

Expertise effects in the Moses illusion: detecting contradictions with stored knowledge

Allison D. Cantor and Elizabeth J. Marsh

Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

ABSTRACT

People frequently miss contradictions with stored knowledge; for example, readers often fail to notice any problem with a reference to the Atlantic as the largest ocean. Critically, such effects occur even though participants later demonstrate knowing the Pacific is the largest ocean (the *Moses Illusion*) [Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning & Verbal Behavior*, 20, 540–551]. We investigated whether such oversights disappear when erroneous references contradict information in one's expert domain, material which likely has been encountered many times and is particularly well-known. Biology and history graduate students monitored for errors while answering biology and history questions containing erroneous presuppositions ("In what US state were the forty-niners searching for oil?"). Expertise helped: participants were less susceptible to the illusion and less likely to later reproduce errors in their expert domain. However, expertise did not eliminate the illusion, even when errors were bolded and underlined, meaning that it was unlikely that people simply skipped over errors. The results support claims that people often use heuristics to judge truth, as opposed to directly retrieving information from memory, likely because such heuristics are adaptive and often lead to the correct answer. Even experts sometimes use such shortcuts, suggesting that overlearned and accessible knowledge does not guarantee retrieval of that information.

ARTICLE HISTORY

Received 16 December 2015
Accepted 5 February 2016

KEYWORDS

Memory; knowledge;
expertise; heuristics

In 2011, Sarah Palin made an off-the-cuff remark that Paul Revere fired warning shots on his famous ride through Boston to warn the "British that they weren't going to be taking away our arms". While the media had a heyday making fun of Palin's version of history, the actual video footage does not show any reaction from the other people present while Palin was talking. Of course, the others in attendance may not have been listening, but another (albeit surprising) possibility is that they simply did not notice her error, even though most Americans are familiar with the story of Paul Revere. Our interest is in understanding why people sometimes miss errors in questions, statements, stories, and other materials, despite having the requisite knowledge stored in memory (*knowledge neglect*; see Marsh & Umanath, 2013 for a review). In particular, we tested the assumption in the literature that expertise would make someone more likely to notice references that contradict stored knowledge (e.g., Reder & Kusbit, 1991), given that knowledge in one's domain of expertise has likely been encountered many times and is easily accessible.

While one might assume that well-known information is retrieved automatically, people fail to leverage stored knowledge in many situations. People often answer questions containing false presuppositions (e.g., "How many

animals of each kind did Moses take on the Ark?"; the *Moses Illusion*; Erickson & Mattson, 1981), fail to notice contradictions in texts (e.g., a reference to "burying survivors"; Barton & Sanford, 1993), and miss factual inaccuracies embedded in stories (e.g., a reference to St. Petersburg as the capital of Russia; Marsh & Fazio, 2006). Critically, in all of these examples the requisite information was stored in memory, as measured in another part of the experiment or estimated via norms. Such lapses matter—perhaps most obviously for the task at hand, like when one blindly follows a recipe even though it contradicts cooking 101 basics. More surprisingly, there are downstream consequences: people sometimes reproduce the errors later, even though they should know the truth. In other words, people who could once answer, "Who built and sailed the Ark?" become more likely to respond with "Moses" (Bottoms, Eslick, & Marsh, 2010), readers increase their belief in statements like "mental illnesses are contagious" (Prentice, Gerrig, & Bailis, 1997), and the Atlantic becomes the largest ocean (Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013). All of these examples are instances of *knowledge neglect*, whereby stored knowledge is not retrieved and/or not applied when needed (Marsh & Umanath, 2013).

Why does knowledge neglect occur? Relevant research on language processing, decision-making, error detection, and truth judgments all converge on the same point: we often rely on a “toolbox” of shortcuts (e.g., Gigerenzer & Goldstein, 1996), both for cognitive efficiency and because these shortcuts normally lead us to the correct answers. For example, when evaluating information, people are biased to assume it is true (Gilbert, Krull, & Malone, 1990) and interpret easy processing (fluency) as evidence for truth (Unkelbach, 2006). These heuristics reflect the world as we normally experience it: in most situations, information is more likely to be true than false (yielding a truth bias) and any one truth has been encountered more frequently on average than any particular falsehood (leading to a fluency heuristic). Even when people do attempt to directly retrieve their knowledge, they often only check whether incoming information generally “matches” their stored knowledge, instead of requiring a perfect match (*Partial Match Theory*; Reder & Kusbit, 1991). Accepting close matches is a practical strategy since information rarely presents itself in the exact same form each time. This issue is exemplified in speech, which is surprisingly error-filled, with slips and ungrammatical references and pauses (e.g., Fox Tree, 1995). The result is that people use past experience to fill in the gaps when processing language, for example, interpreting “the dog was bitten by the man” as referring to a human victim (Ferreira, Bailey, & Ferraro, 2002). Shallow representations filled out by past experience are likely more accurate as well as more efficient than algorithmically processing language.

Situations that encourage the use of these shortcuts are more likely to yield the Moses Illusion. For example, people are less likely to notice an error in a question when it appears in an easy-to-read font, as the easy processing is interpreted as evidence for truth rather than a perceptual byproduct (Song & Schwarz, 2008). Additionally, people are less likely to notice errors that are highly semantically related to the truth (e.g., *Moses* is harder to catch than *Adam*; van Oostendorp & de Mul, 1990), as *Moses* is “close enough” to *Noah* to engender a feeling of fluency. Our reliance on these heuristics is even stronger in environments where much of the information is true, with the consequence being that we are less likely to notice rare contradictions (Bottoms et al., 2010).

Our question is how expertise affects susceptibility to the Moses Illusion. On the one hand, cognitive heuristics are relied upon in many different tasks, regardless of expertise (e.g., Tversky & Kahneman, 1971). Experts make errors, even within their expert domain—for example, one study investigated how the importance of a topic affected the rate with which experts overlooked methodological flaws in research studies (Wilson, DePaulo, Mook, & Klaaren, 1993). But little research has examined the relative benefits of expertise, once differences in basic knowledge are taken into account. That is, it is not surprising if an expert catches a contradiction at a higher rate than the novice if the novice simply does

not have the relevant knowledge stored in memory. However, less clear is whether expertise conveys additional benefits, given that both the expert and novice know something. Put simply, does “knowing something” mean something different for the novice and the expert? That is, even if a novice knows the same fact, it may be less well-learned, less connected to other knowledge, and otherwise qualitatively different from that of the expert. There is an intuitive appeal to claims such as “biblically trained people [would] not false alarm to the Moses question” (Reder & Cleeremans, 1990, p. 249) and that a seasoned cook would notice a problem with a recipe (see also Reder & Kusbit, 1991). Furthermore, as described below, the larger literature on expertise supports the idea that expert knowledge does differ from that of novices, allowing the prediction that expert knowledge may be more protective than that of novices. We tested this idea in three experiments, using graduate students with expertise in biology or history, but not both, asking whether (for example) the biology student would be more likely to question a reference to animal cells containing cell walls than would the history students.

Critically, expert knowledge differs from that of novices’ in three relevant ways. First, an expert knows the information very, very well—she has encountered it many times, in different scenarios, repeatedly retrieving and applying it (Chi, Glaser, & Farr, 1988). For example, chess experts spend approximately 20,000 hours viewing chessboard patterns and can recognise 50,000 patterns (Chase & Simon, 1973). Second, an expert’s concepts are more differentiated from one another. For instance, bird experts are quicker than novices to identify whether a bird is a warbler or finch and can name more attributes of each (Johnson & Mervis, 1997). Third, experts tend to focus on the “deep structure” of information while novices attend to surface features. For example, when sorting physics problems into groups, experts categorise them based on underlying principles (e.g., problems pertaining to “Conservation of Energy”), whereas novices get stuck on literal features (e.g., “problems that have something rotating”) (Chi, Feltovich, & Glaser, 1981).

These principles suggest that experts will be less likely to miss erroneous references within their domain of expertise. First, experts should have more resources available for monitoring, given that they quickly recognise and retrieve information. This is certainly true for visual monitoring; American football experts are more likely to notice changes in complex game scenes than are novices (Werner & Thies, 2000). Additionally, the expert’s heightened differentiation of concepts and tendency towards deeper processing imply that they will be less likely to deem a match “close enough” (as hypothesised by Reder & Cleeremans, 1990). A bible expert, for instance, probably represents “Moses” and “Noah” more distinctly than the novice, making the expert less likely to accept “Moses” as a substitute for “Noah”.

However, expertise can also backfire. For example, experts on American football will correctly recall more animal team names from a presented list (e.g., broncos, dolphins, rams)—but they will also incorrectly report remembering additional animal team names that were not actually studied (e.g., eagles) (Castel, McCabe, Roediger, & Heitman, 2007). Additionally, expert programmers form abstract representations of a programme (i.e., what the programme does) while novices focus on concrete details (i.e., how the programme works), allowing novices to outperform experts on concrete questions (Adelson, 1984). In other words, because experts easily apply their knowledge to new situations, they may be more likely to automatically misread errors as if they were the truth. That is, a bible expert might reflexively respond “two” to the Moses Illusion because she automatically processed the question as if it referred to “Noah”. Reder and Kusbit (1991) ruled out such encoding failures for non-experts, showing that people spent as much time (if not more) reading the erroneous term as the rest of the sentence (see also Kamas, Reder, & Ayers, 1996). However, experts may be more prone to this encoding problem, given the speed of processing differences between experts and non-experts.

To answer these questions, we recruited graduate students with expertise in biology or history, but not both, allowing a within-subjects manipulation of expertise. We created biology and history questions containing false presuppositions that contradicted information all participants should know, regardless of expertise. For example, most people know that the forty-niners were searching for gold and thus have sufficient knowledge to notice the error in the question, “In what state were forty-niners searching for oil?” To confirm our assumption about the materials, we checked that all critical facts were included in introductory textbooks, ensured that pilot undergraduates knew the facts, and individually measured knowledge in our graduate student participants. Accordingly, knowledge in the non-expert domain was akin to that normally examined in past experiments, whereas the knowledge in the expert domain allowed us to examine whether stronger knowledge protected against the illusion.

Experiment 1

Methods

Participants

A total of 131 students in graduate biology and history programmes participated in the online study. As explained below, 16 participants were excluded for indicating on the final survey that they failed to follow instructions (e.g., they looked up answers); 2 additional participants were excluded because their undergraduate minor or double major was in the domain designated as the non-expert control (e.g., a history graduate student who minored in biology). The biology graduate students

($n = 59$, mean year in programme = 3.31) were from Tulane University and Emory University. The history graduate students ($n = 54$, mean year in programme = 3.00) were from Duke University, University of Chicago, University of North Carolina Chapel Hill, and Yale University. Participants received a \$5 Amazon E-Gift Card for completing the 30-minute study.

As instructed by our Institutional Review Board (IRB), we contacted the IRBs of each university for permission to recruit students. After receiving permission, we recruited students in two ways: through Biology and History department chairs (who forwarded the Qualtrics link to their graduate listserv) or via direct contact with students whose email addresses were publically available. In this experiment and the ones that follow, the institutions varied as (1) not all institutions had sizeable graduate programmes in both Biology and History and (2) not all department chairs responded. More generally, the departments varied widely in size and response rate.

Design

The experiment used a 2 (question topic: biology, history) \times 2 (question type: undistorted, distorted) \times 2 (area of graduate study: biology, history) mixed-factorial design with repeated-measures on the first and second factors.

Materials

Six Moses Illusion questions were modified from published papers (Bottoms et al., 2010; Burke, MacKay, Worthley, & Wade, 1991; Büttner, 2007; Park & Reder, 2004; Reder & Kusbit, 1991) and 54 were originally constructed. Half were about Biology and half about History; of the 60 questions, 40 were designated as critical items and 20 as filler items. The filler items never contained incorrect presuppositions and were used to lower the overall prevalence rate of errors since the illusion rate increases when errors are rare (Bottoms et al., 2010).

For each critical item, we created an undistorted question (e.g., *In what US state were forty-niners searching for gold?*), which allowed a correct answer. We changed one critical term from each question to create the distorted version (e.g., *In what US state were forty-niners searching for oil?*); the placement of the critical term within the sentence (i.e., beginning, middle, or end) varied across items. Distorted questions were technically unanswerable as they contained false presuppositions. Each participant only saw either the undistorted or distorted version of each question; assignment to question type was counter-balanced across participants. Biology and history questions were matched on word count ($M = 15.98$, $t(38) = 1.44$, $p = .157$), Flesch reading ease ($M = 54.83$, $t(38) = 1.27$, $p = .211$), and Flesch-Kincaid grade level ($M = 9.52$, $t < 1$). Appendix lists each Moses Illusion question as well as knowledge check accuracy and the illusion rate for each question.¹

Each critical question required knowledge of one fact in order to detect the error (e.g., knowing that *the forty-niners*

searched for gold). To ensure that students would be likely to know the facts, we confirmed that each appeared in introductory-level college textbooks on biology (Campbell et al., 2008), United States history (Boyer et al., 2000), or Western Civilisation (Noble et al., 2002). Furthermore, we ensured that the majority of an undergraduate sample ($n = 19$) knew the facts ($M = .92$) in a pilot study similar to Experiment 1. Finally, 40 multiple-choice knowledge check questions assessed whether each participant knew each critical fact. Each knowledge check question stem (*What precious resources were the forty-niners seeking in California?*) was paired with three alternatives: the correct answer (*gold*), the error from the distorted version (*oil*), and “I don’t know”.

Due to the online nature of the study, an eight-item survey was created to identify participants who did not follow instructions. Specifically, participants were asked to endorse any applicable items such as “I did not complete the experiment alone”, “I am not in a graduate-level History or Biology program”, and “I looked up answers”. Participants also reported their undergraduate majors and minors to allow exclusion of participants with expertise in their non-expert domain.

Procedure

The online experiment began with the error detection phase. There was no attempt to “trick” participants; they knew their job was to look for errors embedded in questions. Participants were instructed that they would be answering a series of general knowledge questions and told that they would encounter questions containing errors; they also received a specific example of the type of error to look for: “You might be asked, ‘In what mythology was Venus known as the Goddess of War?’ However, Venus was the Goddess of Love, not War”. Upon detecting an error, participants were instructed to type “wrong”. Participants were warned against guessing and were told to type, “don’t know” if they did not know the answer. Questions were presented in one of two fixed random orders during the error detection phase. After the error detection phase, participants took the multiple-choice knowledge check. They were warned against guessing and instructed to select “I don’t know” if they did not know the answer. Knowledge check questions were presented in one of two fixed random orders (different from the orders in the error detection phase).

On the final survey, participants were encouraged to respond honestly and were told that their responses would not affect their compensation. Finally, participants were debriefed and instructed to email the experimenter to receive their compensation.

Results and discussion

Coding

Responses to distorted questions were coded into one of three categories: the illusion (a miss; meaning the error

was missed and the question was answered, for example with “California”), detected (a hit; the participant responded with “wrong”), or “don’t know”. Responses to undistorted questions were also coded into three categories: false alarm (the participant responded with “wrong”), answered (the question was answered), or “don’t know”. The first author coded all responses, a second coder coded 10% of trials, and inter-rater reliability was very high ($\kappa = .98$).

Analytic choices

As noted in prior research, “don’t know” responses during the error detection phase are ambiguous (e.g., Bredart & Modolo, 1988; Shafto & MacKay, 2000) and therefore were excluded from analyses. Across experiments, “don’t know” responses during the error detection phase were infrequent (E1: $M = .06$, E2: $M = .06$, E3: $M = .07$). They were slightly higher in participants’ non-expert domain, and this expertise pattern was similar for undistorted and distorted questions.

Analyses in all experiments collapsed across question topic and area of graduate study to allow within-subjects comparisons of expert versus non-expert performance. The same patterns of results were obtained when question topic and area of graduate study were included as factors.

Finally, all illusion rate analyses only included those items that participants responded correctly on the multiple-choice knowledge check, on an individual basis.² Therefore, participants demonstrated requisite knowledge to detect the error for each of the questions included.

Knowledge check

Overall, as desired, knowledge check performance was very high with participants demonstrating knowledge for .92 (95% CI = [.91, .93]) of critical facts. While participants demonstrated more knowledge in their expert domain ($M = .96$, 95% CI = [.95, .97]), performance in the non-expert domain was very high ($M = .88$, 95% CI = [.86, .91]), $t(112) = 7.49$, $SEM = .01$, $p < .001$, $d = .84$.

Moses illusion

How often did participants exhibit the illusion (i.e., miss errors, answering distorted questions as if they were undistorted) in their expert and non-expert domains? These data appear in Figure 1 (left panel). Collapsing across the student’s domain of expertise, participants missed erroneous references in almost half of the distorted questions ($M = .43$; 95% CI = [.38, .47]). Critically, expertise helped: illusion rates were higher in participants’ non-expert domain, $t(111) = 4.78$, $SEM = .02$, $p < .001$, $d = .40$.³ While expertise attenuated knowledge neglect, experts still showed the illusion .37 (95% CI = [.32, .42]) of the time, a rate significantly different from zero, $t(112) = 14.96$, $p < .001$, $d = 1.41$.

To confirm that expertise conferred an improvement in detection rather than a change in response bias (Kamas et al., 1996), we checked whether false alarm rates were

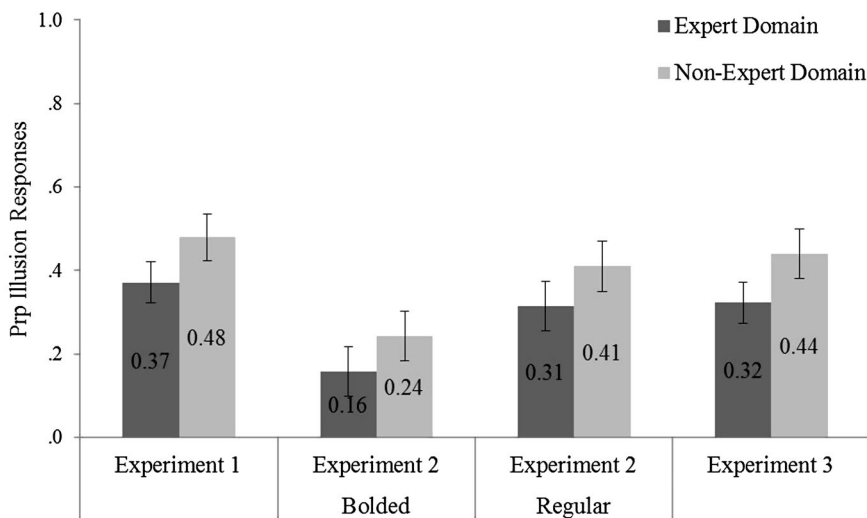


Figure 1. Proportion of distorted questions answered as if they were undistorted during the error detection phase. Data are conditionalised upon accurate knowledge check performance. Proportions exclude “don’t know” responses. Error bars represent 95% confidence intervals.

higher in participants’ expert domain. Overall, participants very rarely responded with “wrong” to an undistorted question ($M = .04$, 95% CI = [.03, .05]), and this rate did not differ as a function of expertise, $t < 1$.

Experiment 1 uncovered a new benefit of expertise: experts were less susceptible to knowledge neglect. Participants were less likely to answer questions that contradicted their expert knowledge and better able to label them as containing errors. However, experts were far from perfect, failing to bring their knowledge to bear about a third of the time.

Experiment 2

Why did participants sometimes fall for the illusion in their expert domain? Given that expertise overall conveyed a benefit, one possibility is that experts sometimes “read over” errors, processing the questions as if they contained the correct references. As discussed earlier, encoding failures are not believed to account for the illusion in non-experts because (non-expert) participants spend as much time reading the error as other words in the question (Reder & Kusbit, 1991). Furthermore, physically highlighting errors does not eliminate the illusion in non-experts (Kamas et al., 1996) even though such signalling cues are excellent directors of attention (Lorch, Lorch, & Klusewitz, 1995).

However, it would be premature to dismiss the encoding failure hypothesis for experts. Because experts process information quickly and relatively automatically, they might be especially prone to accidentally “reading over” errors. To test these ideas in Experiment 2, we manipulated between-subjects whether the critical term in each question was bolded and underlined. If drawing attention to errors eliminates the illusion in experts, or at least benefits experts more than non-experts, this would

suggest that encoding failures play a role when experts exhibit the illusion. Experiment 2 also allowed a replication of Experiment 1 with graduate students from different institutions.

Methods

Participants

A total of 113 graduate students in Biology and History programmes participated in the online study; the same recruitment methods were used as in Experiment 1. The data from 19 participants were excluded based on their final survey responses. The biology ($n = 49$, mean year in programme = 3.24) and history graduate students ($n = 45$, mean year in programme = 3.84) were from the Ohio State University, University of South Carolina, University of California Santa Barbara, University of Minnesota, University of Washington, and University of California, Berkeley. History graduate students were also recruited from Temple University. Participants were compensated with a \$5 Amazon E-Gift Card for the 30-minute study.

Design

The experiment used a 2 (question topic: biology, history) \times 2 (question type: undistorted, distorted) \times 2 (area of graduate study: biology, history) \times 2 (critical term: bolded, regular) mixed-factorial design with repeated-measures on the first and second factors.

Materials

The materials and counterbalancing were the same as in Experiment 1. In the bolded condition, the critical term from each question was bolded and underlined. For distorted questions, the critical term was the incorrect presupposition (e.g., “In what US state were the forty-niners searching for oil”); in undistorted questions, the critical

term was the corresponding accurate presupposition (e.g., “In what US state were the forty-niners searching for **gold**”). One important term from each filler question was bolded and underlined.

Procedure

The procedure was identical to Experiment 1. Participants received no additional instructions about the bolded terms as prior research demonstrated this was unnecessary (Kamas et al., 1996).

Results and discussion

Coding

The coding scheme was the same as Experiment 1. The first author coded all responses; a second coder scored 10% of trials and inter-rater reliability was very high ($\kappa = .97$).

Knowledge check

Participants demonstrated knowledge for .93 (95% CI = [.92, .94]) of critical facts. We computed a 2 (expertise: expert, non-expert) \times 2 (critical term: bolded, regular) ANOVA on the proportion of knowledge check questions answered correctly. Consistent with Experiment 1, participants correctly answered more questions from their expert ($M = .97$, 95% CI = [.96, .98]) than non-expert domain ($M = .90$, 95% CI = [.88, .92]), $F(1,92) = 49.56$, $MSE = .01$, $p < .001$, $\eta^2 = .35$. No other effects were significant ($F_s < 1$).

Moses illusion

Experiment 1 demonstrated that people fell for the illusion less frequently in their expert domain; our question here was whether drawing attention to the errors magnified the benefits of expertise. We computed a 2 (expertise: expert, non-expert) \times 2 (critical term: bolded, regular) ANOVA on the illusion rate. Replicating Experiment 1 and as shown in Figure 1 (centre panel), participants fell for the illusion less frequently in their expert domain, $F(1, 92) = 17.73$, $MSE = .02$, $p < .001$, $\eta^2 = .16$.

Drawing attention to the error helped, in that participants fell for the illusion less often when the error was bolded ($M = .20$, 95% CI = [.15, .25]) than when it was not ($M = .36$, 95% CI = [.31, .41]), $F(1, 92) = 18.99$, $MSE = .07$, $p < .001$, $\eta^2 = .17$. Interestingly, this benefit of bolding was similar regardless of whether participants were answering questions from their expert or non-expert domain, $F < 1$. Thus, while drawing attention to the error attenuated the illusion, it did so similarly regardless of expertise. Furthermore, even when the error was bolded and underlined, experts still sometimes fell for the illusion ($M = .16$, 95% CI = [.10, .22]), at a rate significantly different from zero, $t(46) = 6.63$, $p < .001$, $d = .97$.

To rule out the possibility that expertise and bolding reduced the illusion due to a shift in response bias, we also analysed false alarm rates. Participants responded with “wrong” to undistorted questions .04 of the time

(95% CI = [.03, .06]), and this rate was unaffected by expertise, $F(1, 92) = 1.35$, $MSE = .01$, $p = .248$, or by bolding, $F(1,92) = 2.98$, $MSE = .01$, $p = .088$.

Drawing attention to the erroneous presuppositions helped—but this benefit was similar regardless of expertise. This result means that encoding failures cannot explain all instances of the illusion in one’s expert domain. Rather, the data fit well with claims that heuristics are used in many situations to judge truth, as opposed to directly retrieving information from memory. Furthermore, people sometimes rely on these heuristics even when evaluating overlearned and accessible knowledge; expertise does not guarantee retrieval of that information.

Experiment 3

Experiments 1 and 2 established that expertise confers a protective benefit, but that the illusion still occurs in the face of expertise. It is more likely, however, that expert knowledge protects against downstream consequences of exposure to errors. That is, even if experts miss a few contradictions, we expect them to be able to retrieve the truth when later asked directly. We base this prediction on experiments with older adults who can be considered “knowledge experts” because they have acquired their knowledge over a lifetime (e.g., Perlmutter, 1988), often outperforming young adults on vocabulary and general knowledge tests (e.g., Botwinick & Storandt, 1980; Mitchell, 1989; Perlmutter, 1978). These natural “knowledge experts” sometimes miss contradictions with stored knowledge, but rarely reproduce errors on later tests (e.g., Umanath & Marsh, 2012). We predict that experts will behave similarly to older adults—even though experts will occasionally fail to detect errors, they should be less likely than non-experts to reproduce errors. We added a short-answer general knowledge test to Experiment 3 to assess this prediction.

Methods

Participants

A total of 102 graduate students in Biology and History programmes participated in the online study; recruitment was the same as in Experiments 1 and 2. The data from 17 participants were excluded based on their final survey responses. The biology ($n = 51$, mean year in programme = 3.26) and history graduate students ($n = 34$, mean year in programme = 3.92) were from University of Michigan, University of Illinois at Urbana-Champaign, University of Texas at Austin, and Harvard University. Participants were compensated with a \$7 Amazon E-Gift Card for the 45-minute study.

Design

The experiment used a 2 (question topic: biology, history) \times 2 (question type: undistorted, distorted) \times 2 (area of graduate study: biology, history) mixed-factorial design with repeated-measures on the first and second factors.⁴

Materials

The same materials and counterbalancing were used as in Experiment 1 except that two biology and two history questions were removed as it was not possible to generate suitable short-answer questions for these items. Two filler questions were also removed to maintain the error prevalence rate from Experiment 1.

The 36-item short-answer general knowledge test was used to assess whether participants would reproduce the errors. For the majority of the questions, the question stem from the multiple-choice knowledge check was used (e.g., *What precious resources were forty-niners seeking in California?*). Some question stems were slightly modified to allow a short-answer format; we used any modifications for the multiple-choice knowledge check questions as well.

Procedure

The procedure was identical to Experiment 1 except for the addition of the short-answer general knowledge test between the error detection phase and the multiple-choice knowledge check. Again, participants were warned against guessing on all tests and were instructed to respond “don’t know” when they did not know the answer.

Results and Discussion

Coding

The same coding scheme was used as in Experiments 1 and 2. Short-answer responses were coded as correct (e.g., “gold”), the error embedded in the question (e.g., “oil”), another wrong response (e.g., “silver”), or “don’t know”. The first author coded all responses; a second coder coded 10% of trials and inter-rater reliability was very high ($\kappa = .96$).

Knowledge check

Overall, participants demonstrated knowledge for most of the critical facts ($M = .91$, 95% CI = [.90, .93]). Participants demonstrated more knowledge in their expert ($M = .96$, 95% CI = [.94, .97]) than non-expert domain ($M = .87$, 95% CI = [.84, .90]), $t(91) = 5.58$, $SEM = .02$, $p < .001$, $d = .74$.

Moses illusion

Participants fell for the illusion .38 (95% CI = [.33, .43]) of the time. The data appear in Figure 1 (right panel). Again, even though participants fell for the illusion more frequently in their non-expert domain, $t(91) = 4.40$, $SEM = .03$, $p < .001$, $d = .44$, they still showed some susceptibility to the illusion in their expert domain ($M = .32$, 95% CI = [.27, .37]), a rate significantly different from zero, $t(91) = 12.95$, $SED = .02$, $p < .001$, $d = 1.35$. The benefits of expertise reflected improvements in detection rather than changes in response bias, since false alarm rates were low ($M = .04$,

95% CI = [.03, .05]), and unaffected by expertise, $t(91) = 1.22$, $SEM = .01$, $p = .227$.

Short-answer general knowledge test

Did exposure to errors have any downstream consequences? We computed a 2 (question type: undistorted, distorted) \times 2 (expertise: expert, non-expert) ANOVA on the proportion of target error responses reported on the short-answer general knowledge test. Table 1 shows these data, as well as the proportion of questions answered correctly. Although most questions were answered correctly, enough questions were answered with the target errors to show a clear pattern. Participants almost never produced the error if they were not exposed to it during the error detection phase ($M = .002$, 95% CI = [.000, .004]), and this rate was not significantly different from zero, $t(92) = 1.73$, $p = .086$. Rather, participants were more likely to produce the error after reading it in the error detection phase ($M = .05$, 95% CI = [.03, .06]), $F(1, 91) = 48.60$, $MSE = .02$, $p < .001$, $\eta^2 = .16$. This pattern held in both expert $t(91) = 4.03$, $SED = .01$, $p < .001$, $d = .61$ and non-expert domains, $t(91) = 6.02$, $SED = .01$, $p < .001$, $d = .90$. However, the memorial consequences of exposure to errors were larger in participants’ non-expert domain as reflected in the expertise \times question type interaction, $F(1, 91) = 9.02$, $MSE = .01$, $p = .003$, $\eta^2 = .03$.

Not only were experts less likely to exhibit the illusion, they were also less likely to use erroneous presuppositions to answer later general knowledge test questions. However, experts did occasionally repeat errors, albeit at a rate that was low ($M = .03$, 95% CI = [.02, .04]), but statistically different from zero, $t(91) = 4.74$, $p < .001$, $d = .49$.

General discussion

Expertise helped, meaning that having strong and well-connected knowledge stored in memory promotes error detection. However, expertise did not completely solve the problem: graduate students sometimes answered distorted questions, failing to leverage their expert knowledge about a third of the time and even occasionally reproducing the errors on a later test. Bolding and underlining errors helped, but the benefits were the same regardless of expertise, arguing against an explanation based on encoding failures. Furthermore, item effects cannot explain the results; collapsing across experiments, experts were less likely to answer distorted questions than non-experts for 35 of the 40 questions (see Appendix).

Instead, our results support arguments about the frequency with which people use heuristics to judge truth (Unkelbach, 2006), as opposed to directly retrieving information from memory—even experts sometimes use such shortcuts. Exhibiting the illusion should be viewed as a side effect of an otherwise adaptive system. That is, it is beneficial for a system to be able to handle imperfect input, and to do so quickly. Most of the time, these shortcuts lead one to the correct answer, more efficiently that

Table 1. *Memorial consequences.*

	Correct		Target error	
	Undistorted	Distorted	Undistorted	Distorted
Experts	.96 (.01)	.92 (.01)	.003 (.002)	.03 (.006)
Non-experts	.92 (.01)	.83 (.02)	.001 (.001)	.06 (.01)

Note: Proportion of correct and target error responses to short-answer general knowledge questions after viewing the undistorted and distorted versions of the questions. Data are from Experiment 3. Proportions do not sum to 1 because “don’t know” and other wrong responses are not included in the table. Standard error is noted in parentheses.

attempting direct retrieval. Knowledge neglect, in this case, is a side effect of our otherwise adaptive tendency to handle distorted inputs.

We do not believe that using graduate students (rather than tenured biologists and historians) drove our results. Even if these graduate students had not yet had the 10 years of intense preparation needed to be considered an expert (e.g., Ericsson & Charness, 1994), they were well on their way to developing expertise. Our definition of expertise is similar to (or stronger than) that used in many other studies (e.g., Arkes, Hackett, & Boehm, 1989; Boehm, 1994; Brandt, Cooper, & Dewhurst, 2005; Van Overschelde & Healy, 2001; Weiser & Shertz, 1983). Regardless of one’s definition, one can always make an argument that true “knowledge” is stronger than the way it was measured, but the point is that the graduate students clearly should have had much stronger knowledge in their field than the typical undergraduate. The results clearly show having more knowledge in a field helps, although it does not guarantee all errors will be caught or that they will never be reproduced.

Finally, these results highlight that the effects of expertise on cognition are not as simple as one might think. The literature often focuses on understanding experts’ high levels of performance (see Chi et al., 1988); as discussed earlier, explanations of expert performance include not only a greater quantity of knowledge—rather, their knowledge is more differentiated and stronger in memory, and often guides them to use different strategies. However, these characteristics do not mean that expert cognition is always perfect. For example, the instructional techniques that help novices can actually hinder learning for domain experts (*expertise reversal effect*; Kalyuga, Ayres, Chandler, & Sweller, 2003). The same processes that can help experts outperform novices can occasionally lead them astray. Even though domain experts may remember more, their ability to “fill-in-the-gaps” will also cause them to recall related but not-presented information (Castel et al., 2007). Similarly, in the present research, expertise helped in general, but did not mean that experts never relied upon heuristics—occasionally leading them to miss errors.

Notes

1. When analysed at the question level, illusion rates did not differ significantly based on question characteristics including word

count, reading ease, grade level, or position of the erroneous term within the question (i.e., beginning, middle, or end).

2. In all three experiments, the pattern of results and significant effects did not change when data were not conditionalised upon accurate knowledge check performance.
3. One participant was excluded for having zero observations in the non-expert domain. This participant answered six knowledge check questions in the non-expert domain correctly, and for each, responded “don’t know” during the error detection phase.
4. Note that we did not include a neutral condition where the questions were not presented in the error detection phase (see Bottoms et al., 2010), in order to maximise observations for the undistorted and distorted question types. The benefit of including a neutral condition is that these questions can serve as a baseline of misinformation answers produced on the short-answer-test. Importantly, Bottoms and colleagues found no difference in the proportion of misinformation responses produced between the undistorted and neutral conditions (both $M_s = .01$). That is, reading the undistorted questions did not decrease misinformation responding. Thus, we decided to use the undistorted condition as a comparison to the distorted condition, with any differences in misinformation responses assumed to reflect an increase from baseline for the distorted condition rather than a decrease from baseline for the undistorted condition.

Acknowledgements

We thank members of the Marsh Lab and Lynne Hasher for helpful comments on the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by a Collaborative Activity Award from the James S. McDonnell Foundation’s Twenty-first Century Science Initiative in Bridging Brain, Mind and Behavior (EJM).

References

- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 483–495.
- Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, 2, 81–94.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21, 477–487.
- Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin*, 20, 285–293.
- Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, 18, 670–678.
- Botwinick, J., & Storandt, M. (1980). Recall and recognition of old information in relation to age and sex. *Journal of Gerontology*, 35, 70–76.
- Boyer, P. S., Clark, C. E., Kett, J. F., Salisbury, N., Sitkoff, H., & Woloch, N. (2000). *The enduring vision*. Boston, MA: Houghton Mifflin Company.
- Brandt, K. R., Cooper, L. M., & Dewhurst, S. A. (2005). Expertise and recollective experience: Recognition memory for familiar and unfamiliar academic subjects. *Applied Cognitive Psychology*, 19, 1113–1125.

- 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 in young and older adults? *Journal of Memory and Language*, *30*, 542–579.
- Büttner, A. C. (2007). Questions versus statements: Challenging an assumption about semantic illusions. *The Quarterly Journal of Experimental Psychology*, *60*, 779–789.
- Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2008). *Biology*. San Francisco, CA: Pearson, Benjamin Cummings.
- Castel, A. D., McCabe, D. P., Roediger, H. L., & Heitman, J. L. (2007). The dark side of expertise domain-specific memory errors. *Psychological Science*, *18*, 3–5.
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York, NY: Academic Press.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.
- Chi, M. T., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747.
- 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., Barber, S. J., Rajaram, S., Ornstein, P. A., & Marsh, E. J. (2013). Creating illusions of knowledge: Learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General*, *142*, 1–5.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709–738.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*, 601–613.
- Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*, 248–277.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31.
- 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.
- Lorch, R. F., Lorch, E. L., & Klusewitz, M. A. (1995). Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology*, *20*, 51–64.
- 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., & Umanath, S. (2013). Knowledge neglect: Failures to notice contradictions with stored knowledge. In D. N. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Cambridge, MA: MIT Press.
- Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 31–49.
- Noble, T. F. X., Strauss, B. S., Osheim, D. J., Neuschel, K. B., Cohen, W. B., & Roberts, D. D. (2002). *Western civilization: The continuing experiment*. Boston, MA: Houghton Mifflin Company.
- van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74*, 35–46.
- 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: Psychology Press.
- Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology*, *14*, 330–345.
- Perlmutter, M. (1988). Cognitive potential throughout life. In J. E. Birren & V. L. Bengtson (Eds.), *Emergent theories of aging* (pp. 247–268). New York, NY: Springer.
- Prentice, D. A., Gerrig, R. J., & Bailis, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review*, *4*, 416–420.
- Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: The Moses illusion revisited. In A. Graesser & G. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 233–258). New York, NY: Academic 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.
- 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.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.
- Umanath, S., & Marsh, E. J. (2012). Aging and the memorial consequences of catching contradictions with prior knowledge. *Psychology and Aging*, *27*, 1033–1038.
- Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*, 339–345.
- Van Overschelde, J. P., & Healy, A. F. (2001). Learning of nondomain facts in high- and low-knowledge domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1160–1171.
- Weiser, M., & Shertz, J. (1983). Programming problem representation in novice and expert programmers. *International Journal of Man-Machine Studies*, *19*, 391–398.
- Werner, S., & Thies, B. (2000). Is “change blindness” attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Visual Cognition*, *7*, 163–173.
- Wilson, T. D., DePaulo, B. M., Mook, D. G., Klaaren, K. J. (1993). Scientists' evaluations of research: The biasing effects of the importance of topic. *Psychological Science*, *4*, 322–325.

Appendix

List of all Moses Illusion questions used in the error detection phase. For each question, we report knowledge check performance and the illusion rate collapsing across the three experiments. For the knowledge check, the reported means indicate the proportion of correct selections on the multiple-choice knowledge check measure. For the illusion rate, the reported means indicate the proportion of distorted questions answered as if undistorted, conditionalised upon accurate knowledge check performance and excluding “don't know” responses.

Domain	Question	Undistorted term	Distorted term	Knowledge check experts	Knowledge check non-experts	Moses illusion experts	Moses illusion non-experts
Biology	Water contains two atoms of XXX and how many atoms of oxygen?	Hydrogen	Helium	0.99	.99	.21	.43
Biology	Food that is left uncontained for too long can begin to grow a XXX known as?	Fungus	Algae	0.96	.84	.26	.50
Biology	What is the biological term for breathing, in which humans exchange carbon XXX for oxygen?	Dioxide	Monoxide	0.99	.91	.28	.38
Biology	During which trimester can a single embryo split to form XXX twins?	Identical	Fraternal	0.96	.85	.47	.62
Biology	What is the name used to classify meat-eaters, who act as XXX and eat other animals?	Prey	Predators	0.99	.99	.39	.30
Biology	A half-life refers to the number of years it takes for what percentage of a sample of an isotope to XXX?	Decay	Grow	1.00	.92	.05	.15
Biology	The XXX of which vitamin, commonly found in citrus fruits, causes impaired immunity and weakness?	Deficiency	Excess	0.97	.99	.11	.14
Biology	What is the male sex hormone produced in the XXX that promotes increased muscle, bone mass, and hair growth?	Testes	Prostate	0.75	.66	.55	.63
Biology	Which defining feature of primates, from the class of XXX mammals, enables them to grasp and handle objects?	Vertebrate	Invertebrate	1.00	.95	.33	.48
Biology	What is the name of the process in which plants use the green chemical XXX and sunlight to make food?	Chlorophyll	Chlorine	0.99	.98	.19	.50
Biology	What is the name of the negatively-charged particles that orbit the atomic nucleus, which contains a mix of neutrons and XXX?	Protons	Positrons	0.98	.94	.24	.44
Biology	XXX resources such as coal, petroleum, and natural gas are known as what type of fuels?	Non-renewable	Renewable	0.98	.96	.21	.29
Biology	What term describes the profound change in an organisms' form, like as a caterpillar changes to a butterfly and a tadpole changes to a XXX?	Frog	Lizard	1.00	.98	.45	.51
Biology	A gill is a XXX organ found in what aquatic animal?	Respiratory	Digestive	0.98	.98	.15	.24
Biology	During fertilisation, an egg unites with what to form a XXX?	Zygote	Embryo	0.74	.44	.79	.88
Biology	What is the name for a shot, made from a weakened or killed pathogen, which helps XXX diseases such as small pox?	Prevent	Cure	0.99	.96	.63	.67
Biology	What is the name for the breakdown of food, accomplished in humans by organs such as the oesophagus and the XXX intestines?	Two	Three	0.92	.91	.30	.38
Biology	Sex chromosomes, one of the XXX pairs of human chromosomes, can either be xx for females or what for males?	23	22	0.91	.55	.23	.44
Biology	On which continent did our own species known as Homo XXX originate?	Sapiens	Erectus	1.00	.98	.38	.56
Biology	What is the rigid outer boundary of an XXX cell called?	Plant	Animal	0.98	.66	.26	.50
History	Which famous Paul is credited with alerting Colonial militia of approaching British forces during the XXX War?	Revolutionary	Civil	0.99	.93	.24	.26
History	In what US state were the forty-niners searching for XXX?	Gold	Oil	0.96	.91	.27	.30
History	What is the nationality of XXX inventor of the telephone?	Alexander Graham Bell	Thomas Edison	0.94	.94	.55	.54

(Continued)

Continued.

Domain	Question	Undistorted term	Distorted term	Knowledge check experts	Knowledge check non-experts	Moses illusion experts	Moses illusion non-experts
History	What name describes the form of government of the US, the birthplace of which was XXX?	Athens	Sparta	0.94	.81	.28	.28
History	Which famous stone is used to interpret ancient Egyptian writing known as XXX?	Hieroglyphics	Cuneiform	0.95	.90	.48	.65
History	Who was the leader of the Nazi party which was responsible for the deaths of about 6 million XXX Jews?	European	American	1.00	.99	.18	.29
History	What is the former name of Istanbul, which was used when Christians occupied the XXX?	City	Country	0.96	.89	.79	.90
History	What soft, fluffy, white plant was a major cash crop in the Pre-civil war XXX?	South	North	0.99	.99	.19	.16
History	The British XXX imposed fees on the colonies' sugar, tea, and stamps that were called what?	Parliament	Congress	0.96	.83	.52	.59
History	Which car, produced by XXX Ford's Ford Motor Company, is generally regarded as the first affordable automobile?	Henry	Gerald	0.99	.92	.43	.63
History	During World War II, the US used which weapon against Japanese cities Hiroshima and XXX?	Nagasaki	Nagano	1.00	.95	.33	.37
History	Which president, paralysed by polio, enacted a series of economic programmes called the XXX Deal?	New	Square	0.99	.87	.06	.24
History	What year did Thomas Jefferson write the XXX?	Declaration of Independence	Constitution	0.94	.78	.15	.20
History	What name refers to the severe worldwide economic depression that lasted throughout the XXX?	1930s	1920s	0.91	.76	.17	.31
History	The figure, "XXX the Riveter" symbolised which type of people who assumed industry jobs during World War II?	Rosie	Rosa	0.99	.83	.40	.41
History	The town of Salem, XXX held trials to accuse women of being what?	Massachusetts	Connecticut	0.99	.91	.31	.38
History	Uncle XXX Cabin was a novel written about what form of labour in the 1800s?	Tom's	John's	0.99	.93	.29	.34
History	What was the name of the informal network of secret routes by which Harriet XXX led slaves to freedom?	Tubman	Beecher Stowe	0.94	.87	.34	.29
History	Which of XXX epic poems details the account of Odysseus' journey?	Homer's	Virgil's	0.96	.91	.20	.28
History	Who was the female African-American civil rights activist who refused to give up her XXX seat to a white passenger?	Bus	Train	1.00	.97	.22	.42