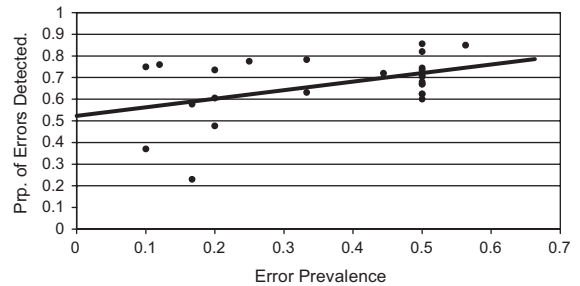


**Figure 1.** Positive relationship between the proportion of experimental questions that were distorted (error prevalence) and the proportion of errors detected. Each data point represents a different experiment, and papers contributing data points are marked with asterisks in the reference section.

stringently. To test this cross-experiment observation, we directly manipulated error prevalence in a single experiment, so that errors were relatively rare for some participants (25% prevalence) and more common for others (50% prevalence).

A second goal was to examine any memorial consequences of the Moses illusion, because in other situations exposure to factual errors affects performance on later tests. For example, reading a story that refers to a prune as a dried date (instead of a plum) increases the likelihood that participants will later answer "*What type of fruit is dried to make a prune?*" with "*date*" (Fazio & Marsh, 2008; Marsh, Meade, & Roediger, 2003). Similar effects occur following multiple-choice testing. For example, answering "*Tenerife is part of what island group: Bahamas, Bikini Islands, Canary Islands, or Cook Islands?*" increases the likelihood that students will later answer that question with one of the previously read multiple-choice lures (Marsh, Roediger, Bjork, & Bjork, 2007; Roediger & Marsh, 2005). In both cases, negative memorial consequences can be linked to a failure to notice the error in the first place: the reader fails to detect the error embedded in the story (e.g., Marsh & Fazio, 2006) and the student selects the wrong answer on the multiple-choice test (e.g., Butler, Marsh, Goode, & Roediger, 2006). The error then comes to mind fluently on later tests, and that retrieval ease is interpreted as truth (Kelley & Lindsay, 1993). Returning to the Moses illusion, failure to detect erroneous presuppositions in questions should have similar effects, leading them to come to mind fluently on later tests. To test this prediction, we added a general knowledge test (with short-answer questions) after the error detection phase and compared performance on new questions to

[1] Based on the published literature, we reported the correlation between error prevalence and detection ability, implying a linear relationship between the two variables. However, an anonymous reviewer correctly noted that this relationship might follow a more complicated quadratic or curvilinear function. Our study does not test this possibility, but rather involves a first step comparing detection of rare versus common errors. Future research should more closely examine detection abilities across varying intermediate levels of error prevalence, in order to evaluate the possibility of a more complicated non-linear relationship between the two variables.

questions corresponding to items that had appeared during the error detection phase.

If the error detection phase affects performance on later tests, it suggests a methodological problem for Moses illusion experiments. Specifically, as described earlier, most Moses illusion experiments conclude with a knowledge check in which participants are explicitly tested on their ability to identify the correct presuppositions. Typically only items answered correctly on the knowledge check are analysed because these are the only items that the experimenter assumes the participant knew during the error detection phase (and should have been able to detect; Erickson & Mattson, 1981). However, if reading errors in the error detection phase influences performance on later tests, then the knowledge check responses are unlikely to be a valid measure of this prior knowledge. Relevant data come from Kamas, Reder, and Ayers (1996), who reported unpublished data showing that the incorrect presuppositions were reproduced on later tests. To examine this more fully, we added a baseline condition to assess knowledge for items never seen in the experiment. If the Moses illusion has memorial consequences, performance on knowledge check questions should differ following undistorted or distorted questions as compared to items appearing for the first time. Furthermore, we analysed our results both the typical way (conditionalised upon correct knowledge check performance) and without conditionalising, to see if the conclusions about error prevalence changed.

To preview, the experiment had three critical phases: an error detection phase, a short-answer general knowledge test, and a final knowledge check. During the error detection phase we manipulated whether errors were rare or common to examine if increased error prevalence aided error detection. During the short-answer test and knowledge check phases we evaluated whether the error detection phase influenced performance on these later tests, by comparing performance on these tests for items that had appeared during the error detection phase to baseline items that had never appeared in the experiment. Overall, our goals were to examine if error prevalence affected error detection and whether failure to detect errors had memorial consequences, with an emphasis on the possible contamination of the prototypical knowledge check.

# METHOD

## Participants

A total of 96 Duke University undergraduates participated for monetary compensation.

## Design

The experiment had a 2 (error prevalence: rare, common) × 3 (error detection question: undistorted, not presented, distorted) × 2 (short-answer: tested, not tested) mixed design. Error prevalence was manipulated between-participants whereas the other variables were manipulated within-participants. Three main dependent variables were analysed: error detection ability, short-answer test performance, and knowledge check selections.

## Materials

A total of 71 Moses illusion questions were modified from published papers (Bredart & Modolo, 1988; Burke, MacKay, Worthley, & Wade, 1991; Buttner, 2007; Erickson & Mattson, 1981; Frick-Horbury & Guttentag, 1998; Hannon & Daneman, 2001; Park & Reder, 2004; Reder & Kusbit, 1991), and 29 were originally constructed. Of these 100 questions, 60 were designated as critical questions and 40 were designated as filler questions. All questions are available from the authors upon request.

Each undistorted critical question (e.g., "*Juneau is the capital of what state?*") was paired with a distorted version that contained a plausible but misleading reference (e.g., "*Anchorage is the capital of what state?*"). Across participants, we counterbalanced which specific items were distorted, undistorted, or not presented during the error detection phase. All participants answered 40 critical questions, but 40 filler questions (all undistorted) were added in the rare error condition (see Table 1). Question order was randomised during the error detection phase and all other tests.

Short-answer general knowledge questions and multiple-choice knowledge check items were created for each critical question. Short-answer questions prompted a specific answer (e.g., "*What is the capital of Alaska?*"). Knowledge check questions paired the short-answer prompt with three

**TABLE 1**
Number of questions presented in each experiment phase

| Error prevalence | Question type | Experiment phase | | |
|---|---|---|---|---|
| | | Error detection | Short-answer | Knowledge check |
| Rare | | | | |
| | Undistorted | 20 | 10 | 20 |
| | Not Presented | 20 | 10 | 20 |
| | Distorted | 20 | 10 | 20 |
| | Filler | 40 | 0 | 0 |
| Common | | | | |
| | Undistorted | 20 | 10 | 20 |
| | Not Presented | 20 | 10 | 20 |
| | Distorted | 20 | 10 | 20 |
| | Filler | 0 | 0 | 0 |

Question type refers to phrasing during error detection phase.

alternatives: the correct answer (e.g., *Juneau*), the misinformation embedded in the distorted question (e.g., *Anchorage*), and "*don't know*". As shown in Table 1, half of the critical questions were tested on the short-answer test (10 each of initially distorted, undistorted, and not presented items). All critical questions were tested on the final multiple-choice knowledge check.

## Procedure

The experiment had four phases: error detection, a filler task, the short-answer test, and the final knowledge check.

At the beginning of the error detection phase, participants were warned that some of the questions contained errors and saw an example: "*You might be asked, 'In what mythology was Venus known as the Goddess of War?' However, Venus was the Goddess of Love, not War.*" Participants were instructed to type "*wrong*" in response to distorted questions and to type "*don't know*" or their answer for all undistorted questions.

During the filler task participants completed Sudoku puzzles for 2 minutes.

Next, participants completed the general knowledge short-answer test. They were warned against guessing and instructed to type "*don't know*" if they did not know an answer.

Finally, participants completed the multiple-choice knowledge check and were debriefed. The experiment took less than an hour to complete.

## RESULTS

Unless otherwise noted, the analyses included only those items answered correctly on the knowledge check and differences were significant at the $p < .05$ level.

## Error detection

Did error prevalence affect performance in the error detection phase? To answer this question, responses were coded as correct, incorrect, detected, or "don't know". Responses were labelled as correct if participants provided the correct answer, as if there were no errors in the question. For example, for the question "*Anchorage is the capital of what state?*", "*Alaska*" would be considered correct and "*Oregon*" would be considered incorrect. If the error was noticed, the response was coded as detected. Two independent coders rated responses (Cohen's kappa = .99), and a third coder resolved all discrepancies.

During the error detection phase participants answered 71% of undistorted questions correctly, and this did not differ across the two error prevalence conditions ($t < 1$). The Moses illusion was observed: participants answered 35% of the distorted questions as if they were correct (e.g., answering "*Anchorage is the capital of what state?*" with "*Alaska*" instead of "*wrong*"). The illusion was similar in the two conditions, with participants in the rarer prevalence condition answering 37% of distorted questions as if they were undistorted, as compared to 32% in the common prevalence condition, $t(94) = 1.42$, $SE = .04$, $p > .15$.

While participants detected some of the errors embedded in questions, they overlooked more than half (although they later demonstrated knowledge for these items). A 2 (error prevalence: rare, common) × 2 (question: undistorted, distorted) ANOVA was computed on the proportion of questions labelled as containing an error. As expected, participants perceived more errors in distorted questions ($M = .42$) than in undistorted questions ($M = .02$), yielding a main effect of question type, $F(1, 94) = 304.56$, $MSE = .03$, $\eta_p^2 = .76$. In addition, participants were more likely to label questions as erroneous when errors were common ($M = .26$) than when they were rare ($M = .19$), yielding a main effect of error prevalence, $F(1, 94) = 7.70$, $MSE = .03$, $\eta_p^2 = .08$. Most importantly, question type and error prevalence
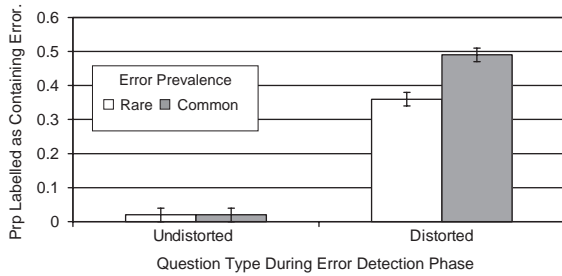
**Figure 2.** Greater error prevalence increases the proportion of distorted questions labelled as containing an error.

interacted, $F(1, 94) = 8.39$, $MSE = .03$, $\eta_p^2 = .08$. As shown in Figure 2, participants rarely labelled undistorted questions as containing errors, and this did not differ across conditions ($t < 1$). In contrast, participants in the common error condition were more likely to detect errors in distorted questions ($M = .49$) than were participants in the rare error condition ($M = .36$), $t(94) = 2.88$, $SE = .05$.

## Memorial consequences: Short-answer responses

Of interest was whether the error detection phase influenced responses on the subsequent general knowledge test. Each short answer was scored as being correct, the misinformation embedded in the distorted question, another wrong answer, or "don't know". Returning to the example "*What is the capital of Alaska*?", "*Juneau*" would be scored as correct, "*Anchorage*" as misinformation, and "*Fairbanks*" as another wrong answer. Two independent coders scored the answers (Cohen's kappa = .99), and a third coder resolved all discrepancies.

Table 2 shows the proportion of short-answer questions answered correctly (left panel) versus with misinformation (right panel). Separate 2 (prior error prevalence: rare, common) × 3 (prior question type: undistorted, not presented, distorted) ANOVAs were computed on the proportion of short-answer questions answered correctly versus with misinformation. For both of these analyses (and the relevant follow-up analyses), one participant was excluded due to missing data. The only significant effects were main effects of prior question type, for both correct, $F(2, 186) = 20.27$, $MSE = .03$, $\eta_p^2 = .18$, and misinformation answers, $F(2, 186) = 18.88$, $MSE = .01$, $\eta_p^2 = .17$. Having answered undistorted questions increased correct responses on the short-answer test ($M = .87$) as compared to the not-presented baseline ($M = .73$), $t(95) = 6.66$, $SEM = .02$. Most importantly, having answered distorted questions increased misinformation answers on the short-answer test ($M = .06$) over that observed for not presented baseline items ($M = .01$), $t(94) = 4.03$, $SEM = .01$.

No other effects were significant. Prior error prevalence did not affect the proportion of short-answer questions answered correctly, $F(1, 93) = 1.51$, $MSE = .05$, $\eta_p^2 = .02$, $p = .22$, nor with misinformation ($F < 1$). No interactions were significant, $Fs < 1$.

Of particular interest was whether successful error detection reduced the likelihood of answering later short-answer questions with misinformation. A 2 (prior error prevalence: rare, common) × 2 (error detected: yes, no) ANOVA was computed on the proportion of short-answer questions answered with misinformation (this analysis was restricted to initially distorted items). A total of 18 participants were excluded because they had missing data for at least one cell in the ANOVA (that is, they might have detected every distorted question, or failed to detect a single distorted question). Overwhelmingly, detection mattered. After failing to detect an error, 9% of short-answer questions were answered with misinformation, as compared to zero following

**TABLE 2**
Proportion of short-answer questions answered correctly (left panel) or with misinformation (right panel), as a function of error prevalence and question type during the error detection phase

| Error prevalence | Proportion correct answers | | | Proportion misinformation answers | | |
|---|---|---|---|---|---|---|
| | *Undistorted* | *Not presented* | *Distorted* | *Undistorted* | *Not presented* | *Distorted* |
| Rare | .86 (.02) | .71 (.03) | .74 (.03) | .01 (.01) | .01 (.01) | .06 (.02) |
| Common | .88 (.02) | .76 (.03) | .76 (.03) | .01 (.01) | .01 (.01) | .05 (.02) |
| *M (SE)* | .87 (.01) | .73 (.02) | .75 (.02) | .01 (.01) | .01 (.01) | .06 (.01) |

Standard errors in parentheses.

successful error detection, $F(1, 76) = 18.37$, $MSE = .02$, $\eta_p^2 = .20$. Neither the main effect of prior error prevalence nor the interaction reached significance (both $Fs < 1$).

## Memorial consequences: Knowledge check

The analyses reported thus far followed the standards in the literature, examining error detection and short-answer performance for items successfully recognised on the knowledge check. However, because the error detection phase affected later short-answer responses, it is important to assess whether it also contaminated the knowledge check. Thus all items were included in the following analyses, to determine whether the earlier phases of the experiment affected performance on the knowledge check.

Table 3 shows the proportion of knowledge check questions answered correctly (left panel) versus with misinformation (right panel). We computed separate 2 (prior error prevalence: rare, common) $\times$ 3 (prior question type: undistorted, not presented, distorted) $\times$ 2 (prior short answer: tested, not tested) ANOVAs on the proportion of knowledge check questions answered correctly versus with misinformation. The table collapses over prior short-answer testing, as that had no effect on the results. The only significant effects were main effects of prior question type for both correct, $F(2, 188) = 79.87$, $MSE = .02$, $\eta_p^2 = .46$, and misinformation answers, $F(2, 188) = 91.89$, $MSE = .02$, $\eta_p^2 = .49$.

We begin by describing correct answer selections. After answering undistorted questions, participants selected more correct answers on the knowledge check ($M = .86$) as compared to the not presented baseline ($M = .78$), $t(95) = 6.66$, $SEM = .01$. More importantly, having answered

distorted questions dropped correct answers ($M = .67$) below the not presented baseline ($M = .78$), $t(95) = 7.03$, $SEM = .02$.

Turning to misinformation selections, having answered undistorted questions decreased misinformation selections on the knowledge check ($M = .04$) as compared to baseline ($M = .07$), $t(95) = 3.41$, $SEM = .01$. Most importantly, after answering distorted questions, misinformation selections increased ($M = .22$) as compared to baseline ($M = .07$), $t(95) = 9.42$, $SEM = .02$.

Of additional interest was whether detecting the error during the error detection phase reduced misinformation selections on the knowledge check. A 2 (prior error prevalence: rare, common) $\times$ 2 (error detected: yes, no) $\times$ 2 (prior short-answer: tested, not tested) ANOVA was computed on the proportion of knowledge check questions answered with misinformation (this analysis was restricted to initially distorted items). A total of 22 participants were excluded due to missing data (that is, they either detected every distorted question, or failed to detect a single distorted question). Again, detection mattered, $F(1, 72) = 140.54$, $MSE = .03$, $\eta_p^2 = .66$. Participants were much more likely to select the misinformation if they had missed the error earlier in the experiment ($M = .27$) than if they had detected the error ($M = .02$), $t(73) = 12.48$, $SEM = .02$. No other effects reached significance.

## Re-analysis of error detection and short-answer data

The error detection phase affected knowledge check selections, calling into question the typical practice of only examining the Moses illusion for facts successfully identified in the knowledge check phase. The error detection analyses are typically restricted to facts correctly identified on

**TABLE 3**

Proportion of knowledge check questions answered correctly (left panel) or with misinformation (right panel), as a function of error prevalence and question type during the error detection phase

| Error prevalence | Proportion correct answers | | | Proportion misinformation answers | | |
|---|---|---|---|---|---|---|
| | *Undistorted* | *Not presented* | *Distorted* | *Undistorted* | *Not presented* | *Distorted* |
| Rare | .86 (.01) | .78 (.02) | .66 (.02) | .04 (.01) | .07 (.01) | .21 (.02) |
| Common | .87 (.01) | .78 (.02) | .68 (.02) | .05 (.01) | .08 (.01) | .22 (.02) |
| *M (SE)* | .86 (.01) | .78 (.01) | .67 (.02) | .04 (.01) | .07 (.01) | .22 (.02) |

Standard errors in parentheses.

the knowledge check in order to show that participants should be able to access the knowledge needed to detect the contradictions. Initially, this logic seems compelling; however, it is complicated by the finding that the error detection phase affected responses on the knowledge check.

In contrast to the earlier analyses that only included items successfully recognised on the knowledge check, we re-analysed the error detection and short-answer data with all questions, regardless of whether the relevant knowledge was correctly identified on the knowledge check. The resulting pattern of data was largely consistent with the results reported thus far. The only inconsistency is noteworthy. As shown in Table 2, answering a distorted question did not reduce later ability to correctly answer short-answer questions ($M = .75$) below the not-presented baseline ($M = .73$). However, when we removed the criterion of answering the knowledge check successfully, previously answering a distorted question reduced performance on the short-answer test ($M = .53$) as compared to baseline ($M = .59$), $t(95) = 2.92$, $SEM = .02$.

## DISCUSSION

Participants detected a greater proportion of errors when errors were more common, but the Moses illusion was equally robust across conditions: Participants answered a third of distorted questions as if they were undistorted. Importantly, exposure to distorted questions had memorial consequences: Participants used the errors embedded in the initial questions to answer later general knowledge questions (including the knowledge check). The general pattern of results was very similar when we analysed all items in contrast to analysing just those answered correctly on the knowledge check (as is typically done). Nevertheless, the knowledge check cannot be considered a pure measure of prior knowledge.

Because we wanted to control for the overall number of errors seen in each condition, we added filler questions to the rare error condition to manipulate error prevalence (see Table 1). In the error detection phase all participants answered 20 distorted and 20 undistorted questions, but only participants in the rare error condition answered the 40 filler questions (which were intermixed with the critical questions). In other words, the length of the error detection task varied across conditions; it was twice as long in

the rare error condition as in the common error condition. It is possible that participants in the rare error condition detected fewer errors because their test was longer, leading to increased fatigue or a failure to remember to monitor for errors. However, participants in the rare error condition detected just as many errors in the last quarter of the test ($M = .34$) as they did in the first quarter ($M = .34$), $t < 1$. Participants consistently detected fewer errors in the rare error condition, even at the very beginning of the test, suggesting that test length did not drive error detection rates.

Although error prevalence affected error detection, it did not affect the memorial consequences of exposure to errors. The right panels of Tables 2 and 3 show similar misinformation production after answering distorted questions, regardless of error prevalence. Given that successful error detection greatly reduced reproduction of errors on later tests, why didn't participants in the common error condition benefit from having detected more errors? Although failure to detect an error was a prerequisite for memorial consequences, most unnoticed errors did not persist to the final tests. Only 9% of undetected errors were reproduced on the short-answer test, and 28% of undetected errors were selected on the knowledge check. Given that the persistence rate was relatively low, error prevalence would have needed to have a larger effect on detection rates to influence memorial consequences. Future research may identify manipulations that affect error detection more dramatically, which in turn should have larger consequences for performance on later tests.

More generally, the finding that the Moses illusion has memorial consequences is an important one. It has long been known that incorrect presuppositions in questions about episodic memories can change people's memories for original events. For example, answering "*How fast was the white sports car going when it passed the barn while travelling along the country road?*" increases later reports of having seen a non-existent barn (Loftus, 1975, p. 566). Our results demonstrate a similar finding within the domain of general knowledge, and add to a growing literature on how people may come to have errors in their knowledge bases. Exposure to a factual error in a story (see Marsh & Fazio, 2007, for a review), a multiple-choice test (see Marsh et al., 2007, for a review), or a question (the present data) can all increase mistakes on later general

knowledge tests. Critically, prior knowledge does not protect people against these semantic illusions. We use the term "knowledge neglect" to describe the phenomenon whereby people have the relevant knowledge to detect an error and yet fail to notice the contradiction.

Finally, it is important to consider the methodological implications of our findings. Consistent with the findings of Kamas et al. (1996), knowledge check performance improved after answering undistorted questions in the error detection phase but was impaired after answering distorted questions. The traditional process of analysing only those items answered correctly on the knowledge check therefore excludes items that participants may have been able to answer before the experiment, and includes items that they may not have been able to answer before the experiment. It is likely that this method *underestimates* the size of the Moses illusion since the percentage of items that may be inappropriately excluded from the analysis is greater than the percentage of items that may be inappropriately included in the analysis. Although analysing the data based on knowledge check performance may not change conclusions about the effects of a manipulation like error prevalence, people's relative inability to catch semantic errors may actually be greater than previously reported.

# REFERENCES

Baker, L., & Wagner, J. L. (1987). Evaluating information for truthfulness: The effects of logical subordination. *Memory & Cognition*, 15, 247–255.*

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21, 477–487.*

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67, 135–144.*

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, 30, 542–579.

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. III. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, 20, 941–956.

Buttner, A. C. (2007). Questions versus statements: Challenging an assumption about semantic illusions. *The Quarterly Journal of Experimental Psychology*, 60, 779–789.*

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning & Verbal Behavior*, 20, 540–551.*

Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, 15, 180–185.

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology*, 111, 43–62.

Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-differences perspective. *Memory & Cognition*, 29, 449–461.*

Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition*, 24, 687–699.*

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1–24.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7, 560–572.

Marsh, E., Meade, M., & Roediger, H. L. III. (2003). Learning facts from fiction. *Journal of Memory and Language*, 49, 519–536.

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, 34, 1140–1149.

Marsh, E. J., & Fazio, L. K. (2007). Learning facts from fiction. In J. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 395–411). New York: Psychology Press.

Marsh, E. J., Roediger, H. L. III, Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, 14, 194–199.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowledge ratings. *Journal of Verbal Learning & Verbal Memory*, 19, 338–368.

Park, H., & Reder, L. M. (2004). Moses illusion: Implications for human cognition. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment, and memory* (pp. 275–292). Hove, UK: Psychology Press.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, 30, 385–406.*

Roediger, H. L. III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1155–1159.

Shafto, M., & MacKay, D. G. (2000). The Moses, Mega-Moses, and Armstrong illusions: Integrating language comprehension and semantic memory. *Psychological Science*, 11, 372–378.*

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*, 26, 791–799.

van Jaarsveld, H. J., Dijkstra, T., & Hermans, D. (1997). The detection of semantic illusions: Task-specific effects for similarity and position of distorted terms. *Psychological Research*, 59, 340–727.*

van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, 74, 35–46.*