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

Memory's Knowledge of Its Own Knowledge: The Accessibility Account of the Feeling of Knowing

Asher Koriat

This chapter contrasts two theoretical approaches to the feeling of knowing. According to the commonly held trace-access approach, when people fail to recall a target from memory, they can nevertheless provide feeling-of-knowing (FOK) judgments by monitoring the presence of the target's trace in store. This approach assumes a two-stage, monitoring-and-retrieval process, where people first ascertain the availability of the target in store before attempting to retrieve it. An alternative single-process account advocated in this chapter is that FOK is computed during the search and retrieval process itself, relying on the overall accessibility of partial information about the target. The implications of this approach for the analysis of the accuracy and inaccuracy of FOK are discussed, and some supportive experimental evidence is presented. This evidence suggests that people have no privileged access to information about the target's presence in store that is not already contained in the output of the retrieval attempt.

What Do We Know When We Don't Know?

There are two general properties of memory that are readily demonstrated both in everyday experience and in the laboratory. First, die information that we can retrieve at any one moment represents only a fraction of what we actually know. In the terminology of Tulving and Pearlstone (1966), more information is *available to* people than is *accessible* to them. The second property is that memory is not an all-or-none matter. Thus, even when we fail to retrieve a specific target from memory, we may still be able to say something about it.

The information that we can often supply about an unrecallable target is of two different sorts. First is a *feeling-of-knowing* (FOK) judgment, conveying our subjective assessment that we "know" the target to the extent of being able to recall or recognize it in the future. The second consists of some *partial or generic information about* the target. For example, even when we fail to recall the name of a person, we may still be able to tell what it sounds like.

A question that naturally arises concerns the *validity* of the information supplied regarding the unrecallable target. Interestingly, both FOK judgments and partial information tend to be quite accurate, suggesting that people can somehow "get a glimpse" of the unrecalled target. Consider FOK judgments first. Many studies confirmed that these judgments are accurate in predicting the likelihood of recalling the target in the future, producing it in response to clues, or identifying it among distractors (e.g., Freedman & Landauer, 1966; Gardiner, Craik, & Bleasdale, 1973; Gruneberg & Monks, 1974; Gruneberg & Sykes, 1978; Hart, 1965a, 1967a, 1967b; Leonesio & Nelson, 1990; Nelson & Narens, 1990; Schacter, 1983).

With regard to partial information, the classic study by Brown and McNeill (1966) has indicated that the information that comes to mind in the tip-of-the-tongue (TOT) state tends to be accurate. Thus, subjects were able to guess various features of the inaccessible word, such as the initial letter, the number of syllables, the location of the stress, and so on (see also Brown, 1991; Koriat & Lieblich, 1974; 1975; Smith, this volume). Other studies still indicate that subjects can also gain accurate information about some of the word's *semantic* attributes (Schacter & Worling, 1985; Yavutz & Bousfield, 1959). For example, in an unpublished study in our laboratory (Erdry, 1990), subjects unable to recall the translation of a so-called Somali word were accurate in judging its connotative meaning with regard to the three dimensions of the semantic differential, good-bad, active-passive, and strong-weak. Their judgments were accurate even after a 1-week period.

The present chapter focuses on FOK judgments, but I shall use some of the observations regarding partial information to help clarify the mystery surrounding the FOK phenomenon. Two questions about the FOK suggest themselves. First, what is the *basis* for the feeling of knowing? Second, what makes such subjective feelings *valid predictors* objective memory performance? These two questions are, of course, related, because a satisfactory model of the basis of FOK judgments must also provide an explanation for their validity.

What Is the Referent for FOK and Partial Information?

I would like to relate a personal experience and use it to highlight some of the issues pertaining to FOK and partial information: During one of the conferences on memory, I tried to recall the name of the author of a particular book, a book that I had read man}' years earlier. I tried hard, but for some strange reason I could not retrieve it. Only some letters came to mind, and these made me all the more frustrated for not being able to home in on the name: I felt quite sure that the name contained Wand *N,* and was somewhat less confident about a third letter, *S.* I struggled with the name for a whole day, trying to play with various permutations of the letters to help retrieve the entire name.

In the evening, I went for a walk with a friend, an expert on the TOT phenomenon, who saw me in my anguish and offered his help. I described to him what I knew about the book — that it was a Penguin book on thinking, with a bluish cover — and also communicated to him the letters that I was able to access. Luckily, he remembered a Penguin book that roughly fits the description, as well as the name of the author: Wason! At that point I had some insight about what was happening: I knew the Penguin book edited by Wason (Wason & Johnson-Laird, 1968), and 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. However, I also realized where the partial information was coming from: It was most probably coming from *Wason!* I made an effort to put aside the letters that came to mind, and after a while I successfully retrieved the name: It was McKellar!

This example illustrates one of the accounts of the TOT state. According to that account, the failure to retrieve the target in the TOT state stems, in part, from the interfering effect of "blockers" or "interlopers" that come to mind (Burke, MacKay, Worthley, & Wade, 1991; Jones, 1989; Reason & Lucas, 1984). Such interlopers represent plausible candidate answers that interfere with accessing the correct target.

Let us assume that "Wason" constitutes such an interloper, and "McKellar" represents the correct, ultimate target. The example described above then presents a dilemma: When we fail to access the full target, but are able to provide partial information and FOK judgments, which is the actual *referent* for these responses? In other words, when I cannot recall an item and yet can access some information, what is that information about? With regard to partial information, the example mentioned above indicates that the phonological clues that came to mind were quite accurate in predicting the *wrong* referent (Wason), and were way off as far as the correct target is concerned (the name "McKellar" does not contain any of the letters that came to mind). With regard to FOK judgments, however, it is not clear which of the two targets was being monitored. Evidently, throughout the entire search process I had a very strong positive FOK, and this turned out to be valid, because I ultimately succeeded in recalling the correct name (McKellar). Thus, is it possible that a *dissociation* exists between partial information and FOK, so that FOK continues to monitor the availability of the *correct* target in store, even when we receive "vibrations" from a related, but *incorrect* target?

To complicate the story further, when preparing the references for this chapter I discovered to my surprise that McKellar's book was *not* a Penguin book. I thought I had the book, so I went to look for it in the place where it was supposed to be, but the book I pulled out from the shelf was not McKellar's. Rather, it was a blue-cover Penguin book by Thomson en tided *The Psychology of Thinking.* So, perhaps, it was this book that gave rise to the partial attributes "bluish" and "Penguin" (I do not know even now what was the color of the cover of Wason's or McKellar's books). "Thomson" may have been also responsible for some of the letters accessed (S and N), though I must admit that I had no recollection of having ever read Thomson's book.

The example cited above helps illustrate some of the theoretical dilemmas raised by memory-blockage states such as those associated with a strong TOT and FOK. These states are of particular interest because they combine 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 does a person know that he/she knows the answer in the face of being unable to produce it? In what follows I shall contrast two general accounts of the FOK that attempt to address this question, the trace-access account and the accessibility account.

The Trace-Access Account of FOK

A simple and elegant model that explains both the *basis* of FOK judgments as well as their *accuracy* is the trace-access model. This model, first advocated by Hart (1965a, 1967a, b; see Nelson, Gerler, & Narens, 1984; Yaniv & Meyer, 1987), assumes that FOK judgments directly monitor the *availability* of the solicited target in store. These judgments are seen to represent the output of a specialized *memory-monitoring module* that can directly inspect the stored memory traces, and determine whether the target's trace is there or not. Thus, whenever a person is required to recall a target, the monitoring module is activated to make sure that the target is present in store before attempting to retrieve it. Such a monitor, then, can save the time and effort looking for a target that is not there.

This monitor-and-retrieve model can best be illustrated by drawing an analog)' to the manner in which information is organized in computerized systems. If you have had some experience with computers you must have some knowledge about *directories.* A directory is a file that catalogues other files; it contains a listing of the *names* of the files stored in a computer's memory as well as their addresses. Thus, when the computer is requested to retrieve a file from memory (analogous to a memory query), the process is something like that depicted in figure 6.1. First, the *directory* is inspected to see whether it contains the *name* of the file. If the name cannot be located, the computer returns the response "File not found" (analogous, perhaps, to "I don't know"). Note that this "don't know" response is outputted 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.*

Although it is not claimed that human memory is organized in a similar manner, the directory analog)' contains the basic ingredients of the trace-access model: First, this model postulates a special mech-

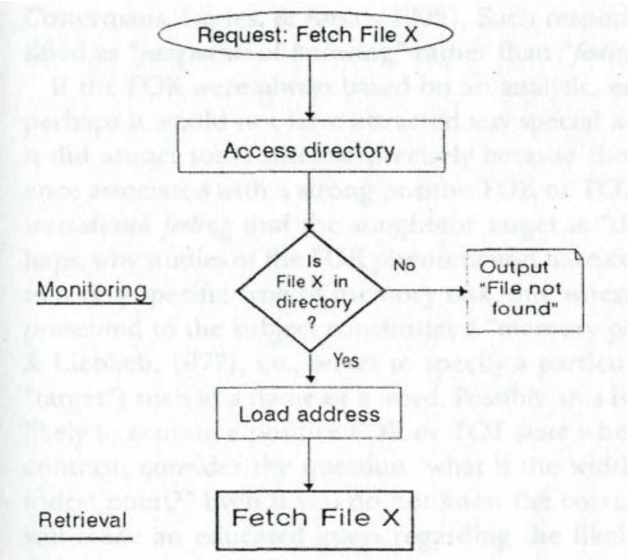

Figure 6.1
Retrieving a file in a computerized system: An illustrative implementation of a two-stage monitoring-and-retrieval process.

anism for detecting the presence of the sought for item *without having to retrieve it.* This mechanism also allows the person to reach a "don't know" decision in a way other than by failing to retrieve the target. Second, the process of answering a question is conceived as a *two-stage* process: The person first ascertains that the solicited target is available in store (analogous to consulting the directory listing) and only then embarks on an attempt to retrieve it (analogous to accessing the file itself). Such utilization of the memory-monitoring mechanism can save the time and effort searching for something that is not there. Thus, while a positive FOK can drive the search process, a negative FOK would discourage it (see Nelson & Narens, 1990; Reder, 1988). Finally, because FOK judgments rest on a process that is *independent of* that required to retrieve the target itself, a *dissociation* may be expected between the outputs of the two processes. Such dissociation should possibly be more prevalent in the fallible human memory than in computerized systems. Consider, for example, what happens when retrieval is misled by "interlopers." The dissociation

between retrieval and monitoring implies that although such interlopers (like "Wason," in die example cited above) may lead the search astray, the monitoring process continues to detect the *coned* (eventually retrieved) target ("McKellar"), despite the misleading clues that come to mind.

The strongest support for the trace access view comes precisely from the *accuracy* of FOK judgments in predicting correct recall or recognition of the target. How else would people know that they know the correct target if they cannot retrieve it, or worse, when the partial information that they access is wrong? Thus, evidence indicating that FOK is -accurate in predicting target recognition is normally seen to also constitute support for the trace-access account of FOK.

FOK as Based on Inference

The trace-access model assumes that the information pertaining to the feeling of knowing is directly available in a *ready-made format.* An alternative view, however, is that FOK judgments, like many judgments concerning future events, rest on an *inferential* process, conscious or unconscious, where several cues are utilized to assess the likelihood that a momentarily inaccessible target will be recalled or recognized at some later time. Nelson et al. (1984) listed a number of cues that can feed into the FOK, such as familiarity with the general topic and retrieval of pertinent episodic information.

Inference-based mechanisms underlying the FOK may be roughly classified into two general types, analytic and nonanalytic (see Jacoby & Brooks; 1984; Jacoby & Kelley, 1987). Analytic inferences are those in which a variety of considerations are explicitly considered and weighed to reach a probability estimate that the solicited target will be subsequently recalled or recognized. For example, in trying to recall the name of a person, I may retrieve the episode in which that person was first introduced to me, or in which I later introduced that person to a friend, and *deduce* that I must have known the name at some time. Such analytic inferences are possibly not very different from those underlying probability judgments in general. In fact, in such cases subjects may prefer to phrase their judgments as "I *must* know" or "I *believe that* I know" rather than as *"I feel I* know" (see also

Costermans, Lories, & Ansay, 1992). Such responses are better classified as *"judgments* of knowing" rather than *"feelings* of knowing."

If the FOR were always based on an analytic, educated inference, perhaps it would not have attracted any special attention. However, it did attract some interest precisely because the subjective experience associated with a strong positive FOR or TOT state is that of an *unmediated feeling* that the sought-for target is "there." This is, perhaps, why studies of the FOR phenomenon have confined themselves to a very specific type of memory task, one where the memory cue presented to the subject constitutes a "memory pointer" (see Koriat &: Lieblich, 1977), i.e., serves to specify a particular *memory entry* (a "target") such as a name or a word. Possibly, this is the situation most likely to activate a positive FOR or TOT state when retrieval fails. In contrast, consider the question "what is the width of the Nile in its widest point?" Even if you do not know the correct answer, you can still make an educated guess regarding the likelihood of selecting the correct answer from among distractors. However, it is hard to think of such a judgment as being based on an immediate *feeling* of knowing. The point that I wish to emphasize here is that "knowledge" comes in many different forms: We know the names of people and the words designating various concepts, but we also know that canaries are yellow, what the map of Italy looks like, and when America was discovered. Note, however, that the latter type of questions are not typically included in FOR studies (though they are included in studies of subjective confidence, see, e.g., Koriat, Lichenstein, & Fischhoff, 1980). This should, perhaps, be telling about the FOR phenomenon itself.

In fact, from a phenomenological point of view, the experience associated with a positive FOR or TOT is often quite similar to what is implied by the trace-access view (see James, 1890): We sometimes *sense* the unrecalled target, and can even *feel* its emergence into consciousness. Therefore, if the *feeling* of knowing is based on an inference, possibly that inference must be nonanalytic in nature, involving a global, automatic, and effortless process, where several inarticulate and undifferentiated cues contribute en masse to the FOR. Indeed, two of the accounts of FOR that have been considered in recent work represent nonanalytic heuristics, cue familiarity and accessibility. According to the cue-familiarity hypothesis (see Metcalfe, 1993; Metcalfe, this volume; Metcalfe, Schwartz, & Joaquim; 1993; Miner & Reder, this volume; Nelson et al., 1984; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992) when a person is presented with a memory query that is intended to cue a particular target, FOR is based not on the availability or retrievability of the target, but on the familiarity of the cue itself. This view has been supported by several findings indicating that FOR judgments can be enhanced by advance priming of the cues, but not by the priming of the target (Reder, 1987, 1988; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992). The accessibility hypothesis, which will be presented in detail below, assumes that FOR monitors the overall accessibility of the information pertaining to the target.

The Accessibility Account of the Feeling of Knowing

According to the accessibility model there is no need to invoke a separate monitoring module that taps directly the presence of the solicited target in store when retrieval fails. Rather, the cues for the FOR are to be found in the products of the retrieval process itself. Whenever we search our memory for a solicited target, a variety of clues often come to mind (see Lovelace, 1987). These may include fragments of the target, semantic attributes, episodic information pertaining to the target, and a variety of activations emanating from other sources. Such clues are often not articulate enough to support an analytic inference. Furthermore, they tend to have a "nonaddressable" quality that makes it difficult to attribute them to their proper source, or to judge their dependability, for example, by pitting them against each other (e.g., think of the letters that come to mind during the TOT state). However, they may act en masse to give rise to the subjective feeling that the target is "there," and is worth searching for. Thus, even when retrieval of the target fails, the scattered debris that is left behind can foster a positive feeling of knowing, a feeling that the target will be recalled or recognized in the future. The feeling of knowing, then, is based on a nonanalytic inference that considers the *overall accessibility of partial information* pertaining to the target, i.e., the overall amount and intensity of the clues that come to mind. Essentially, this accessibility heuristic represents an attempt to extrapolate from the processes that occur during the early stages of one retrieval episode to future retrieval episodes: If a memory pointer activates many associations, it is likely to eventually lead to die recollection of the target. If it leaves one "blank," chances are that it will continue to bring nothing to mind. This account of FOK is similar to the availability heuristic postulated by Tversky and Kahneman (1973) to explain how people estimate proportions or frequencies.

In sum, in contrast to the trace-access model, which implies a dissociation between monitoring and retrieval, the accessibility account assumes a *single* retrieval-and-monitoring process: It is through the process of attempting to search for the solicited target that one assesses the likelihood that it is available in store and can be recollected. FOK judgments, then, do *not* monitor memory *storage* (see Hart, 1967a). Rather, they are *computed and updated on-line,* on the basis of clues accumulated during the initial stages of search and retrieval. The monitoring process, then, is *not* independent of the retrieval process; if the latter goes astray, so will the former.

We are now in the position to take up the questions raised earlier in connection with the McKellar-Wason example. As noted earlier, according to the trace-access model, the feeling of knowing continues to tap the trace of the *correct* target (McKellar) even when the partial information that comes to mind emanates from other, misleading sources. The target that I eventually recalled, and that I recognized as the correct one was "McKellar," rather than, say, "Wason," or "Thomson." Therefore, if the feeling of knowing monitors storage rather than retrieval, it must be this target that has served to drive FOK throughout the entire search process. In contrast, according to the accessibility position, the partial information accessed in the course of the retrieval process *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 tend to contribute to the enhancement of FOK unless (and until) it is proven to be wrong or irrelevant. This implies, in a sense, that the strong feeling of knowing that I had about McKellar stemmed, in fact, from the partial information accessed about Wason!

According to the accessibility account, then, the FOK is based on the overall accessibility of information, *regardless of its source.* Thus, both *correct* and *incorrect* clues contribute-equally to the FOK. This assumption of the accessibility account distinguishes it from the *target retrievability* explanation of FOK. According to this explanation (see Nelson et al., 3984; Schwartz & Metcalfe, 1992), FOK is based on partial recall of the *target proper.* The assumption is that although subjects sometimes fail to retrieve the entire target, they may retrieve parts of it, and these are sufficient to activate a positive FOK. This can also explain the *accuracy* of the FOK, because FOK is seen to be narrowly tuned to the partial recall of the actual, *correct* target. In terms of the example used above, this would mean that the FOK emanates specifically from those clues that pertain to McKellar, implying that subjects can monitor directly the *accuracy* of the information that comes to mind.

Explaining the Accuracy and Inaccuracy of FOK

Let me now turn to the question of the *validity* of FOK in predicting actual memory performance. As noted earlier, a desirable feature of the trace-access model is that it affords a straightforward account of the *accuracy* of FOK: FOK is assumed to tap directly the trace of the inaccessible target, and hence its accuracy in predicting subsequent recognition memory. This is also true of the target retrievability account just described, where FOK is seen to tap the partial information that is specifically due to the correct target.

In contrast, it is not immediately clear how the accessibility account can explain the accuracy of FOK in predicting *correct* memory performance. In fact, the basic tenet of this account is that not only are subjects incapable of monitoring the *availability* of information in store, but they are also incapable of monitoring directly the *accuracy* of the accessible information. Therefore, if monitoring is based on the by-products of the retrieval process, then one must seek an explanation for its validity in the nature of the partial information that comes to mind during a retrieval episode.

As indicated in the introduction of this chapter, when unable to retrieve a target from memory, subjects can sometimes provide partial information about it, and this information tends to be *accurate* (e.g., Erdry, 1990; Schacter & Worling, 1985; Yavutz & Bousfield, 1959). It is proposed that the validity of FOK in predicting future memory performance stems directly from the validity of the partial information recollected. Assuming that FOR judgments rest on the mere *amount* of partial information retrieved, it can be shown that such judgments would tend to be valid as long as that information contains more correct than incorrect elements.

Indeed, the typical result with most free-report memory tests is that correct responses represent a much larger proportion of the total number of responses reported than incorrect responses (see Koriat & Goldsmith, 1993). This is also true of the partial information retrieved. This, of course, derives from a fundamental property of memory, that an item that has been committed to memory is more likely to give rise to correct than to incorrect (full or partial) reports. Under such conditions, a monitoring mechanism that relies on the mere accessibility of information is bound to be predictive of subsequent recall or recognition performance, because most of that information is correct. Of course, there are "deceptive" items that tend to produce more incorrect than correct responses (see Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, 1976; Nelson et al., 1984), and these may result in an unwarranted feeling of knowing (Koriat & Lieblich, 1977). However, these (perhaps, like the McKellar-Wason example) are the exception, not the rule.

This brings us to the question of the *inaccuracy* of feeling of knowing. The trace-access account implies that FOK judgments would be highly accurate in predicting recognition performance. However, the correlations reported in the literature, although generally positive, are low to moderate in size (Nelson & Narens, 1990). Therefore we must examine the conditions that contribute to FOK's *inaccuracy.* These can be derived from figure 6.2. As sketched in this figure, the validity of FOK in predicting subsequent memory performance depends on the correlation between (1) the *quantity* of information accessible at time *tl* and (2) the *accuracy* of memory performance (e.g., recognition) at time *12.* Thus, there are two factors that should contribute to the inaccuracy of FOK: the discrepancy in the *property* concerned (accessibility vs. accuracy) and the *time lag.*

Consider the first factor. As noted above, the accuracy of FOK judgments depends largely on the *correctness* of the partial information retrieved. Therefore, *monitoring accuracy* should be intimately tied to *memory accuracy,* so that conditions that improve memory accuracy should tend to enhance monitoring accuracy (Carrol & Nelson, 1993; Lupker, Harbluk, & Patrick, 1991; Nelson & Narens, 1990). Note that what matters according to the present formulation is not how many of the partial attributes of the target are recalled, but how many attributes recalled are correct (in the terminology of Koriat & Goldsmith, 1993, these two indices correspond to input-bound and output-bound measures of memory performance, respectively). Indeed, the analysis of memory pointer (word definitions) reported by Koriat and Lieblich (1977) supports this contention. This analysis was motivated by the observation that the exact same memory pointers tended to precipitate a TOT state in many subjects. Therefore, it seemed important to investigate the nature of these pointers. In their study, subjects were presented with word definitions, and were asked to recall the corresponding word. The word definitions were then classified in terms of the memory states that they tended to precipitate. These memory states (e.g., "don't know," "know-incorrect," "TOT-got it-correct") were defined in terms of both subjective and objective indices of knowing. Some of the word definitions were found to consistently elicit *accurate* positive or negative feelings of knowing across all subject. Examination of these definitions indicated that they typically provided an articulate specification of the target through a set of converging operations that allowed the search process to zero in on the target (or on the memory region where it resides). Such "focused" memory pointers, then, induce selective tuning to the correct target, resulting in a larger ratio of correct to

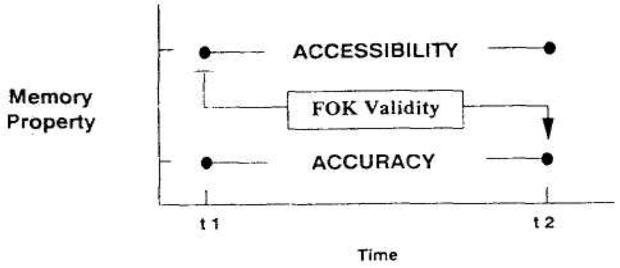

Figure 6.2
A conceptual framework for the analysis of FOK accuracy and inaccuracy.

incorrect partial clues. Therefore, they allow subjects to know that they know the answer when they actually know it, and to know that they do not know, when the target is not available to them.

Other memory pointers, in contrast, tend to produce a wealth of partial clues early in the search process, many of which are incorrect. This may occur either because the word definition itself is not focused or specific enough, or because the lexical entry corresponding to the solicited target is difficult to single out from other potential candidates. With such memory pointers the high accessibility of information does not guarantee the subsequent recall or recognition of the correct target. Therefore, such pointers tend to foster a false positive feeling of knowing.

With regard to the effects of *time lag*, one source of inaccurate FOKs derives from the systematic changes that occur over time in the amount and kind of information accessed. The search for a solicited memory target apparently begins with a rapid, shallow analysis of the question or word definition (see Reder & Ritter, 1992), which gives rise to a diffuse, nondeliberate summoning of pertinent clues from a broad memory region (see Kohn, Wingfield, Menn, Goodglass, Gleason, & Hyde, 1987). Gradually the search becomes more focused and controlled, and entails a more detailed evaluation of the information retrieved. These systematic differences between the information that comes to mind when memory is first queried and that which is ultimately used to support target retrieval will generally contribute to FOK's inaccuracy. In the analysis of Koriat and Lieblich (1977), pointers that resulted in a discrepancy between knowing and feeling of knowing were typically of two types, those that activated rich associations early in the search process, which later proved ineffective in supporting retrieval, and those that brought to mind few associations initially, followed later by a spontaneous retrieval of the answer.

Consider the former first. Because the initial inspection of memory covers a broad region, some of the clues that come to mind originate from misleading "interlopers" in the entire region. Such clues are difficult to discard because their source cannot be specified (unless the "interloper" itself— like "Wason" — is retrieved and identified). Therefore, their accessibility inflates preliminary FOK, even if the

correct target is eventually recognized or retrieved. Thus, a critical determinant of FOK accuracy is the "density" of memory entries in the broad memory region initially inspected. Indeed, on the basis of their analysis of word definitions, Koriat and Lieblich (1977) concluded:

> The presence of responses which approximately satisfy the definition seems to raise the rate of false positive feeling of knowing even when the correct target is zeroed in on. This latter effect may suggest that the preliminary analysis of the definition involves a cursory inspection of a broader region of memory including many entries, some of which satisfy the definition only grossly. The ease with which entries from this region come to mind then effects the estimate that the correct target will be found, (p. 161).

Interestingly, a false positive feeling of knowing was also precipitated by short definitions, as well as by the presence of redundant information in the word definition. Both of these were seen to affect FOK through the same mechanism mentioned above — by facilitating the emergence into mind of likely candidates during the stage of preliminary analysis.

In contrast, other memory pointers tend to be associated with a positively accelerated rate of information accrual, resulting in a false preliminary "don't know" response. Such pointers induce a search process similar to that involved in solving insight problems (see Metcalfe. 1986a; Metcalfe & Wiebe, 1987): The information does not accumulate gradually, but rather the answer appears to pop up suddenly, sometimes because of a spontaneous restructuring or paraphrasing of the question (see Koriat &: Lieblich, 1977). Such word pointers ma)' lead to the peculiar sequence of events characteristic of the *"don't know-got it-correct"* state (Koriat & Lieblich, 1974).

In general, then, FOK is assumed to be computed and updated on-line according to the information accessible at that point in time. However its accuracy will depend on the correlation between (1) the accessibility of information at the time of soliciting FOK judgment, and (2) the accuracy of memory performance at the time of administering the criterion test (e.g., recognition). Systematic differences that are due to memory property (accessibility vs. accuracy) and time lag may contribute to the impression that monitoring and retrieval are dissociable, independent processes.

One implication of the results of the analysis of memory pointers is that characteristics of the *question* (e.g., the amount and kind of initial activations it precipitates) may sometimes be more critical for preliminary FOK judgments than the recallability of the *answer*. This implication is consistent with that which derives from the cue familiarity hypothesis (Metcalfe et al., 1993; Reder, 1987; 1988, Reder & Ritter, 1992; Schwartz & Metcalfe, 1992).

An Accessibility Model of FOK and Some Empirical Evidence

In the present section I shall briefly sketch a process model of the feeling of knowing, and present some illustrative results of experiments designed to test it (figure 6.3). The model and the experimental work are described in detail elsewhere (Koriat, 1993), and here only a brief report will be included.

The model assumes that when searching memory for a solicited target a variety of clues come to mind. Some of these emanate from the target proper and represent "correct partial information," while others represent "wrong partial information" that may stem from a

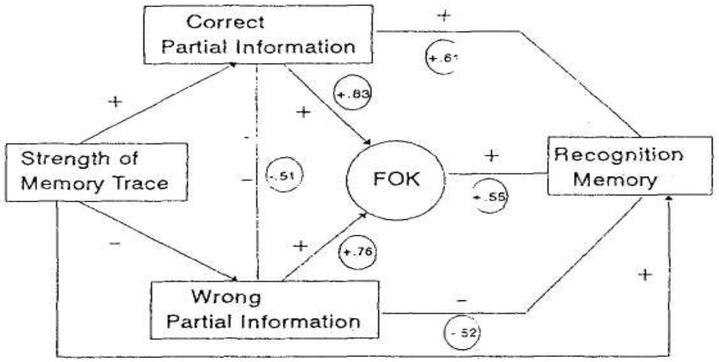

Figure 6.3
An accessibility model of the feeling of knowing. The positive and negative correlations postulated by the model are denoted by plus and minus signs, respectively. The correlations marked within circles are based on the results of experiment 1 of Koriat (1993)

variety of sources. In general, the higher the memory strength of the target, the more likely it is to give rise to correct partial or full recall, as well as to accurate recognition. In contrast, the stronger the memory trace, the lesser 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.

Turning next to the *feeling* of knowing, the core assumption of the model is that FOK depends on the accessibility of partial information *regardless of its correctness.* Accessibility includes two factors, the *amount* of information retrieved, as well as its *intensity* (its ease of access, its persistence, etc.). FOK is assumed to increase with increasing accessibility of *both* correct as well as incorrect partial information. It is important to stress that the distinction between these two components is assumed *not* to be directly available to the subject, i.e., subjects cannot directly monitor the *accuracy* of the partial information that comes to mind. Therefore what matters is only the overall accessibility of information.

The pattern of relationships noted above between partial information, FOK, and recognition memory implies that the dependence of FOK on the accessibility of *correct* partial information is responsible for its *success* in predicting correct recognition, whereas its dependence on the accessibility of wrong partial information is responsible for its *inaccuracy.*

This pattern raises the question of why FOK is nevertheless generally *accurate* in predicting recognition? There are two main reasons for that. First, as noted earlier, under most common conditions, the partial or full information that comes to mind is more likely to be correct than incorrect. Therefore, correct partial information tends to constitute the largest portion of the total amount of accessible information, and to account for the bulk of its variance. The overall result is that of a *positive* correlation between the *total amount* of accessible information and recognition memory.

The second reason has to do with the *intensity* of the information recalled. Not only does a memory target tend to give rise to more correct than incorrect partial clues, but also correct clues tend to emerge into consciousness with a greater *intensity*. Therefore, although subjects may not be able to monitor *directly* the accuracy of the partial information retrieved, they can do so *indirectly on* the basis of its intensity. An important intensity cue that is utilized by subjects is the *ease with* which information comes to mind (see Jacoby, Kelley, & Dywan, 1989; Jacoby & Kelley, 1991; Jacoby, Lindsay, & Toth, 1992), and this cue can be expected to contribute to both the FOK as well as its accuracy.

I shall now present some illustrative results from experiment 1 of Koriat (1993), which is a modification of that employed by Blake (1973; see also Hart, 1967a). In each trial, subjects memorized a four-letter string (e.g., *TLBN).* They were then presented with a filler task for 18 seconds, and were then 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. Thirty subjects participated in the experiment, and each was presented with 40 such trials.

A methodological note is in order. Although this procedure generally conforms to the recall-judgment-recognition paradigm (Hart, 1965a) that has been typically used in most FOK studies, some of its unique features should be noted, because they are critical for the accessibility account of FOK. First, unlike most previous studies, FOK judgments were always solicited here regardless of the subject's performance on the initial recall test. The common practice of soliciting FOK judgments only when the subject's answer is incorrect (or when the subject fails to produce any response) reflects the assumption that the subject has direct access to the correctness of his/her answer. From the point of view of the accessibility model, however, it would seem odd to eliminate from the study of FOK all of the subject's responses which the *experimenter* knows are right. Second, unlike some of the previous studies that tested the partial-recall hypothesis (Blake, 1973; Eysenck, 1979; Schacter & Worling, 1985), where partial knowledge was assessed through a forced-report procedure, here subjects were allowed the option to report as many letters as they could remember. This was necessary to allow assessment of the *amount* of partial information accessible to them. When a forced-choice procedure is used, it is the *experimenter* who must determine how many of the letters reported by the subject are *correct,* and it is not clear at all that that information is accessible to the subject. In fact, a finding that FOK ratings rest on the number of correct letters retrieved provides little insight into the basis of FOK judgments, because it leaves us with a no less intriguing question: How does the subject know how much he or she knows?

The procedure described above allows evaluation of some of the predictions of the model pertaining to the amount of partial information retrieved. Figure 6.3 also includes (in circles) the estimated correlations between some of the components of the model. These estimates were derived from the empirical data using complex procedures that will not be described here (see Koriat, 1993). Note that correct partial information was defined in terms of the number of correct letters reported by the subject, whereas wrong partial information was defined in terms of the number of incorrect letters reported. Each of these could range from 0 to 4, with their sum never exceeding 4. It can be seen that the correlational pattern conforms to the model. Notably, FOK increased as a function of increasing number of correct letters recalled (+ .83), but it also increased with increasing number of *wrong* letters accessed (+ .76). While the former was positively correlated with recognition (+ .61), the latter was negatively correlated (-.52). Thus, it would seem that the number of correct letters retrieved should contribute to the *accuracy* of FOK, whereas the number of incorrect letters should contribute to its *inaccuracy.*

However, despite the conflicting contributions of correct and wrong partial recalls to the validity of FOK, the overall correlation between FOK and recognition was positive (+ .55; figure 6.3). The reason is that the great majority (89%) of the letters recalled were correct. Therefore the mere number of letters recalled is a sufficiently good predictor of recognition memory even if subjects cannot tell correct from incorrect recalls.

If monitoring effectiveness derives from the effective retrieval of correct partial information, then subjects exhibiting better memory accuracy should also evidence better metamemory (see Lichtenstein & Fischhoff, 1977). Indeed, when subjects were divided in terms of their overall recognition memory performance into a High-Recognition and a Low-Recognition group, the average correlation between FOK and recognition performance was significantly higher

(+ .67) for the former group than for the latter (+ .40). Examination of the partial recall performance of the two groups explained why: high-recognition subjects produced a higher proportion of correct to incorrect letters than the low-recognition subjects, and this was probably responsible for the higher validity of their FOK judgments.

The results presented above support the claim that the predictive validity of FOK derives solely from the diagnostic value of total partial information. If such is indeed the case, then the latter should be no less predictive of recognition memory than the subject's own feeling of knowing. 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, the feeling that one "knows" the target, was not any more diagnostic of the "availability" of the solicited target than the mere amount of information accessed. This implies that subjects' monitoring responses do not have privileged access to information that is not already contained in the output of the retrieval attempt.

Additional results (experiment 2; Koriat, 1993) indicated that subjects can further improve their monitoring by taking into account factors having to do with the *intensity* of the partial information retrieved. When ease of access was indexed by the latency of recalling the letters of the target, it was found, first, that ease of access is diagnostic of the *correctness of* the information retrieved. That is, recall latency was shorter for correct than for incorrect partial recalls, even when the total number of letters recalled was held constant. Second, FOK judgments increased with increasing ease of access, suggesting that the feeling of knowing rests not only on the amount of partial information recalled, but also on its ease of access. Thus, reliance on ease of access can also contribute to FOK validity in predicting recognition.

In conclusion, the present chapter contrasted the trace-access model of FOK with the accessibility model. The former model postulates a special monitoring mechanism that taps directly the presence in memory of an unrecallable target. This mechanism provides for the validity of FOK judgments. The accessibility account, in contrast, denies the necessity of invoking such a mechanism, and shows

how both the accuracy and inaccuracy of FOK judgments can be explained by assuming that FOK judgments merely monitor the overall accessibility of partial information regarding the target in question.

Acknowledgments

The research reported in this chapter was supported by the Israel Science Foundation administered by the Israel Academy of Sciences and Humanities. The experiments were conducted at the Institute of Information Processing and Decision Making, University of Haifa.

Metacognition

Knowing about Knowing

edited by Janet Metcalfe and Arthur P. Shimamura

A Bradford Book The MIT
Press Cambridge,
Massachusetts London,
England

This book was set in Baskerville by DEKR Corporation and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Metacognition : knowing about knowing / edited by Janet Metcalfe and
Arthur P. Shimamura. p.
cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-13298-2
1. Metacognition. 2. Cognition. I. Metcalfe, Janet. II.
Shimamura, Arthur P. BF311.M4489 1994
153—dc20 93-37597
CIP