Koriat, A. (1995). Our knowledge of our own knowledge: Monitoring and control processes in memory . In K. Pawlik (Ed.), Bericht uber den 39. Kongress der Deutschen Gesselschaft für Psychologie in Hamburg 1994 (pp. 95-113). Göttingen: Hogrefe.

OUR KNOWLEDGE OF OUR OWN KNOWLEDGE: MONITORING AND CONTROL PROCESSES IN MEMORY

Asher Koriat Department of Psychology, University of Haifa, Israel

We are constantly engaged in monitoring our knowledge. Even when we fail to recall a target from memory, we may have a "feeling of knowing" about it, and even feel that it is "on the tip of the tongue". The present paper focuses on the question of how we succeed in monitoring our knowledge. According to the commonly held trace-access view, we do so by using a special monitoring mechanism that can detect the presence of the target in memory before attempting to retrieve it. The alternative, accessibility account advocated in this paper postulates that people have no direct knowledge of their own knowledge. Rather, the accuracy of metaknowledge in mirroring knowledge is a by product of the accuracy of knowledge itself. Thus, the feeling of knowing about an unrecallable target is based on the overall accessibility of partial information about the target. Because this information is generally accurate, the feeling of knowing is also accurate. However, when memory is inaccurate, metamemory will also be inaccurate. Results supporting this view are reviewed.

The subjective monitoring of knowledge

The study of subjective experience

Experimental psychology began its life in Germany as the science of consciousness. Its goal was to analyze the content of subjective experience as it reveals itself to introspection. The advent of behaviorism practically banned the study of subjective experience, focusing only on the experimental investigation of overt behavior. The last two decades, however, have witnessed a renewed interest in the subject of conscious experience and its possible role in information processing and behavior. This interest extends over several divergent areas of research, such as social psychology (e.g. Lewicki, 1986; Nisbett & Wilson, 1977), cognitive psychology (e.g., Bans, 1988; Marcel & Bisiach, 1988), decision making (e.g., Montgomery & Svenson, 1989), and neuropsychology (e.g., Shallice, 1988). Various aspects of conscious experience have gained a reputable status in both research and theorizing.

Despite the large body of evidence that has accumulated during the past few years regarding the correlates of consciousness, some of the fundamental issues remain open. For example, recent work has clearly demonstrated the wide prevalence of unconscious and implicit modes of information processing. On the one hand, this work suggests that consciousness is not very critical to the processing of information. On the other hand, the contrast between implicit and explicit processes suggests that consciousness does make a significant contribution to the quality of cognitive operations. Thus, the old issue of the status of consciousness in human behavior is still open.

Here I would like to focus on the subjective monitoring of knowledge, that is, on our knowledge of our own knowledge. Although this topic covers only a very small aspect of subjective experience, for some theoreticians it touches upon a defining property of consciousness. For those theoreticians, consciousness implies not only that we know something, but also, that we know that we know it. Thus, consciousness binds together knowledge and metaknowledge. Indeed, the study of metacognition can help illustrate some of the very basic issues about

consciousness, as well as demonstrate the possible interactions between consciousness and behavior.

Varieties of memory monitoring

Let me begin by illustrating some of the monitoring and control processes that occur during the course of learning and remembering: When we read a text, we make sure that we comprehend what we read, otherwise we try to read it again. Studies concerned with the calibration of comprehension indicate that people are very bad at assessing their comprehension. They tend very often to have an illusion of comprehension. Thus, a student preparing for an exam may sense that he has mastered the material and can go to the movies, and then be surprised when he gets a low grade.

In everyday life we constantly monitor the future recallability of newly acquired information, and allocate learning and rehearsal resources to it accordingly. When we have a scheduled appointment, sometimes we make a great effort not to forget it. At other times we may judge that there is no need to write down the appointment in our diary, because we will remember it anyway, and consequently end up missing that appointment. Even when we do remember to perform a planned action, such as calling the doctor, we have to remember that we already did it in order not to repeat the action once again. Older people are known for their inability to keep track of what they have done. They take a medicine more often than prescribed, having forgotten that they already took it, or tell the same joke over and over. We examined what happens when we tell a joke to an older person, and then try to tell him the same joke again. The older person would invariably stop us, complaining that we already told the joke. It would seem from our studies (Koriat, Ben-Zur, & Shelter, 1988) that older people have difficulty monitoring their own actions, but are not very deficient in monitoring the actions of others. Older people are not the only sufferers; most of us face a problem in monitoring our actions. A few years ago I sent a preprint of a paper entitled "Telling the same story twice" to a friend , and after a week or so I received a note from him: "Thanks, but you have already sent me a preprint of that paper several days ago". Too often, after performing a routine action, we wonder whether we indeed did it. Have we locked the door? Have we turned off the oven? In a survey study that we conducted (Koriat & Ben-Zur, 1988), subjects reported that in 99% of the cases in which they returned to check whether they had locked the door, they found the door locked. So, we do not trust our memories that much when it comes to memory for routine actions that we perform. Sometimes excessive checking behavior reflects a failure of control: A compulsions repeat an act. But more often it simply derives from imperfect monitoring: uncertainty about whether we have already performed the act, or perhaps just thought about performing it.

Although we are particularly aware of deficiencies in our monitoring, it should be stressed that by and large we are quite successful in monitoring the correctness of our memories. When we are presented with the question "What is the phone number of Woody Allen's parents?" We can immediately reply "I have no idea" without bothering to search for the number in our memory. In other cases, even when we fail to remember some piece of information, for example, the name of somebody, we may experience a "feeling of knowing": a feeling that we do know the name, and are likely to recall it later. Sometimes we may feel that the name is "on the tip of the tongue", and can even monitor its emergence into consciousness. When the name does emerge into consciousness, we can generally tell that it is indeed the one we have been searching for.

These are only a few examples of the monitoring processes that occur in the course of learning, remembering and acting. These processes raise the following questions: How do we

know that we know? Does subjective knowledge accord with actual knowledge and why? What are the consequences of acting on one's own subjective judgments of knowledge?

The challenge of memory monitoring

Over two decades ago, Tulving and Madigan (1970) concluded their review of the verbal learning literature with the following statement (p. 477):

> "What is the solution to the problem of lack of genuine progress in understanding memory? It is not for us to say, because we do not know. But one possibility does suggest itself: why not start looking for ways of experimentally studying, and incorporating into theories and models of memory, one of the truly unique characteristics of human memory: its knowledge of its own knowledge. No extant conceptualization, be it based on stimulus-response associations or an information processing paradigm, makes provisions for the fact that the human memory system cannot only produce a learned response to an appropriate stimulus or retrieve *a* stored image, hut it can also rather accurately estimate the likelihood *of* its success in doing it.
>
> Hart has reported some experimental data in support of the reality of what lie calls the memory-monitoring process and has thus pointed the way to the study of the most important and the least understood aspect of human memory. We cannot help but feel--- that if there is ever going to be a genuine breakthrough in the psychological study *of* memory, one [hat would save the students of ecphoric processes from the fate we talked about at the beginning of this chapter, it will. among other things, relate the knowledge stored in an individual's memory to his knowledge of that knowledge."

Although a great deal of work has been carried out in the last two decades on the feeling of knowing, only a small part of it has direct bearing on the theoretical question raised by Tulving and Madigan, and that is the question that I would like to focus on. This question becomes particularly acute when we consider memory blockage states, such as the tip-of-the tongue state. This state is of particular interest because it combines two conflicting features, the *subjective* conviction that I know the answer, and the actual, *objective* failure to retrieve it. The question that naturally arises is how do I know that I know the answer in spite of being unable to produce it?

Much of the research to date has focused on a somewhat different question, that of demonstrating the *accuracy* of feeling of knowing (FOK) judgments, that is, the validity of subjective indexes of knowing in mirroring actual knowing. Indeed, these studies have established that when subjects are unable to retrieve a target from memory, their FOK judgments about the target predict their success in recalling the target in the future, or identifying it in a recognition test.

The standard paradigm for assessing FOK accuracy was introduced by Hart (1965). The subject is first presented with a question requiring a one-word answer, and asked to recall that answer. For example, "What is the capital of Finland?", "What is the name of Tarzan's girl friend?" "What was the nickname of Louis Armstrong?" If the subject succeeds in retrieving the correct answer, the trial is terminated, and the next question is presented. However if he fails to find the answer, he is asked to judge whether or not he "knows" the answer to the extent of being able to recognize it among distractors. Finally, the subject is tested on that question using a forced-choice recognition memory test The results generally indicate that FOK judgments following recall failure are predictive of recognition success. The correlations are generally not very high, but they are nevertheless sizeable and significant.

More recently, there has been greater interest in two related questions. First, what is the *basis* of the FOK. Where do people get the sense that they know something even though they are not able to retrieve it? Second, why is the FOK *accurate* in monitoring actual knowledge? This seems to be the puzzle that Tulving and Madigan stress: How does metamemory succeed in mirroring

memory? I shall begin my examination of this question by an illustrative study which concerns the validity of subjective judgments in a somewhat unusual task.

A demonstrative study: The puzzle

That study (Koriat, 1975) dealt with the idea of a universal phonetic symbolism. A simple demonstration of this idea was provided by Koehler (1947): When subjects are asked to match the two nonsense words Takete and Maluma with two shapes, a rounded shape and angular shape, they almost invariably assign the name Maluma to the rounded shape and Takete to the angular shape. What interested me about this phenomenon is that subjects expressed a strong conviction in the correctness of their match. Of course, it is not meaningful to ask whether the subjects' FOK is accurate, because we are dealing here with artificial, nonsense stimuli. So I wondered what would happen if we use real words from remote natural languages for which there is a correct answer?

Indeed, there have been a few studies in which subjects were asked to guess the meaning of words from noncognate foreign languages, such as Chinese, Yoruba, etc. The procedure is simple: A subject is presented with a pair of antonyms from the foreign language, and asked to match them with the respective English antonyms. For example, the Chinese pair *ching* and *chung*, would be presented with the English pair light-heavy, and the question is which is which. Chances of getting the translation correct are 50%. In several studies using different foreign languages subjects were found to do slightly but significantly better than chance. These results suggest that phonetic sounds have intrinsic semantic connotations which are universally shared, and that traces of these sound-meaning associations survive in all natural languages. I found a similar better than chance performance when Hebrew-speaking subjects matched Chinese ideographs with their corresponding Hebrew words, and when subjects were asked to guess whether a letter in a foreign alphabetic string designated the vowel / or the vowel *a*.

In carrying out these studies I was again impressed by the fact that subjects sometimes responded with great confidence, as though the appropriate match felt intuitively self-evident. The question that I asked was whether subjective conviction, or the sense of "self evidence", is diagnostic of the correctness of the match. Indeed, some philosophers assume that universally shared notions have the quality of self evidence; They strike you as being right.

The experiment was simple. I took advantage of items that were compiled by Dan Slobin (1968). These consisted of antonymic pairs from three languages, Thai, Kannada, and Yoruba, and their English translations. The pairs were chosen to represent the three major dimensions of the semantic differential, and included both sensory and abstract dimensions. American students were asked to match each foreign pair with its corresponding English pairs. In addition, for each match, subjects were asked to indicate their FOK on a 4-point scale, with 1 indicating "a totally wild guess", and 4 indicating that the answer is "reasonably likely to be right".

What were the results? Consider first translation accuracy. Mean percentage of correct translation was 58.1, which was very significantly higher than the 50% chance level. A better than chance translation was found for both sensory and abstract dimensions, and for antonyms pertaining to all three dimensions of the semantic differential.

Of more interest to us here are the FOK ratings. The percentage of correct translation appeared to vary systematically with these ratings. When subjects made a rating of "1", that is, "a totally wild guess", their accuracy was 53.6%, which was still better than chance. However, translation success increased steadily with FOK ratings, reaching 66% for a rating of 4. The difference was very highly significant. The correlation between FOK ratings and translation

accuracy was quite pervasive: Correct translations were assigned higher FOK ratings than incorrect translations, and this was generally true across sensory and abstract attributes, and across all three dimensions of the semantic differential.

These results present a puzzle, because it is not clear how subjects succeed in monitoring the correctness of their answers. There are many problems and riddles in which once I have arrived at the solution, I can easily check whether I am right or wrong, but this is not true for the word-matching task. Here there is no simple algorithm by which I can determine whether I am right or wrong. In fact, this is also true for many memory tasks: When I retrieve the name of an acquaintance from my memory, there is no simple procedure by which I can determine that the name is indeed correct. Thus, perhaps, there is some general principle that underlies the accuracy of monitoring in a variety of tasks, including memory tasks and word-matching tasks such as the one I have just described.

How, then, do people monitor the correctness of their guesses about the meaning of words from an unknown language? I am going to leave the puzzle open, and will return to it after examining some results from memory research.

The feeling of knowing

Let us now focus on the typical memory situation in which a FOK is experienced. In this situation the solicited target is apparently available in memory but is not immediately accessible. This is a rather common situation, because only a fraction of the information that is available in store is accessible to us at any moment.

What do we know when we don't know?

Even when we fail to retrieve a particular target from memory, we can sometimes access two types of information about it. First, we can make *a feeling of knowing* judgment, representing our subjective conviction that the target is indeed available in memory. Second, we may be able to provide some *partial information* about the unrecallable target. For example, when we cannot remember the exact meaning of a rare word, we can still make some guesses about its generic meaning (Eysenck, 1979). Also, when we fail to recall the name of a person, we may still be able to tell what it rhymes with.

Interestingly, both types of information tend to be accurate. The feeling of knowing is relatively accurate, and also the partial information retrieved, both phonological and semantic, tends to be correct (A. Brown, 1991; R Brown and McNeill, 1966; Koriat & Lieblich, 1974; Schacter & Worling, 1985). A question that presents itself concerns the relationship between these two types of information, the FOK, and the retrieved pieces of information. As we shall see, this is an important question which can help distinguish between two types of theoretical approaches to the FOK. I would like to illustrate this question with an episode from my own personal experience (cited in Koriat, 1994a).

During one memory convention, 1 was asked for the reference of a book on thinking that I had read many years earlier. I tried in vain to recall the name, but only a few letters came to mind: Wand *N*, and (with somewhat less confidence) *S*. After struggling with the name for a whole day, I was able to get some help from a friend. After communicating to him all the partial information I was able to access, including the letters that came to mind, and some details about he book, he came up with the answer: Wason! When I heard the name, I remembered the

Penguin book edited by Wason (Wason & Johnson-Laird, 1968), but it was immediately clear to me that it was *not* the book I had in mind, and Wason was *not* the name that I was searching for. However, I also understood that the partial information that I was accessing was apparently coming from *Wason*. Now that I found out that I was on the wrong track, I made an effort to put aside the letters that came to mind, and it took me no more than a few minutes to retrieve the correct name: It was McKellar!

This example is consistent with one the accounts of the TOT state, according to which, we fail lo retrieve the solicited target in the TOT state because other related targets, referred to as "blockers" or "interlopers", come to mind and interfere with accessing the correct target (Burke, MacKay, Worthley, & Wade, 1991; Jones, 1989; Reason and Lucas, 1984). "Wason" possibly represents an interloper, and "McKellar" represents the correct target.

The question that emerges is the following: when we fail lo recall a target, what is the *source* of (he partial information and the FOK judgments that we can provide about it? Clearly, the letters that came to mind apparently emanated from "Wason" rather than from "McKellar". Is it possible, then, that the positive FOK that I experienced actually originated from "McKellar"? If so, would this mean that the FOK continues to monitor the availability of the *correct* target in store, even when the partial information accessed steins from an interloper, that is, from an *incorrect* target?

This question can help us distinguish between two very different accounts of the FOK, the trace-access account, and the accessibility account.

The trace-access account of the feeling of knowing

The trace-access model offers a simple answer to the question of how do we know that we know something that we are unable to retrieve. This elegant model was first proposed by Hart (1965, 1967a, 1967b), and has been implicitly endorsed in many discussions since (see e.g., Yaniv & Meyer, 1987). The model postulates the existence of a special *memory-monitoring module* that directly detects the presence of the target in the store. This module is assumed to have a privileged access to memory traces; it can detect the *availability* in the memory store of an otherwise inaccessible target. Thus, whenever a person is required lo recall a target, the monitoring module is activated to make sure that the target is present in memory before attempting to retrieve it.

Hart stressed the functional value of having such a built-in monitor, given the fallibility of the memory system (see also Yaniv & Meyer, 1987). In such a system FOK "can serve as an indicator of what is stored in memory when the retrieval of a memory item is temporarily unsuccessful or interrupted. If the indicator signals that an item is not in storage, then the system will not continue to expend useless effort and time at retrieval; instead, input can be sought that will put the item into storage. Or if the indicator signals that an item is in storage, then the system will avoid redundantly inputting information that is already possessed" (Hart, 1965; p. 214). The assumption, then, is that subjects have *direct access* to the information pertaining lo the presence of the solicited item in memory.

At first sight, this solution lo the question of how one knows that one knows appears lo raise the *homunculus* problem: How does the monitor itself "know"?. However, the idea becomes much less far-fetched when we draw an analogy from the manner in which information is organized in computerized systems. Computer users arc familiar with the concept of a *directory*. A directory contains only the *names* of the files stored on a computer disk; not the content of the files. Therefore, when a computer is asked to retrieve a file from memory, the first step is to consult the *directory* to see whether it contains the *name* of the file. This step is analogous to "monitoring". If the name cannot be located, the computer can respond "file not found", without having to search the contents of the memory store. Only when the name of the file is found in the directory, will an attempt be made to retrieve the *file itself*. This step is analogous to "retrieval" (see Koriat, 1994a). This is a very effective way of organizing information, and perhaps human memory is also similarly organized.

The directory analogy illustrates the two basic assumptions of the trace-access model: First, the assumption of a special monitoring module which allows to directly detect the presence of the target in the memory store. Second, the assumption of a *two-stage* process, in which monitoring always precedes retrieval. Consistent with this assumption is the finding that subjects spend more time searching for a target when the initial FOK is high than when it is low (see Nelson & Narens, 1990).

An elegant feature of the trace-access model is that it also offers a straightforward explanation for the *accuracy* of the FOK. Clearly, if the FOK directly monitors the presence of the trace in memory, then it ought to serve as a valid predictor of actual memory performance. In fact, if this view is endorsed, it should be the inaccuracy rather than the accuracy of the FOK that would need explanation!

Lei us now see how the trace-access model handles the tip-of-the-tongue example that I gave earlier concerning Wason and McKellar. The assumption of two separate mechanisms, one for monitoring and one for retrieval, allows for a dissociation between their outputs. Although the retrieval mechaiysm keeps coming up with information from Wason, the monitoring process continues to detect the presence of the *correct target* ("McKellar"). The assumption, then, is that even when the retrieval process is misled by a variety of intruding clues emanating from wrong targets, the FOK still monitors the presence of the *correct* target in memory.

This assumption, in fact, has guided some of the common experimental practices in the study of the FOK. To an outsider these practices would seem to entail *a confusion between the perspective of the subject and that of the experimenter* (see Koriat, 1993). Consider a typical FOK experiment. The subject is presented with the question "What is the capital of California?" If the subject fails to provide any answer, then the experimenter solicits FOK judgments, and tests for target recognition. However, suppose that the subject gives the correct answer, Sacramento. The experimenter would then typically move on to the next item because there is no sense in soliciting FOK judgments when the subject knows the answer. However, who is making the decision here? It is the *experimenter*. He knows that the subject knows the answer. However, our interest is not to test the experimenter; we want to determine whether the *subject* himself knows that he knows and how. Yet in this procedure a significant subset of the memory instances are eliminated from further analysis merely because the *experimenter* knows that the subject is right.

In contrast, assume that the subject's response is "San Francisco". Now the experimenter will inform (he subject that the answer is wrong, and will proceed to solicit FOK judgments about the prospects of recognizing the correct answer. Again the *experimenter* knows that the subject is wrong, and therefore defines the situation as involving a "retrieval failure", that is, a failure to recall the *correct* target. Thus, even though the subject may insist on San Francisco, the experimenter still asks for FOK judgments. Why? Because such judgments are implicitly assumed to have access to the "real thing", to the trace of the *correct* target. The assumption, then, is that the FOK has privileged access to information that is beyond the reach of retrieval.

The accessibility account of the feeling of knowing

An alternative account that I have proposed is the accessibility account (Koriat, 1993, 1994a, 1994b). According to this account, people have no knowledge of their own memory over and above what they can retrieve from it. They cannot monitor directly the presence of information which they cannot access. However, they can still tale advantage of what they can retrieve to make inferences about what they cannot access. Thus, I here is no separate monitoring module that has privileged access to information that is not already contained in the output of retrieval. Rather, the cues for the FOK reside in the products of the retrieval process itself. Whenever memory is searched for a solicited target, a variety of clues often come to mind (see Blake, 1973; Gardiner, Craik, & Bleasdale, 1973; Lovelace, 1987; Read & Bruce, 1982), including fragments of the target, semantic attributes, episodic information, and a variety of activations emanating from other sources. Although such clues may not be articulate enough to support an analytic, calculated inference, they can still act in concert to produce the subjective feeling that the target is "there." Even when retrieval fails, the scattered debris that are left behind can foster a positive FOK. Importantly, it is assumed that subjects cannot directly tell whether a partial clue is right or wrong. Thus, I could not tell whether the letters S or N that came to my mind during the TOT state were correct or incorrect. Therefore, both correct and incorrect clues contribute to the enhancement of the FOK. What matters for the FOK is only the overall accessibility of information, that is, the number of clues that come to mind and the ease with which they come to mind. Essentially, the FOK represents an attempt to extrapolate from the processes that occur during one retrieval episode to future retrieval episodes: If a question activates many associations, then I have the feeling that it will eventually lead to the recollection of the target, but if it leaves me "blank," I feel that it will continue to bring nothing to mind. This account of FOK resembles the availability heuristic postulated by Tversky and Kahneman (1973) to explain how people estimate proportions or frequencies.

The accessibility account of the FOK contrasts with the trace-access model in two respects. First, whereas the trace-access model assumes a modular organization in which monitoring precedes retrieval, the accessibility account assumes an interactive process that combines retrieval and monitoring: It is by attempting to search for the solicited target that I can know whether I know it or not. Thus, the FOK monitors the accessible information in short term memory rather than its availability in long-term memory.

Second, because monitoring is not independent of retrieval, if retrieval goes wrong, so will monitoring. Thus, retrieval may be fooled by a variety of clues deriving from many sources, such as neighbouring targets, priming, misleading postevent information, and so on. In that case monitoring too will go wrong.

Going back to the Wason-McKellar example, according to the trace-access model, the FOK taps the trace of the *correct* target (McKeilar) even when the partial information that comes to mind stems from another, misleading source. In contrast, according to the accessibility position, the partial information accessed *is* the very basis for the FOK. Because the FOK is computed on line, it must reflect the overall accessibility of information at every point in time. Therefore, every clue that comes to mind will enhance of the FOK unless and until it is proven to be wrong or irrelevant. This implies that the strong FOK that I had about McKeilar stemmed in part from the partial information accessed about Wason!

Explaining the accuracy and inaccuracy of the feeling of knowing

Let us turn next to the accuracy of the FOK: The question is why is the FOK accurate in predicting actual memory performance if it merely monitors the overall accessibility of information regardless of whether it is correct or wrong'.'

The answer is simple: A basic tenet of the accessibility account is that the accuracy of metamemory stems directly from the accuracy of memory itself. Memory is by and large accurate in the following sense:

Information that comes to mind during retrieval is more likely to be correct than wrong. This is almost part of the definition of memory: If you learn the name of a person, you are more likely to recall that name in the future than to recall another name instead. Indeed in free recall tests, for example, most of the items that a subject reports are correct, and only a small proportion represent commission errors. The same is true for partial recall. We used the term "output-bound accuracy" to designate the likelihood that what I do recall is correct (Koriat & Goldsmith, 1994, in press). Because most of the information that comes to mind is correct, a monitoring mechanism that relies solely on the accessibility of information, as such, is bound to be predict actual recall and recognition performance.

In sum, it is argued that we have no access to information that is not already contained in the output of our retrieval attempts. Furthermore, we cannot directly monitor the correctness of the information that we do retrieve. Ail we can do is use the overall accessibility of information to form a FOK judgment about the presence of the target in the store. This inference tends to be accurate, because memory itself is generally accurate.

The accessibility model of feeling of knowing: A demonstrative study

I will now sketch an accessibility model of the FOK and examine some predictions. The model (see Figure 1) assumes that when we search our memory for a particular target, many clues come to mind. Some of these originate from the target and form "correct partial information," whereas others stem from many other sources and represent "wrong partial information". In general, the stronger the memory strength of the target, the larger the amount of correct partial information and the stronger the likelihood of accurate recognition. Also, the stronger the memory trace, the less the likelihood that misleading clues will intrude. Thus, positive correlations are expected between the three components representing "objective knowing" - memory strength, correct partial information, and recognition - and all should be negatively correlated with the accessibility of wrong partial information.

The FOK depends on the accessibility of partial information *regardless of its correctness.* Both correct and incorrect partial information contribute to the FOK. Therefore correct partial information is responsible for the accuracy of FOK. whereas wrong partial information is responsible for its inaccuracy.

Let us examine some illustrative results from one experiment (Experiment 1; Koriat, 1993). On each trial, subjects memorized a four-letter string (e.g., *BLMD).* Following a filler task which lasted 18 seconds, they were asked to report the full target or as many letters as they could remember. Finally they provided FOK judgments about the probability of identifying the target among distractors, and their recognition memory for the target was tested.

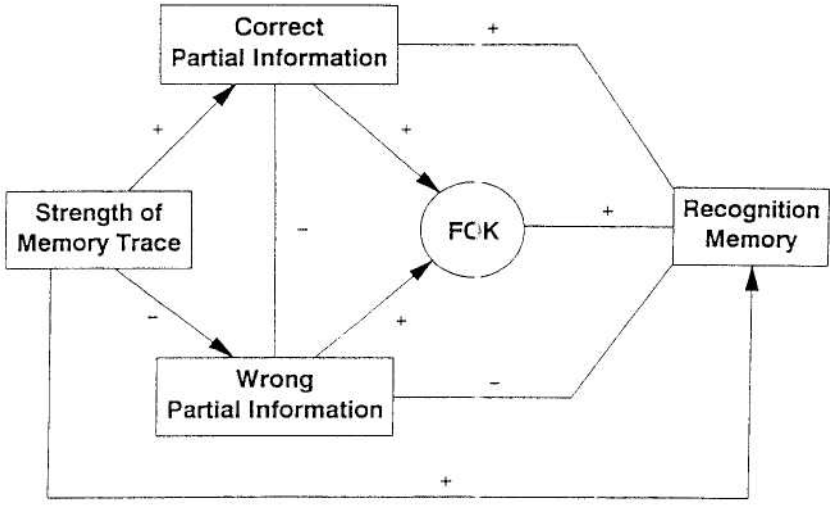

Figure 1. The accessibility model of the feeling of knowing. The + and - signs indicate the postulated positive or negative links between the respective components.

effects of partial information on the FOK. Correct partial information was simply defined as the number of correct letters that the subject was able to access, whereas wrong partial information was defined as the number of incorrect letters reported (like the letters *W* and *N*, when "McKellar" was the target). The FOK increased significantly as a function of the correct partial information accessed: The estimated correlation for the grouped data was +.83. However, it also increased significantly with the amount of *incorrect* partial information accessed, the respective correlation being +.76. As far as recognition memory is concerned, however, this was positively correlated with the amount of correct partial information, +.61, but negatively correlated with tile amount of incorrect partial information, -.52. Thus, correct partial information contributes to the accuracy of the FOK, whereas incorrect partial information contributes to its inaccuracy.

Nevertheless, despite the conflicting contributions of correct and wrong partial recalls to the validity of the FOK, the overall correlation between the FOK and recognition was positive +.55. The reason is that the information accessed tends to be correct, by and large: About 90% of the letters recalled were correct. Therefore although subjects cannot tell whether the partial information retrieved is correct or wrong, the total amount of information retrieved is a sufficiently good predictor of recognition memory.

To return to the question *of privileged access.* Do FOK ratings have access to information that is not contained in the retrieved information? Apparently not. In fact, the subject's FOK judgments had no greater predictive validity than (he mere amount of information retrieved. Indeed the correlation between number of letters recalled (regardless of their correctness) and recognition memory was .58, which is about the same as that between FOK and recognition 55.

Thus, I do not need the subject's FOK judgments; I can piedict his recognition performance no less successfully than his FOK judgments do, provided that I know the amount of information that is accessible to him. In this sense FOK judgments have no privileged access to information that is not already contained in the output of the retrieval attempt.

Dissociating knowledge and metaknowledge *Relating*

metaknowledge accuracy and knowledge accuracy: Back to phonetic symbolism

I shall return now to the puzzle with which I started. As I noted earlier, subjects were able to guess the meaning of foreign words with a belter-than-chance success. Furthermore, they were found to accurately monitor their performance. The question, again, is how can people tell whether they are right or wrong in this task?

According to the view that I have proposed here, people have no way of monitoring *directly* the accuracy of their knowledge. However, they can do that indirectly, by focusing on cues that are correlated with accuracy. In the case of memory, I have shown that the amount of information retrieved is indeed a valid cue, because *memory* itself is generally accurate. Hence the accuracy of metamemory is inherently tied to the accuracy of memory.

The same argument can be made with regard to the phonetic symbolism results. These results indicated two findings which, at first sight appear unrelated: First, subjects were generally accurate in guessing the meaning of foreign words. This aspect may be referred to as "knowledge accuracy". Second, subjects were able to monitor the correctness of their guesses. This may be termed "metaknowledge accuracy". It is argued that the two findings are related, because metaknowledge accuracy derives directly from knowledge accuracy.

The argument is simple: We can talk about the *strength* of a match, the extent to which a certain translation strikes you as being more correct than the alternative translation. A simple index of strength is intersubject consensus: the more compelling is the match the larger the number of subjects who endorse it.

Now the basic finding concerning "knowledge accuracy" is that correct matches are more compelling than incorrect matches; they are consensually endorsed by the majority of subjects. As far as metaknowledge accuracy is concerned, however, we do not know whether the FOK monitors directly the correctness of the match. Perhaps it monitors only the *strength* of the match, but because the more compelling matches tend also to be correct, it succeeds in predicting the accuracy of the match.

In order to examine this possibility, it is necessary to destroy knowledge accuracy and see what happens. Thus, we must find items for which most subjects actually agree on the *wrong* translation, that is, items in which the *wrong* match happens to be **the** more compelling, and see whether for such items the FOK monitors the correctness of the match or its consensuality.

To do that, I searched for such items across several studies including mine, and compiled a list of 85 items from 6 languages, with a heavy representation of deceptive items, items for which people tend to consensually endorse the wrong translation (Koriat, 1976). One hundred English speaking subjects were asked to match the members of each foreign pair with their English equivalents, and to indicate their FOK ratings. On the basis of the translation results I classified the items into three classes according to whether the majority of subjects significantly agreed on

labeled consensually-correct (CC), consensually-wrong (CW), and nonconsensual (NC), respectively.

Turning now to the FOK ratings, the relationship between these ratings and the correctness of the translation was found to differ markedly for the three classes of items. For the CC class, correct answers were indeed associated with stronger FOK ratings, as we found before. However, for the CW class it was the *wrong* answer that was associated with the stronger ratings. For the NC class FOK judgments were unrelated to the correctness of the match.

Conversely, when we examine percentage of correct translations as a function of FOK ratings, we see that for the CC items the correlation was positive and strong. For example, when subjects assigned a FOK rating of 4, they were accurate in 85% of the cases! For the CW items, in contrast, the correlation was negative: A high FOK rating predicted the wrong answer. The NC class yielded again no correlation.

These results clearly demonstrate the dependence of metaknowledge accuracy on knowledge accuracy: As knowledge accuracy improves, metaknowledge accuracy improves. Thus, subjects can monitor the correctness of their matches only when these matches are more likely to be correct than wrong. In contrast, when the matches art: consensually wrong the more one feels that one knows the more likely that one is wrong.

Should we conclude then that subjects cannot monitor their knowledge?. Of course not. Subjects in a phonetic symbolism experiment do know when they are right and when they are wrong. This is what we find when we use a representative sample of items. However, subjects monitor their knowledge indirectly, by relying on cues that arc correlated with accuracy.

An analysis of memory pointers

Let us return now to the situation in which the FOK is typically elicited - the retrieval of real-world knowledge from long term memory. Is it possible that knowledge and metaknowledge are also dissociable for this task?

In what follows, I shall use the term "memory pointer" (see Koriat and Lieblich, 1977), to designate any cue that is intended to specify a particular memory entry. This may include a general-information question that specifies a particular memory entry, a word definition, or a stimulus word in a paired-associate task. Some of our results suggest that a great deal of information about the FOK can be gained from an analysis of memory pointers.

Some pertinent data come from a study on the tip-of-the-tongue phenomenon that we conducted many years ago (Koriat & Lieblich, 1974, 1975; 1977). That study indicated that people who report a TOT state are more successful in retrieving correct partial information than people who report a "Don't Know" state, suggesting that subjective judgments monitor the correctness of the information retrieved.

In that study we used the list of word definitions that had been used in the classic study of Brown and McNeill (1966). When I received a copy of the stimulus list from Roger Brown, I noted a peculiar comment: Some words had an asterisks next to them saying "Words especially effective for eliciting a TOT state". This appeared to me rather odd. 1 had thought that the TOT state represents a peculiar state of mind that depends on idiosyncratic variables that vary from situation to situation and from person to person. But here it is implied that the likelihood of a TOT state is a reliable property of the pointer. We decided to examine this reliability on our TOT data (Koriat & Lieblich, 1977). We divided our sample of subjects randomly into two groups, and for each group we calculated the frequency of subjects signalling a TOT state for each pointer. The correlation across pointers between the two group frequencies was .62 *(p<.000l).* This is

quite surprising, because it implies that you can predict rather well whether a person will experience a TOT state simply from knowing what question was presented to him.

In order to clarify what pointers elicit a TOT state, we also calculated the frequency with which each pointer elicited other memory states, such as "know-correct", "tip of the tongue - incorrect", etc., and submitted these frequencies to a multidimensional analysis. The analysis indicated that memory pointers differ along two dimensions: The first is actual knowledge: the effectiveness in suggesting or eliciting the correct target. The second is metaknowledge: the degree of initial feeling of knowing elicited. What is surprising is that the two dimensions are *orthogonal,* suggesting that the FOK rests on aspects of a pointer that are different from those which determine the success of retrieving the correct target. That is, some pointers tend to produce a high FOK or a low FOK independent of their tendency to elicit the correct answer.

An analysis of the pointers that tend to produce an overly high FOK clearly suggests that the critical factor is the amount of information they tend to elicit. This amount seems to depend both on characteristics of the question as well on characteristics of the solicited answer. I will mention just two examples: As far as the question is concerned, questions that contain redundancies and repetitions tend to produce inflated feelings of knowing, possibly because they generate a greater amount of activation without increasing the chances of eliciting the answer. As far as the solicited answer is concerned, questions for which the target has many "close neighbours" tend to produce inflated FOK even when the subject ultimately retrieves the correct target. It would seem that the FOK is based on an unfocused scanning of a broad region of memory in which the target is likely to reside. Other memory entries in the vicinity of the target can be mistaken for the target when they are inspected from a distance, although when you get closer you can easily reject them and home in on the correct target.

Dissociating knowing and the feeling of knowing

The correlational analysis of memory pointers suggests that pointers differ reliably in characteristics that are pertinent to both knowledge and metaknowledge. These differences can perhaps be exploited to examine the basic questions about the FOK, and to produce a dissociation between knowing and the feeling of knowing. This is what I did in a recent study which has just been submitted for publication (Koriat, 1994b).

That study addressed two questions: what is the basis of the FOK and what is the basis for its accuracy. As far as the basis of the FOK is concerned, it would seem that pointers which bring to mind many clues should result in a stronger FOK than those eliciting few clues. A simple index of the amount of accessible information elicited by a pointer is the percentage of subjects who produce an answer to that pointer in recall, regardless of whether the answer is correct or wrong. This will be called the accessibility index (ACC). The hypothesis is that high-accessibility pointers will result in a high FOK judgment even among subjects who fail to recall any answer. This is because such pointers are expected to leave behind a large number of clues when recall fails.

What about the accuracy of these FOK judgments? This is assumed to depend on the correctness of the clues that come to mind, as can be inferred from the correctness of the answers provided in recall. The proper index is what we called output-bound accuracy (OBA; see Koriat & Goldsmith, 1994; in press), that is, the percentage of correct answers out of all the answers elicited by a pointer. This index reflects the likelihood that an answer that comes to mind is correct. The prediction is that the FOK will be valid for pointers that produce more correct than incorrect answers. These pointers are representative of memory pointers in general, and will be

designated "Consensually Correct" (CC). However, there are many deceptive questions for which this is not true (Fischhoff, Slovic, & Liclucnstcin, 1977). For example, the question "What is the capital of Australia'.'" elicits more incorrect than correct answers. Such pointers are designated "Consensually Wrong" (CW). If metamemory accuracy depends on memory accuracy, then the FOK will be valid only for the CC pointers, but not for the CW pointers.

To examine these questions it was necessary to compile a sufficiently large number of deceptive pointers. For Experiment 1, 95 general- information pointers were used, with a heavy representation of deceptive pointers. All required a one-word answer. A typical FOK procedure was used: Subjects attempted to recall the answer, then they provided FOK judgments, and finally were tested on a 4-alternative recognition test.

I used the data from the recall phase of the experiment as the basis for classifying the pointers in terms of the two indexes, accessibility and output-bound accuracy.

Consider first *the basis of FOK judgments.* The accessibility index ranged from 6% to 97% across pointers. All pointers were divided into a high-accessibility and a low-accessibility class. In general, FOK judgments were markedly higher when an answer was provided (90.5) than when the trial culminated in omission (59.1). However, high accessibility pointers engendered higher FOK judgments than low-ACC pointers for both omission trials, and commission trials. That is, even if I cannot find any answer, my feeling of knowing is stronger for questions that elicit many answers among other people than for those eliciting fewer answers. Presumably the former questions leave behind a larger amount of activation even when recall fails.

Let me turn next to *the accuracy of FOK,* which is assumed to vary with the output-bound accuracy of the accessible information. The output-bound accuracy index ranged from 0.0, that is, all the answers provided were incorrect (e.g., "Who was the first Roman Caesar") to 100.0, that is, all the answers were correct ("What is the original family name of the singer Bob Dylan?"). 52 pointers were consensually-correct, in that they produced more correct than incorrect answers, and 37 were consensually-wrong, precipitating more incorrect than conect answers across subjects.

I shall focus only on trials in which subjects failed to produce an answer. FOK accuracy was quite accurate in predicting recognition memory in the case of the CC pointers: The within-subject correlation between FOK judgments and recognition memory was .50.In contrast, for the CW class, F'OK judgments were unrelated to actual recognition performance; the correlation was -.05. The interaction was highly significant. In fact, the CW pointers evidenced a very marked illusion of knowing: Whereas FOK judgments averaged about 90% for the high-FOK pointers, recognition performance averaged only about 35%, barely better than chance!

Is it possible to produce a negative correlation between knowing and the feeling of knowing as we did with the phonetic symbolism task? In a second experiment, I focused on the preliminary, immediate FOK that people have before attempting to search for the target deliberately (Reder & Ritter, 1992). In this manner it was hoped to minimize reliance on specific information that subjects may use when FOK judgments rely on an effortful, analytic inference. 37 CC and 37 CW pointers were selected, on the basis of the results of Experiment 1. Of course, a pointer must produce some answers in recall in order to be reliably classified as CC or CW. Therefore these pointers can be said to represent a moderate-to-high accessibility level. Thus, another set of 36 Low-Accessibility (LA) pointers was added, comprising difficult questions that tend to evoke practically no free-recall responses overall.

The procedure was similar to the previous experiment. Subjects were urged, however, to provide a fast, preliminary FOK judgment, but to write down the answer if it comes to mind

spontaneously. A 2-alternative rather than a 4-alternative recognition test was used. Subjects did reach an answer in 6% of the trials and these were eliminated from the analyses.

Consider first *the basis of the FOK.* If we focus only on the CC and CW pointers for which the ACC index was available, we find that preliminary FOK judgments were indeed higher for the High-Accessibility (80.7) than for the Low-Accessibility (72.3) pointers, and this was true for both the CC and the CW pointers. Thus, the number of answers elicited by a pointer in one group of subjects predicts the preliminary FOK in another group of subjects.

Let us turn now to *FOK accuracy.* Consider first the comparison between the means of the three classes of pointers. The results demonstrate a double dissociation between knowing and the feeling of knowing. The CC and CW pointers evoked practically identical preliminary FOK judgments, (76.1 and 76.4, respectively), but differed considerably in recognition memory (73.3 and 43.9, respectively). On the other hand, the CW and LA pointers yielded similar recognition performance (43.9 and 45.1, respectively), although they differed markedly in their preliminary FOK judgments (76.4 and 58.7, respectively).

Turning next to a within-subject analysis of the predictive validity of the FOK, we find again a relatively good monitoring and good calibration for the CC pointers: Recognition memory increased significantly with increasing FOK, and the within-subject correlation averaged .31. In contrast, for the CW pointers recognition memory *decreased* significantly as FOK increased, yielding a significantly negative correlation within individuals, -.18! Thus, the more one feels that one knows, the less likely that one actually knows. The correlation for the LA pointers was .02.

In sum, the results clearly demonstrated that the correlation between the FOK and recognition memory can be positive, negative, or zero depending on the recall characteristics of the set of items included in the sample. These results help illustrate the criticality of using a representative sample of items when making conclusions about the validity of subjective monitoring (see Gigerenzer, Hoffrage, & Kleinbolting, 1991).

Furthermore, the results suggest that it is not useful to conceive of preliminary FOK judgments as emanating from a preretrieval stage (e.g., Reder & Ritter, 1992). Rather, these judgments also rely on a fast, shallow inspection of the partial activations elicited by a pointer.

The interpretation of the results of these experiments is based on the idea that the amount and type of answers elicited by a pointer can serve as a rough index of the amount and quality of the partial clues that are left behind when recall fails. Experiment 3 supported this assumption. In this experiment, when subjects failed to recall an answer, they were asked to list the partial clues that they could retrieve about it, for example, its initial letter, the number of syllables it contains, etc. The results indicated the following: First, the amount of partial information accessed was indeed higher for the high-accessibility than for the low-accessibility pointers. That is, a pointer that elicits a full answer among many subjects tends also to evoke a large amount of partial information when recall fails. Second, the accuracy of the partial information retrieved was indeed higher for questions eliciting mostly correct answers than for those eliciting mostly incorrect answers. Such a pattern can explain why the deceptive pointers evoke a strong, unwarranted FOK when recall fails.

The illusion of knowing

The results of this recent study can give us some insight into the origin of the illusion of knowing, and I am going to comment only briefly on this very important issue. Why do I have a strong conviction that I know the answer despite the fact that I do not know it? As I have argued,

metamemory goes wrong when memory itself goes wrong. But in what way does memory go wrong?

We tried 10 study the illusion of knowing by examining the nature of the pointers which produced a strong unwarranted F'OK across subjects and experiments. What is the nature of these deceptive pointers? It would be too simple to assume that deceptive pointers are merely those for which the subject happens to hold the wrong answer in memory. Actually, almost all of the deceptive pointers in our study turned out to evoke more than a one incorrect answer across subjects. In fact 50% elicited four or more different incorrect answers, and two elicited as many as 9 different incorrect answers.

Our analysis of these pointers suggests that a high degree of accessibility is necessary for the illusion of knowing. There are two conditions which contribute to enhanced accessibility. First, the pointer must be familiar enough to lead us to interrogate our memory for an answer, and explore possible answers. This increases the overall accessibility of information that is left behind when we fail to find an answer. A second condition is that the solicited target resides in a memory region with a high density of memory entries. Presumably the FOK monitors the accessibility of information from the entire memory region. Therefore initial FOK is high whether or not we finally succeed in pushing aside some of the interlopers while homing in on the correct target.

A third condition, however is necessary for the pointer to also be truly deceptive; The correct answer must be less familiar and less accessible than one or more of the other candidates.

It is important to stress that what matters is the accessibility rather than the availability of information. As I argued before (Koriat, 1993), the FOK monitors the activated information in short-term memory rather than the stored information in long-term memory. For example, the question "In which state of the USA is Yale University located"? generated 9 different incorrect answers among Israeli subjects, and produced a strong, unwarranted FOK. In contrast, the comparable question "In which of the USA states is the College of William and Mary located?" generated few recall answers and low FOK.

Retrospective confidence

I will end with a few comments about retrospective confidence. Retrospective confidence is measured when the subject is asked first to answer a question, and then to judge how confident he is in the correctness of that answer. Unlike FOK judgments, which are basically *prospective*, expressing confidence in the future recall or recognition of the target, retrospective confidence concerns a decision that has already been made. In terms of the view of subjective monitoring that I have proposed, retrospective confidence should present some interesting problems that touch upon the general issue of privilege access to the determinants of one's judgments and behavior.

The issue of privileged access revisited

Nisbett and his associates (Nisbett & Bellows, 1977; Nisbett & Wilson, 1977) have proposed thai people have little direct introspective access to the actual determinants of their behavior. Rather, when asked to report on the reasons for their behavior, people simply report those reasons that according to their theory constitute *plausible* determinants of their behavior.

Consider, then, a situation in which a person who has just reached an answer to a question, or has just chosen between two alternative answers, is asked to express his confidence that the

answer is correct. For example, a person must decide which city is farther north, Rome or New York. Let us assume, as Nisbett argues, that the person is not aware of the actual determinants of his choice. Actually he chose the answer on the basis of reason x, but he consciously believes that he has done so because of reason y. On which of the two reasons will lie base his subjective confidence?

Some insight into this question conies from a study on retrospective confidence that we carried out several years ago (Koriat, Lichtenstein & Fischhoff, 1980). The study addressed the question of why people are overconfident in the correctness of their knowledge. A general finding in calibration studies is that people overestimate the correctness of their performance. For example, when they assign a probability of .90 the actual probability of being correct is only about .75. We explained this discrepancy by proposing that subjective confidence essentially represents an attempt to justify a decision that has already been made. The answering of forced-choice questions, then, may be conceptualized as involving two stages. In the first stage, subjects interrogate their memory for pertinent considerations in an unbiased fashion until a decision is reached. In the second, *retrospective review* stage, the evidence retrieved in the first stage is reviewed to assess the likelihood that the answer is correct. This retrospective review tends to be *biased* by the decision already reached: it tends to focus on evidence that supports the decision, and to disregard evidence contradicting it, thereby resulting in overconfidence in the decision reached. Thus, subjective confidence rests on a process of self justification.

In an extension of this work (Koriat, Fischhoff, & Slovic, 1990), we suggested that the retrospective review of evidence underlying subjective confidence is based on the *perceived* determinants of one's decision; not on the actual determinants. This implies that when an answer is actually chosen because of reason x, but the person believes that it was chosen because of reason y, it is the latter reason that should affect confidence in the correctness of the answer.

Retrospective confidence: Arbitrating between actual and perceived determinants

This idea assumes that the factors that determine confidence in a decision may differ from those that determine the decision itself, which is, of course, consistent with what I have been saying about the dissociation between knowledge and metaknowledge.

We are presently exploring some possible predictions concerning conditions in which there is a discrepancy between the *actual* reasons and the *perceived* reasons for one's decision (see Slovic, Fleissner, & Bauman, 1972). We argue that in these conditions confidence judgments must arbitrate between the two types of reasons: The decision, which was itself determined by reasons that are not acknowledged by the person, must be justified in retrospect in terms of the perceived, influential factors. Thus, assume that we can estimate the *plausible* decision, that is, the decision that the person should have made had he indeed relied on the perceived influential factors. It may be then hypothesized that retrospective confidence in a decision should decrease as the actual decision diverges from the plausible decision. So far, support for this idea was quite limited, but work is still in progress.

In sum, these comments about retrospective confidence further help illustrate the view of subjective monitoring that I have advocated. We, the subjective monitors of our knowledge, must make our judgments from the outside. We cannot monitor directly the presence of information in the memory store, and our feeling of knowing must therefore rely on a nonanalytic inference that utilizes cues that are present in the output of the memory system. We also cannot monitor directly the correctness of the clues that come to mind, or the correctness of an answer that we have finally reached, but must infer that from cues such as the ease with which information comes to

mind. Finally, even when we form a subjective conviction in the correctness of a decision, we act as outsiders: We watch a decision which originated from factors that are foreign to us, and must now justify that decision in terms of our conscious, subjective theory.

Nevertheless, by and large, we are generally accurate in monitoring the correctness of our knowledge. This is an important achievement which should not be taken for granted.

Baars, B. J. (1988). *A cognitive theory of consciousness.* Cambridge: Cambridge University press.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenome non. *Journal of Verbal Learning and Verbal Behavior, 12,* 711-319.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109,* 204-223.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal, of Verbal Learning and Verbal Behavior, 5,* 325-337.

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

Eyscnek, M. W. (1979). The feeling of knowing a word':; meaning. *British Journal of Psychology, 70,* 243-251.

Fisehhoff, B., Slovic, I, & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552-564.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory <& Cognition, 1,* 213-216.

Gigerenzer, G., Ho ff rage, U., & Kleinbolting, H. (1941). Probabilistic mental models. A Brunswikian theory of confidence. *Psychological Review, 98,* 506-528. Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56,* 208-216.

Hart, J. T. (1967a). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6,* 685-691.

Hart, J. T. (1967b). Second-try recall, recognition and the memory-monitoring process. *Journal of Educalio nal Psychology, 58,* 193-197.

Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition, 17,* 69-76.

Koehler, W. (1947). *Gestalt Psychology.* New York: Lippincott.

Koriat. A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition, 3,* 545-548.

Koriat, A. (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition, 4,* 244-248.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609-639.

Koriat A. (1994a). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J.Mclcalfc & A. P. Shimamura (lids.), *Metacognition; Knowing about knowing* (pp. 115-135). Cambridge, MA: MIT Press.

Koriat, A. (1994b). *Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model.* Manuscript submitted for publication.

Koriat, A. Ben-zur, H. (1988). Remembering that I did it: Processes and deficits in output monitoring. In: Gruneberg, M., Morris, P., and Sykes, R. (Eds.) *Practical aspects of memory; Current research and issues, Vol I,* (pp, 203-208) Wiley.

Koriat, A., Ben-z.ur, II., and Shcffcr, D. (1988). Telling the same story twice: Output monitoring and age. *Journal of Memory & Language, 27,* 23-39.

Koriat, A., Fischhoff, B., & Slovic, P. (1990). *Subjective convictions when intuitions are wrong; A study of subjective confidence inr elation to actual and perceived reasons for behavior.* Unpublished research report to the U.S -Israel Binatioa! Science Foundation.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the

Koriat A., & Goldsmith, M. (in press). Memory as something that can be counted vs. memory as something that can be counted on. In D. Herrman, M. Johnson, C. McEvoy, C. Hertzog, & P, Hertel (Eds.). *Basic and applied memory; Research on practical aspects of memory,* Hillsdalc, NJ.: Erlbaum.

Koriat, A., Lichtenstein, S., & Fischhoff, B, (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107-118.

Koriat, A., & Lieblich, I, (1974). What does a person in a "TOT" state know that a person in a "don't know" state doesn't know? *Memory & Cognition, 2,* 647-655.

Koriat, A. and Lieblich, I. (1975). Examination of the letter serial postilion effect in the "TOT" and the "Don't Know" states. *Bulletin of the Psychonomic Society, 6,* 539-541.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Ada Psychologies, 41,* 151-164.

Lewicki, P. (1986). *Nonconscious social information processing.* New York: Academic Press.

Lovelace, E. (1987). Attributes that come to mind in the TOT stale. *Bulletin of the Psychonomic Society, 25,* 370-372.

McKellar, P. (1957). *Imagination and Thinking.* Mew York: Basic Books.

Marcel, A. J. & Bisiach, E. (Eds.). (1988). *Consciousness in contemporary science.* Oxford: Clarendon Press.

Montgomery, H. & Svenson, O. (Eds.). (1989). *Process and Structure in Human Decision Making.* Chichester: Wiley.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation, advances in research and theory* (Vol. 26, pp. 125-173). San Diego, CA: Academic Press.

Nisbett, R. E, & Bellows, N. (1977). Verbal reports about causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology, 35,* 613-624.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84,* 231-279.

Reag, J. D., & Bruce, D. (1982). Longitudinal tracking of difficult memory retrievals. *Cognitive Psychology, 14,* 280-300.

Reason, J., & Lucas, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J. *E.* Harris & P. E. Morris (Eds.), *Everyday memory, actions and absentmindedness* (pp. 53-69). San Diego, CA: Academic Press.

Reder, L. M. & Ritter, F. E. (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.

Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling of knowing. *Canadian Journal of Psychology, 39,* 467-475.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Slobin, D. I. (1968). Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology, 10,* 301-305.

Slovic, P., Fleissner, D. & Bauman, W. S. (1972). Analyzing the use of information in investment decision making: A methodological proposal. *Journal of Business, 45,* 383-301.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology, 21,* 434-437.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4,* 207-232.

Wason, P.C., & Johnson-Laird, P.N. (Eds.). (1968). *Thinking and Reasoning: Selected Readings.* Hards monds, Middx.: Penguin Books.

Yaniv, I., & Meyer, D. E (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187-205.