15 Moses illusion

Heekyeong Park and
Lynne M. Reder

When asked "How many animals of each kind did Moses take on the Ark?" most people respond "two", even though they know that it was Noah, not Moses, who took the animals on the Ark (Erickson & Mattson, 1981). When a term in a sentence or a question is replaced with a semantically similar but incorrect term, people have difficulty in detecting the distortion. This tendency to overlook distortions in statements is known as the *Moses illusion*.

The Moses illusion was first explored as a scientific issue of inquiry by Erickson and Mattson (1981). They found that people frequently failed to notice the distorted term "Moses" when asked to answer the question, despite reading the question aloud before answering it and despite knowing the name of the correct agent in this role. Even when warned about possible distortions, there was still a great tendency not to note the distortions until they were pointed out. This phenomenon is so robust that it does not require time pressure to elicit the illusion.

Studies of the Moses illusion have focused on when the illusion occurs, what factors influence the illusion, and what mechanisms are responsible for this seeming liability in cognitive performance. That is, what are the mechanisms that underlie the failure to notice mismatches between what is actually presented and what you believe is being asked of you? This chapter will review these issues and present various theoretical accounts that have been proposed to explain these failures, summarizing relevant studies that support or disconfirm each particular account. In the course of this review, the chapter will focus on how people parse questions, query memory, and decide whether the requisite information has been found. We also try to answer the question of why this illusion occurs. Understanding human vulnerability to the Moses illusion can shed light on the memory processes involved in question answering and text comprehension, and will help illuminate the nature of human cognitive architecture.

Text box 15.1 A prototypical Moses illusion experiment

The prototypical Moses illusion experiment described here is based on Experiment 1 of Reder and Kusbit (1991). Participants are asked to answer questions, half distorted and half undistorted. If the question is perceived to be distorted, participants are told to respond "can't say". Otherwise, they are to give the answer to the undistorted question. Accuracy of detecting distorted questions and reaction times to answer them serve as dependent measures.

Method

Participants

A sample size between 20 to 50 participants has provided significant results in the past.

Materials

A list of questions that can be used to produce the illusion is shown in Table 15.1. Each question is listed in two forms, one distorted and one undistorted, along with the answer for the undistorted question. The questions are subject to the following constraints: (1) each substituted term or phrase has to be semantically confusable with the original term; (2) each substituted term has to be syntactically the same part of speech as the original term; (3) the distorted question should not be differently interpretable; (4) the base form of the question should be answerable in the absence of the critical term; and (5) the pair of questions should not differ in length.

Design and procedure

Each participant is presented with only one version of a given question. The questions are randomly assigned for each participant to be presented in either the normal or distorted form, with the constraint that half are seen each way. All participants are told that questions will be presented one at a time on a computer screen and instructed to answer as quickly as possible while remaining accurate. Participants are instructed to treat each question literally and not to give an answer to a question that seems distorted. When a question seems distorted, the participant should respond "can't say". Participants are instructed to respond "don't know" if they do not know the answer to a question.

Analysis

Besides accuracy and response time data, it is necessary to determine whether participants are actually good at detecting distortions. There is a possibility that participants could be biased for calling a question distorted no matter whether the question is distorted or not. Then, the accuracy of distortion detection for a distorted question simply reflects the bias for calling a question distorted rather than sensitivity to distortion detection. In a study by Kamas, Reder, and Ayers (1996), nonparametric measures of sensitivity and bias were calculated. Hit rate (A') is the proportion of "can't say" responses to distorted questions and reflects the proportion of correctly detected distorted questions. False alarm rate (*B'd*) is the proportion of "can't say" responses to undistorted questions and reflects a response bias towards identifying questions as distorted.

Table 15.1 The exemplar questions used in the Moses illusion experiment

	Questions	*Answer*
1	How many animals of each kind did Moses take on the Ark? How many animals of each kind did Noah take on the Ark?	two
2	What country was Margaret Thatcher president of? What country was Margaret Thatcher prime minister of?	England
3	What kind of tree did Lincoln chop down? What kind of tree did Washington chop down?	cherry
4	By flying a kite, what did Edison discover? By flying a kite, what did Franklin discover?	electricity
5	What did Goldie-Locks eat at the Three Little Pigs' house? What did Goldie-Locks eat at the Three Bears' house?	porridge
6	Who found the glass slipper left at the ball by Snow White? Who found the glass slipper left at the ball by Cinderella?	prince
7	What is the name of the Mexican dip made with mashed-up artichokes? What is the name of the Mexican dip made with mashed-up avocados?	guacamole
8	What is the name of the shape whose circumference is "pi-r-squared"? What is the name of the shape whose area is "pi-r-squared"?	circle
9	What country is famous for cuckoo clocks, chocolate, stock markets and pocketknives? What country is famous for cuckoo clocks, chocolate, banks and pocketknives?	Switzerland
10	In the biblical story, what was Joshua swallowed by? In the biblical story, what was Jonah swallowed by?	whale

From "Locus of the Moses illusion: Imperfect encoding, retrieval, or match?", L. M. Reder and G. W. Kusbit (1991). *Journal of Memory and Language*, *30*, 403–406. © 1991 Elsevier Science. Adapted with permission of Elsevier Science.

Results

An analysis of the accuracy and response time results typically showed that undistorted questions were answered much more accurately and quickly than were distorted questions. Moreover the analysis of hit and false alarm rates tended to show that the manipulation of encoding and retrieval did not affect detection of distortions.

THE LOCUS OF THE MOSES ILLUSION: EXPLANATIONS

In this section we review evidence supporting or disconfirming several proposed explanations for the Moses illusion. These explanations include: (1) the cooperative principle – the listener notices the distortion, but believes that the speaker intended the correct term and so ignores the distortion; (2) imperfect encoding – people simply did not read or hear the distorted term in the sentence; (3) imperfect memory retrieval – the question is correctly heard but the information retrieved from memory is incomplete; and (4) imperfect matching of the question terms to memory.

Cooperation hypothesis/conversational postulate

In everyday situations, people often misspeak and it might be considered rude to jump on one's conversation partner and quickly point out every flaw in his or her utterance. From this observation, it might seem reasonable to view the "failure to detect distortions" in the Moses illusion as merely an extension of the everyday behaviour of cooperating with the speaker. In terms of the "conversational postulate" (Grice, 1975), people notice the distortion but choose not to comment on it because they believe that they know what the speaker intended to say. Being cooperative, they know what was meant by the question and therefore respond in a way that reflects the shared knowledge.

Although this explanation seems plausible, it implies that people are explicitly "overlooking" or ignoring a distortion, which means that the task would be easier if the listener/question-answerer did not feel obliged to inhibit correcting or noting the distortion. If so, people should find it easier to detect distortions than to ignore them, and people should find it easy to report a distortion when requested to do so. However, experimental research suggests otherwise: People still exhibit the Moses illusion when explicitly instructed to watch out for any distortion in a sentence (Reder & Kusbit, 1991, Exp. 1). Contrary to the conversational postulate, participants found it harder to detect distortions (responded more slowly and made more errors) when asked to detect distortions than when told to ignore any distortion and just answer the *gist* of the question (as dictated by the conversational postulate). Response times were significantly faster in the so-called gist condition than in the literal condition, in which the participants were asked to monitor for and report any distortion in a question.

Other studies have also called the cooperative principle into question (e.g., Bredart & Modolo, 1988; van Oostendorp & de Mul, 1990). In those experiments, participants were not asked to answer the questions but rather to verify the validity of statements. Clearly, in that situation, politely ignoring distortions would not be appropriate. Nonetheless, the illusion was still found.

Imperfect encoding hypothesis

When information is not encoded, it is not processed, and it cannot be used to make decisions. The second explanation of the illusion assumes that people might not carefully listen to or read the distorted element in a question. It is possible that people already know what the questioner is going to ask once they hear a part of a question. The question would then be understood without encoding distorted information presented later in the sentence. In that case, the Ark question might be processed as "How many animals of each kind were taken on the Ark?" Otherwise, perhaps, encoding might be so expectation-driven that people expect to read or hear "Noah" when they begin to process a question that begins with "How many animals of each kind . . ." In other words, is the failure to notice the mismatch due to imperfect encoding?

If the Moses illusion were due to encoding failure of a distorted word, then a manipulation to ensure encoding of distorted information should eliminate it. In order to investigate this possibility, Erickson and Mattson (1981) required participants to read the sentence out loud before answering the question. Despite this requirement, the illusion still occurred. Conceivably this requirement invoked an automatic reading to speech response, but participants were apparently not *really* processing what they were reading. Perhaps the weak encoding hypothesis could be salvaged if participants are shown to process the critical, distorted word less well when they failed to notice the distortion. In a study of Reder and Kusbit (1991, Exp. 4), word-by-word reading times were collected while participants read and answered questions. The results are displayed in Table 15.2. Reading times for distorted words were faster when participants noticed the distortion than when the distortion was ignored. This result is consistent with the result of van Oostendorp and de Mul (1990), in which failures to detect distortions were found to be slower than detections. If reading time for a critical word is an indication of the amount of time spent encoding that word, the imperfect encoding hypothesis would suggest that participants should have read a distorted word faster when the distortion was not noticed than when the distortion was noticed. The results demonstrate that the illusion is not based

Table 15.2 Mean target reading times and proportion of correct and incorrect responses (in parentheses)

	Literal task		*Gist task*	
	Correct	*Errors*	*Correct*	*Errors*
Normal	525ms (0.79)	515ms (0.21)	429ms (0.82)	618ms (0.18)
Distorted	539ms (0.57)	633ms (0.43)	441ms (0.76)	771ms (0.24)

From "Locus of the Moses Illusion: Imperfect encoding, retrieval, or match?" L. M. Reder and G. W. Kusbit (1991). *Journal of Memory and Language*, *30*, 397 © 1991 Elsevier Science Adapted with permission of Elsevier Science on either encoding failure of the critical word or hasty responding to the question.

Inadequate retrieval hypothesis

Although people correctly encode the distorted question, they might not retrieve the required information to detect the distortion. The information retrieved from memory might be incomplete, sometimes omitting the distorted term, thereby explaining the failure to detect the distortion. For example, people might fall for the Moses question because the retrieved proposition about the number of animals on the Ark would not include the critical information about who took the animals on the Ark. If the illusion were due to imperfect memory retrievals, then one would expect that manipulations improving access to memory should improve detection of distortions as well; however, study results did not support such a notion.

Neither studying nor memorizing the correct version of queried facts before attempting to answer the questions facilitated detection of distortions (Reder & Kusbit, 1991, Exp. 2). Participants studied a series of relevant facts before the questioning, such as "Noah took two animals of each kind on the Ark." Later, after studying half of the facts to be queried, the participants were given half of the questions in their distorted form and half in the undistorted form, making four conditions (studied-distorted, studied-undistorted, not-studied-distorted, not-studied-undistorted) that were crossed with answer instruction types (gist vs literal). Strengthening the memory trace of the correct information should have affected the probability of detecting the distorted term if the illusion were simply based on weak knowledge of the critical term. However, such priming to make relevant information more accessible did not make it easier to detect a mismatch between the underlying information and a distorted question. Participants were just as vulnerable to the illusion after studying the correct information, although the number of wrong answers (e.g., "three" for the Moses question) and "don't know" responses was reduced.

If access to the correct information had affected one's ability to detect mismatches, the gist condition would be expected to suffer relative to the literal condition, because the primed knowledge should make distortions easier to notice and harder to ignore; however, after studying the queried facts prior to answering the questions, participants were much faster and more accurate for studied statements in both conditions. The results reflected an increase in accessibility of relevant knowledge, yet the basic pattern of results was still the same as in other studies: Performance in the literal condition was still slower and less accurate than performance in the gist condition.

Memorizing facts reliably facilitated question answering for undistorted questions both in terms of speed and accuracy of responding; however, a concomitant facilitation was not shown for the distorted questions. This means that the manipulation had an effect, but not the hypothesized effect of reducing susceptibility to the illusion. In sum, given that strengthening the memory trace by familiarizing the relevant knowledge did not reduce the tendency to fall for the illusion, we can reject the imperfect memory retrieval hypothesis as a plausible explanation for the underlying mechanism of the illusion.

Partial match hypothesis

The final explanation we consider suggests that the illusion results from an incomplete or partial match between the probe and the memory structures. That is, as cognitive processors, people make incomplete matches of a complete representation of the question (or memory probe) and a complete representation of the stored proposition that contains the answer. As a question is read, the terms are matched to memory so that the answer may be retrieved. Not every word in the question will be matched exactly to a corresponding memory structure. When the input does not exactly match the memory representation, a term will nonetheless be accepted if it passes a criterion of sufficient match, enabling comprehension. If the degree of match does not reach this criterion, the input query will be regarded as incorrect or incomprehensible. What is the basis for this criterion for comprehension? How much is sufficient to pass the criterion?

With a "game show" paradigm, Reder and her colleagues (Reder & Ritter, 1992; Reder & Schunn, 1996; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997) established that people can erroneously believe that they know the answer to a math problem if it shares features with a problem that they already know. In these experiments, participants were given a math problem, similar to one that they had studied and answered many times in the past, but with the replacement of an operator between the two operands. Thus the correct answer to the altered problem was not available, but the partial match of aspects of the problem led participants to think that it was. This partial match process was modelled within a framework called SAC, for Source of Activation Confusion (Reder & Schunn, 1996; Schunn et al., 1997). This activation-based model of partial matching may represent a prototype for the kind of process involved in question-answering situations that produce the Moses illusion.

It is not unreasonable to think that partial matching might be a general aspect of cognition and as such be a viable explanation for the illusion. Consider, for example, face recognition in a real-world setting. Chances are that a face experienced earlier may be viewed in a different location, with a different expression, and with a different clothes or hairstyle. Despite all of these changes, more likely than not the face will still be recognized.

On the other hand, it would be unreasonable to assume that partial matching would always prevent detection of distortions. For example, it seems likely that people would notice the distortion in a question such as

"How many animals of each kind did Nixon take on the Ark?" Although *Nixon* has the same number of syllables and the same initial phoneme as *Noah*, participants readily noticed the misinformation (Erickson & Mattson, 1981; van Oostendorp & Kok, 1990). While "Nixon" was easy for participants to detect as out of place in that question, not all other names would have the same effect. Then what does affect the detectability of a mismatch? We will consider two proposals, one that the partial match is based on semantic features, and the other that it is based on phonological features.

The case for semantic feature overlap

In the study of Erickson and Mattson (1981), the difficulty of noting a distortion was aggravated when the replaced term in the probe was semantically related to the original term. Erickson and Mattson suggested that a crucial component of the illusion might be the semantic similarity between distorted term and the original term. When the semantic similarity between two terms is high, the replaced term does not seem to flag a mismatch. Conversely, when the semantic similarity between distorted and undistorted terms is low, people more often notice that something is wrong and go on to analyze the critical term in more detail. This, of course, begs the question, in the sense that the original term cannot be in conscious awareness as part of this comparison; if it were, the detection would be trivial. Presumably, this semantic similarity cannot be computed at a lexical level for the two terms.

The view that distortion detection involves a two-pass process – the first to flag a potential mismatch and the second to invoke a careful inspection that might confirm an erroneous term in the question – has support from other types of cognitive tasks. For example, Reder (1987) proposed a two-stage model of question answering that involved strategy selection based on semantic similarity. The first stage consisted of an automatic or implicit evaluation stage in which queries are rapidly assessed for answerability. This initial assessment affects the second stage in which people choose either to search memory in order to answer a question or to base their answer on a plausibility strategy. In the first stage, familiarity and relatedness of terms in the question are evaluated. This assessment is based on semantic relatedness and lexical priming. When the terms seem very familiar (as if the words have been heard recently), people tend to answer the question based on a direct search for the answer; if the terms themselves do not seem as if they were just mentioned, then the question is assessed for general semantic relatedness or familiarity, to decide whether to answer the question in some other way or to decide that it is not answerable.

These ideas may be extendable to question answering for a Moses illusion question. If semantic relatedness among terms in the probe is low, this may suggest the need for further processing before attempting to answer the question. That is, when semantic relatedness between the terms is high, as in the case of the Moses illusion, further processing is less likely to be invoked.

We propose that when a substituted term shares low semantic similarity with an original term, the substitution will be easily detected and thereby cause the checking mechanism to confirm the mismatch. Conversely, substituted terms that bear high semantic similarity to the terms they replace would likely go unnoticed, enabling the adoption of a direct retrieval strategy for finding the answer. Given that *Moses* and *Noah* share many semantic features (e.g., both are central figures in a well-known Biblical story, both stories involved water, they were both old for most of the story, etc.), the substitution of one for the other would likely be undetected. Supporting this suggestion, participants in the study of van Oostendorp and de Mul (1990) frequently failed to notice a distortion when the distorted term was highly semantically related to the original term. Moreover, participants took more time to accurately reject a query in the high semantic similarity condition than in the low semantic similarity condition. This result provides further support that people tend to make more errors when semantic overlap is high and that accurate monitoring of highly semantically related lures is a difficult task.

Semantic cohesion of the critical term with the embedding context or proposition also affects the occurrence of the illusion (van Oostendorp & Kok, 1990). When the distorted terms are totally unrelated to the script that is queried, the discrepancy is readily noticed. On the other hand, when the replaced term is related to the remainder of the proposition or the general context of the query, noticing the distortions is quite difficult. In other words, the more consistent the critical terms in the question are with the script or knowledge structure associated with taking animals on the Ark, the harder it is to notice that the wrong term is used. Moses, a biblical figure, is loosely related to the Ark script, whereas Nixon, a modern politician, is not (Erickson & Mattson, 1981). This is another reason why Moses is frequently accepted in the Ark question, whereas Nixon does not produce (illusory) answers to the question.

Hannon and Daneman (2001) also showed that semantic relatedness of both the words and surrounding context were necessary to elicit semantic illusions such as the Moses illusion. As we mentioned earlier, semantic relatedness might be a function of the number of associations that are shared between two terms. It is more difficult to detect distortions when more terms are related to the theme of the question, suggesting that activation relevant to an answer influences processing of distortions (Reder & Kusbit, 1991). We will discuss this issue in detail in the following section.

Is partial matching based on phonological features?

There is an interesting phenomenon called the "Armstrong illusion" (Shafto & MacKay, 2000) which makes a strong case that a partial matching strategy cannot be based solely on semantic similarity. The Armstrong illusion refers to people's inability to detect the distortion in the question "What was the famous line uttered by Louis Armstrong when he first set foot on the moon?" As in the case of the Moses illusion, people tend to take the question as comprehensible and give an answer to the question, despite knowing that Louis Armstrong was a jazz musician, and that the correct name of the astronaut who visited the moon was Neil Armstrong. Shafto and MacKay argued that the underlying mechanism for the illusion is phonetic similarity between Louis Armstrong and Neil Armstrong, and that phonological input of *Armstrong* and semantic input from the remainder of the question lead to people to overlook the distortion. They went on to argue that the Moses illusion could be explained in the same manner. Although *Moses* is presented in the question, the name *Moses* receives only one source of bottom-up priming from the physical presentation of the name; however the correct, but non-presented name, *Noah* is assumed to receive priming from two sources. The term *Noah* is primed by the terms "Ark" and "animals of each kind" in the question because *Noah* is already pre-associated with those concepts in the Ark script. *Noah* is already primed from the name *Moses* because the two names share many semantic similarities and are strongly associated. That is, although the name actually presented was *Moses*, the name *Noah* receives more priming because of pre-associations in semantic memory and semantic linking between two terms. Since the name *Noah* receives more priming by two convergent priming processes, people often fail to notice that an important term has been replaced. In this framework, the Moses illusion is considered to be the result of miscomprehension of *Moses* as *Noah*.

Further, Shafto and MacKay (2000) proposed that phonetic similarity of a substituted term with an original term primes the original term, thereby facilitating the illusion. While the semantic similarity aspect of their explanation is consistent with experimental results, the phonetic similarity aspect is not. As was mentioned earlier, the name *Nixon*, which is closer phonetically to *Noah* than *Moses*, does not elicit the illusion. Given that phonetic similarity does not always produce the illusion, we need to consider other accounts of the Armstrong illusion. The name "Armstrong" is frequently cited when the topic of the first astronaut who landed on the moon is mentioned. Although the last name is not always needed to identify an individual (e.g., "Elvis") and sometimes the last name is not sufficient to identify an individual (e.g., "Taylor"), the last name is frequently used in a non-familiar context for purposes of identification. Perhaps "Armstrong" boosts the activation of *Neil Armstrong* and semantic cohesiveness of the name, and the remainder of the sentence leads to the illusion due to high-relatedness of the name "Armstrong" and the moon-landing script. The Armstrong illusion by itself could be accommodated by either the phonetic or the semantic partial matching story; however, considering the other data on illusions, it seems that semantic overlap is still the most important factor contributing to the occurrence of the illusion.

Partial match and spreading activation

Let us consider in more detail how much semantic similarity or semantic feature overlap is required between the distorted and original term in order to produce the illusion. One possible mechanism would involve bringing the entire memory trace or schema related to the probed information into working memory. In such a situation, not every term of the memory trace would be carefully matched to the test question before "reading off" the answer. It seems reasonable that this partial matching process could be our default process for memory matching. In most situations the form of a question is not likely to match closely with the memory representation it queries. Slight mismatches would be expected even when the input is a statement rather than a question. Indeed, everything we see is varied from different perspectives, so we need to perform partial matches to recognize virtually anything. Consequently, people are accustomed to being tolerant of discrepancies, and highly similar terms are allowed to slip by or are folded into existing representations. In our view, the normal mode of processing strives to be as effortless as possible, and that includes comprehension.

Partial match is sufficient to retrieve information from memory and is itself an important matching process involved in memory retrieval. The amount of overlap between the working memory representation and the long-term memory structure affects the likelihood of accepting the partial match as sufficient. The degree of acceptable overlap is primarily a function of the amount of activation arriving at the higher-level structure that is being matched (see Reder & Schunn, 1996; Schunn et al., 1997, for more details).

Kamas and Reder (1995) suggested that the Moses illusion might be explained by positing the spreading of activation among related concepts in semantic memory. When a person is asked a question, processes operate on this semantic network in search of the queried element. The more activation that accrues at a concept through its connections to the remainder of the concepts in the question, the more likely the person is to accept the retrieved concept as being acceptable and not a distortion. When a term in the question does not match the stored representation, the probability of detecting this mismatch is a function of the number and strength of connections from the distorted word to the schematic node that is queried. The more connections between the schema and the distorted term, and the stronger those connections, the more likely the distorted term will go undetected. Since Noah and Moses share many features, there would be a large number of connections between Noah and Moses leading the substitution of *Moses* for *Noah* to go unnoticed. In contrast, *Nixon* has no obvious semantic connections with the Noah schema, thereby making the mismatch easier to detect. Further, activation is divided among all the concepts in the probe and is assumed to be finite (e.g., Anderson, Reder, & Lebiere, 1996). Thus a mismatching word with no connections to the remaining concepts in the question (e.g., *Nixon*) takes away activation that could be spread to the relevant script, further facilitating distortion detection.

Hannon and Daneman (2001) proposed that knowledge access regulated processing of the related terms while working memory span regulated processing of the context of the terms involved. In their regression analysis, combined factors of knowledge access for critical terms and working memory span accounted for substantial amount of variance for occurrence of semantic illusion. The portion accounted for by knowledge access was greater than that accounted for by working memory span, further supporting the claim that the illusion is due to semantic similarity.

What affects failure to detect a distortion? Whether or not one assumes that partial matching is the mechanism that causes the Moses illusion, one can still ask what factors modulate the likelihood of a distortion being detected.

Limited cognitive capacity?

Conceivably, our cognitive capacity is sufficiently limited that it is difficult to monitor for distortions. Under the assumption that the process searching for the answer competes with the process searching for distortions in the question, there might not be enough cognitive resources to adequately monitor for distortions. If so, reducing cognitive load by removing the requirement of answering the question, and instead only requiring that distortions be found, might be easier. When participants were required only to monitor for distortions and not to answer the questions, more distorted questions were detected, but there was also an increase in false alarms to undistorted questions (Kamas et al., 1996, Exp. 2). This result suggests that reducing the cognitive demands only affected the response bias, but not participants' ability to detect true distortions.

Failure to focus attention on the relevant terms?

Although the inability to detect distortions seems not to be due to insufficient cognitive capacity, perhaps this cognitive error arises from insufficient attention to the critical terms. Bredart and Modolo (1988) examined whether focus of the sentences affected the Moses illusion, using cleft sentences such as "It was Moses who took two animals of each kind on the Ark" and "It was two animals of each kind that Moses took on the Ark". If the focus of the sentences were to have an effect on the illusion, the illusion rate would have been greater for the statements directing focus on to something other than the distorted information. Participants often noticed the discrepancy when the inconsistent part of the statement was in focus, whereas they were not good at detecting a discrepancy when the consistent part was in focus. With this result, Bredart and Modolo argued that only the terms in the focus of attention were compared to the memory structure, and that attention focus was the factor contributing to the illusion. It is important to note, however, that Bredart and Modolo did not include correct sentences such as "It was Noah who took two animals of each kind on the Ark". Conceivably, in their study too, the manipulation may have affected response bias rather than true sensitivity to the distortions.

Kamas et al. (1996, Exp. 1) tested the same idea as Bredart and Modolo but did include non-distorted versions of each question in order to estimate response bias from the false alarms. They manipulated focus by having participants study the relevant fact before answering the questions and varied which part of the statement was emphasized. Three different types of focus were used, manipulated by the terms that were capitalized in the study sentence: (1) The answer was capitalized (e.g., "Noah took TWO animals of each kind on the ark"); (2) the critical term was capitalized (e.g., "NOAH took two animals of each kind on the ark"); or (3) neither term was capitalized. Participants were significantly *less likely* to notice the distortion if the *answer* had been capitalized during study. On the other hand, emphasizing (capitalizing) the critical word in the study sentence made participants more prone to detect the distortion.

Although these results seem to suggest that the ease of detecting the distortion depends not only on the semantic similarity but also on the amount of attention, the same problem noted earlier occurred here as well. The capitalization manipulation not only increased detection of distortions, but it also increased the error rate for undistorted questions, suggesting that the effect of focus only affects response bias, not sensitivity. A signal-detection analysis confirmed that capitalization of the critical term only affected bias. Moreover the failure of detection of distortions was high even when the distorted word was capitalized (Figure 15.1). Thus, it seems that the cause of the illusion cannot be attributed to insufficient allocation of attention.

In sum, people are not good at adopting an explicit word-by-word checking procedure, and they cannot easily become more vigilant at detecting distortions even when they try very hard. We are left with the conclusion that it is not easy to change the basic nature of the partial match process. The next question to ask is, at what level does the partial match occur?

Word vs feature level for partial matching?

To investigate at which level partial matching occurs, Kamas et al. (1996, Exp. 4) made various types of features of the distorted term salient in a question preceding the critical question phase. Participants were presented with questions that (1) emphasized features shared between the original and distorted terms (e.g., "What religions study the story of Moses?"); (2) emphasized features that distinguished the distorted term from the term it replaced (e.g., "What sea did Moses part?"); or (3) contained irrelevant features to the illusion questions (e.g., "What is the name of the once-outlawed Polish labour union?"). After presenting the primed question, participants

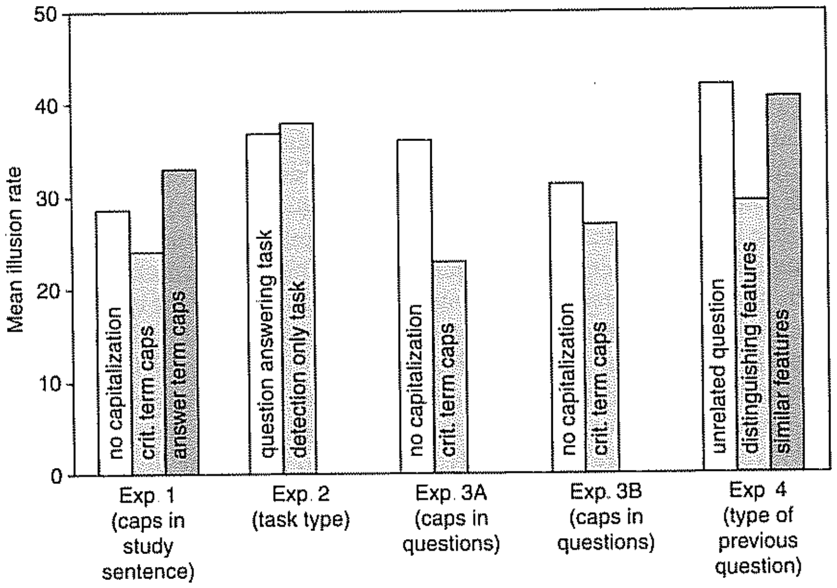

Figure 15.1 Mean illusion rates from four experiments of Kamas et al. (1996) from "Partial matching in the Moses illusion: Response bias not sensitivity" E. N. Kamas, L. M. Reder, and M. S. Ayers (1996) *Memory & Cognition, 24*, 696. © 1996. Psychonomic Society Inc. Reproduced with permission of Psychonomic Society Inc.

were required to answer the critical illusion question. Unlike previously discussed manipulations, detection rates did improve when a preceding question emphasized features that distinguished the original term from its replacement, suggesting that the partial match process operates at the feature level rather than at the word level. It was the focus on the distinguishing features between the original and distorted term that actually improved detection rates as opposed to only affecting response bias.

This result begs the question of why other manipulations at word level did not produce a comparable improvement in detection. It would seem that a manipulation at word level makes more salient both the similar and dissimilar features of the correct and distorted terms, due to their semantic connections in memory. Since the priming at word level does not alter the relative distribution of activation from the word node to its constituent features, the proportion of activation sent from the similar features has not been changed. This explains why the manipulations at the word level did not affect the rate of distortion detection.

In the Moses illusion paradigm, the distorted term shares semantic features with the undistorted term, suggesting that the word is consistent with the basic conceptual representation of the queried information, or at least the features that overlap with it. When the distorted term is activated within the conceptual representation of the question, such as *Moses* in the Ark question, the distorted term has more chance of being accepted in place of the correct term. Then it can also be predicted that the mere presence of the distorted term will affect the schematic representations for a short period of time. Participants in Kamas et al. (1996) were given a *post-test* to ensure that any failures to detect a distortion were not due to a lack of knowledge of the correct information (e.g., Noah took the animals on the ark). Participants were less likely to give the correct answer on the post-test for those questions that had been distorted during the experiment. Tendency to give the correct answer was not improved by getting the undistorted version as compared with a neutral version (e.g., "How many animals of each kind were taken on the Ark?"). One explanation of this result is that the features of the distorted term were already connected to the schema and that these links were strengthened by the previous experimental presentation (see Reder & Schunn, 1996; Reder, Nhouyvanisvong, Schunn, Ayers, Angstadt, & Hiraki, 2000, for more details). Also it has been shown that participants tend to give a wrong answer to a question that has been primed by a semantically similar one (Kelley & Lindsay, 1993). Potter and Lombardi (1990) also demonstrated that priming of a synonym can cause people to intrude the wrong word in a "verbatim" recall of a recently presented sentence. Of course, we expect all these effects to be short-lived; that is, that the probability of giving the distorted term as a response will decrease with time, as activation decays.

CONCLUSION

Research on the Moses illusion demonstrates that people have difficulty in detecting distortions or inaccuracies when a distorted element is semantically related to the theme of the sentence. Why should our cognitive system be so tolerant of distortions and find it so difficult to do careful matches to memory? It might seem that partial matching is a less-than-ideal way to process information; however, the partial match process is not only common and normal but also a necessary mechanism of our cognitive system. This partial match process enables useful communication and comprehension. Very few things that we see or hear will perfectly match the representation that we already have stored in memory. In order to answer questions, we need to be able to use an acceptable match. In order to understand a new situation and map it onto something we have already seen or done, we must accept slight variations. Every day at many levels, we accept slight distortions without even noticing the process. Occasionally we notice a distortion and choose to ignore it, but more frequently, we do not even realize that distortions have occurred. A rigid comprehension system would have a difficult time indeed. Many of our cognitive operations are driven by familiarity-based heuristics rather than careful matching operations. The Moses illusion is an example of how the adaptive, human cognitive system works. Everyday cognitive processing must be based on simple heuristics such as matching sets of features rather than exact matches, as very few tasks require exact matches. Sentences do not match stored information, faces change, voices may change slightly, even our pets and friends change over time. Therefore it makes sense that people do use partial matches in the normal course of matching to memory. Partial matching is immutable because it is the most efficient way for memory to operate, given the nature of the environment in which we live.

SUMMARY

- When a term in a sentence or a question is replaced with a similar but incorrect term, people have difficulty in detecting the distortion. This is called the Moses illusion.
- The illusion results from a partial match process between the memory probe and the memory representation structures.
- The Moses illusion is an example of how human cognition works in an adaptive and efficient way.

FURTHER READING

Erickson and Mattson (1981) was the first paper to explore the Moses illusion. A comprehensive theoretical account of the illusion and empirical support for the explanation are provided in Reder and Kusbit (1991), and Kamas et al. (1996).

ACKNOWLEDGEMENTS

Preparation of this chapter was supported by Grant 2-R01-MH52808 from the National Institute of Mental Health. We thank Kaia Vilberg for comments on an earlier draft of this chapter.

REFERENCES

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thoughts*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221–256.

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

Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*, 540–552.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York: Seminar Press [Originally published from William James Lectures, Harvard University.]

Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-difference perspective. *Memory & Cognition*, *29*, 449–460.

Kamas, E. N., & Reder, L. M. (1995). The role of familiarity in cognitive processing. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 177–202). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*, 633–654.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*, 90–138.

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.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 294–320.

Reder, L. M., & Ritter, F. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–451.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–77). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not calculate: A source activation confusion (SAC) model of problem-familarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1–27.

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

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

van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. *Languages and Cognitive Processes*, *5*, 105–113.