Feelings and Judgments of Knowing: Is There a Special Noetic State?

Janet Metcalfe

Department of Psychology, Columbia University, New York, New York 10027
E-mail: metcalfe@columbia.edu

A. Koriat distinguishes between feeling-based and inferentially based feeling-of-knowing judgments. The former are attributable to partial information that is activated in implicit memory but not fully articulated. They are not, however, attributable to direct access to the target—an hypothesis that Koriat specifically repudiates. While there is considerable merit in the distinction that Koriat draws, and his emphasis on the possibility that people base at least some of their metacognitive judgments on implicit information seems well founded, it is argued that his rejection of the direct access view is premature. There may be a state—a true noetic state—in which people actually know the answer before they are able to express it. A case is made for further consideration of the scientific merits of the direct-access view of the noetic feelings people experience in imminent tip-of-the-tongue (TOT) states.

In his article entitled “The Feeling of Knowing: Some Metatheoretical Implications for Consciousness and Control,” Koriat proposes a distinction between inference-based and experience-based judgments of knowing, and he underlines the importance of implicit-memory feelings that contribute to experience-based judgments. He specifically argues, though, that the feeling state he describes is not one in which the person has direct access to target information. His reasons for rejecting the hypothesis that a state of direct access exists, and some reasons why it might, nevertheless, be worth considering further, form the core of this Commentary. First, though, Koriat’s distinction between experiential- and informational-based metacognitions will be discussed.

The crux of the distinction is that information-based judgments, in Koriat’s framework, are those that are based on expressible partial information—for example, “I know there are only a few words that begin with the letter X, therefore this is likely to be an easy choice, and I’ll give it a high feeling-of-knowing rating.” Experience-based judgments are based on information clues as well, but they are less well articulated—for example, “there’s something, but I’m not sure what (perhaps some aspect of it was presented to me in the previous part of the experiment?), that makes this item stand out, therefore I’ll give it a high feeling-of-knowing rating.” The person is able to articulate what causes the judgment in the informationally based case, but not in the experientially based case. The inability to say what is contributing to the judgments makes the inferences a feeling, in Koriat’s framework.

Commentary on A. Koriat (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. Consciousness and Cognition, 9(2), 149–171. This article is part of a special issue of this journal on Metacognition and Consciousness, with Thomas O. Nelson and Georges Rey as Guest Editors.

1053-8100/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.
Jacoby, and others, have made the case that implicit information may contribute to judgments such as fame (Jacoby, Kelley, Brown, & Jasechko, 1989), likeability, and social judgments (Bargh & Pietromonoco, 1982), and Koriat’s case seems compelling: sometimes this kind of information is used when people are asked metacognitive judgments as well (Benjamin & Bjork, 1996). Indeed, studies such as those based on cue-familiarity (Metcalfe, Schwartz, & Joaquim, 1993; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992), in which activation of the cue produces increases in people’s feelings that they will know the target, even though the cue activation in no way increases the probability of target retrieval, lend support to this view, just as Koriat suggests. My reservation about the idea is not that I deny that implicit information can contribute to feelings of knowing, it is rather that I have slowly become convinced that these are not the only feeling states that are of importance in these judgments, and that, in some cases, people have access to the target itself. Direct access may underlie a feeling seen when a person says he or she is in the state of imminence, at least some of the time.

Although Koriat emphasizes the role of a noetic feeling state (by which one might infer that the person actually knows, and thus, has direct access), his sheer phenomenological “feeling” is an instance of implicit-memory fluency, that, under other circumstances might be interpreted in many different ways. This inferential feeling of fluency may not always be the same as the tormenting tip-of-the-tongue feeling described by William James and others, that seems specific to a feeling of imminent retrieval. The distinction between a feeling state and an inferential state almost disappears when Koriat articulates the basis for feeling-based judgments within his accessibility framework: “In a TOT state, for example, the accessibility of partial information is what precipitates a sheer feeling of knowing, regardless of the source or correctness of the partial information accessed,” or “rather than reflecting direct access to memory traces, noetic feelings are based on inferential heuristics that operate implicitly and unintentionally.” Thus, the feelings turn out to be inferential; their only difference from other inferences is that the person cannot articulate the partial information on which they are based, and so they are considered to stem from implicit memory, or not to have reached the level of consciousness. The lack of consciousness of the background reasons for one’s judgments is crucial in defining the feeling states as opposed to the inferential states, in Koriat’s view, and it follows that they are no longer feeling states once one knows their whys and wherefores.

But, Koriat’s definition of feelings—as partial information below the threshold of consciousness—is not a necessary one. A person could have a feeling of fear or pain or desire—to use classic examples—and be both perfectly conscious of it and able to articulate its whys and wherefores. Consciousness need not stop a person from feeling the pain or fear or desire (though it might allow a person to utilize control processes otherwise unavailable to curb or, in some cases, augment the feelings). A feeling, by this view, is not a feeling because it is weak, subliminal, or inarticulate. Furthermore, being conscious of a feeling does not mean that one must be able to attach a verbal label to it: One can consciously feel pain without saying the word. In addition, feelings, perceptions, judgments, cognitions, and actions are all comprised of subcomponents that are unconscious, inexpressible, and inarticulate, and so this property does not seem either necessary or sufficient to define a feeling. Fi-
nally, Koriat calls his metacognitive feeling states ‘noetic states,’ but it would appear that the so-called ‘noetic’ aspect is really anoetic, since the person does not really know and rather is only making an inference gleaned from implicit information.

DIRECT ACCESS AND KNOWLEDGE OF THE ULTIMATE TRUTH

Koriat states: ‘In the direct-access view, the feeling of knowing is granted a special status, having privileged access to stored information that cannot yet be accessed. Hence, the validity of the feeling of knowing is taken for granted, needing no justification. Traditionally, the notion of ‘self-evidence’ has occupied a central role in intuitionistic theories, which assume that some ultimate truths are directly or intuitively apprehended . . . A direct access view, which takes the validity of the FOK for granted, would assume that the FOK continues to monitor the correct memory target despite the fact that the partial information that a person succeeds to access actually derives from a different source. This implies, in a sense, that metamemory can circumvent the output of memory: It has privileged access to stored knowledge beyond what is accessible to memory’ (my italics on ‘ultimate truth,’ Koriat’s italics on ‘correct’).

It is indisputable that metaknowledge has no privileged access to the Ultimate Truth. However, in contrast to Koriat’s argument against the direct access view, which rests on this characterization of that view, one can hold a direct access view without making this claim. The criterion for determining whether or not a lexical item has been accessed already, at the time the metacognitive judgments is made, is not whether it ultimately proves to be True. No judgment—even if the information was as known, overt, conscious, and accessible as it is possible for knowledge to be within a human frame of reference—could survive the Ultimate Truth test. The criterion for direct access must instead be whether, when the rememberer finally does articulate the lexical entity or it is made known to him or her, it is the same entity that was in mind earlier, before overt expression. Can the rememberer know in advance, with above chance accuracy, whether the rememberer him or herself would call correct the entity that he or she has in mind but cannot yet articulate. (There is also the intriguing possibility that one might have direct access to something that one knows is wrong. For example, I might be thinking of the person who was the person I always confuse with the person who was in Leaving Las Vegas. In this case, the question is not even whether the answer to the question is right—it is only whether it corresponds to the target that I have in mind.) To take the direct access view seriously, then, we must consider subjective not objective truth. Do we know, before we can articulate it, that the solution we are about to articulate, when finally our articulatory programs kick in, will be the solution we are after? I do not think that it is a logical impossibility that we could know it, nor does it have to imply any knowledge of Ultimate Truths, though the feeling state itself might be such that we have a tendency to think that it is true.

A number of linguistics researchers have proposed variants of a dual-stage model of lexical access in which there ‘‘is a moment at which a semantically and syntactically specified lexical representation has been accessed but not its phonological content. This claim about the nature of lexical access is as close to a universally shared position as anything is in cognitive science’’ (Caramazza & Miozzo, 1997, p. 329).
Indeed, some models of lexical access describe a conceptual level at which concepts like ‘tiger’ (with properties such as stripes, wild, animal, and the visual form of the tiger) are at a different level of analysis from the lemma or lexical representation level, which is yet again different from the lexeme level, which then allows selection of phonological segments. Activation in these models sometimes occurs in serial, so that activation at the lemma level, for example, comes before activation at the lexeme level, but sometimes they cascade (Dell, 1986). In models that have such levels of analysis, it is possible that an individual might have excellent knowledge at the semantic level—thus, really knowing what and that they know something, and exhibiting true noetic feelings—but some confusion or breakdown might exist at the phonological levels, or in the connections between one level and another. Even though people truly know, then, they may sometimes be unable to express that knowledge. The model of Burke, Mackay, Worthley, and Wade (1991) applied well to TOT states in older people, and since that time research into the characteristics of anomic patients has suggested more refined variants of the dual stage model. It seems reasonable to suppose that dissociations across levels could occur, even to younger people, some of the time. Indeed, Schwartz (submitted) showed in a diary study that college students exhibited TOT states about 4.7 times in a four-week period.

The (appropriate) complexity of some of the lexical access models leads one to wonder exactly what kind of activation and how much of it is necessary to qualify as direct access? Do all of the features represented at the conceptual level or the semantic (lemma) level have to be activated, for example, and if not, how many are enough? Do all of the features at all of the levels have to be lit up (as seems to be Koriat’s contention)? If one argues, as it appears that Koriat is, that the person must have actually overtly expressed the word, for it to count as having been accessed, do we discount those items that may be pronounced incorrectly as not having been accessed? What about a word that is lisped? It might be possible to say that a person has not really accessed (or directly accessed) a semantic representation unless the entire representation—with all of its features in a perfect form—has been accessed, but we know that the meaning of words varies according to the context in which they are presented and that different features will be highlighted under different conditions of study, and so it would seem that such canonical semantic access to all of the features in a representation is an unattainable ideal. Alternatively, we might take exact reproduction of the features of the event as originally experienced as necessary. But this is an unactualizable ideal. If nothing less than full access, however defined, counts as access, then we are left in the position that people never have access to traces—even when they have overtly recalled the target.

Koriat also claims that people have no knowledge of whether they are right or wrong—they only know what they retrieve from memory and must make the assumption that what is retrieved is correct. In contrast to this hypothesis, though, are the many cases in which people know perfectly well that what they have retrieved is incorrect—as if they had some prototype, or set of constraints that defined correctness, a priori, and knew when the retrieved item was a mismatch. It is almost as if people have a lock which allows them to tell whether or not a particular key is the right one or not. People often retrieve false answers, and then immediately say, covertly or sometimes overtly, “No, that’s not it.” Gehring, Goss, Coles, Meyer, and
Donchin (1993) have even discovered that there is an event-related potential (called the ERN or error-related negativity) that specifically marks the production of a mistake. As Smith (1994) has pointed out, when people are in the TOT state they are sometimes plagued by interlopers which block the correct answer. The point of interest is that people know that the items they are retrieving are interlopers—they do not, typically, mistake them for being correct. When they finally get the right answer, they know that too. The direct access or true noetic view of TOTs says that people tend to be accurate about getting the answer when they are in the state of imminence because in a sense, at least some of the time, they have already gotten the answer. The node is lit up.

One might ask why TOTs are accurate according to Koriat’s accessibility framework? He answers by saying that it is because memory is accurate—begging the question of how it is that (in solving a crossword puzzle, say) we can retrieve any number of wrong answers and know that they are wrong, even though they, too, are retrieved from memory. When we finally retrieve the right answer, somehow (but in a way unexplained by the accessibility account) we know that it is right. In contrast, all of the wrong answers, regardless of how fluently, easily, or persistently they spring to mind, or, indeed, how many of them so flood in upon us, result in metaknowledge not that we know (as the accessibility view would posit) but rather that we do not, as has been experimentally demonstrated by Schreiber and Nelson (1998), Schreiber (1998), and by Maki (1999).

THE FEELING STATE OF KNOWING

I suggest that there may, actually, be two related feeling states associated with knowing, both of which could properly be called noetic feelings. The first is a special feeling state in which the person has a strong intuitive belief that the sought-after information, problem solution, or memory, is known and that it is correct, but it cannot yet be articulated: the tip-of-the-tongue state or a feeling of imminence. The second state is one in which the person retrieves a memory or achieves a solution and has a strong conviction, stemming from a feeling state, that the information retrieved is correct and connected to the self. It clicks. This feeling state is associated with what people have often called explicit memory.

i. TOT’s and Imminence

The first feeling state is the ‘‘gut’’ feeling that people have when they ‘‘know’’ they have an answer, even though they may not yet be able to express it so another person can either understand or verify it. This feeling appears to have compelling motivational implications. It has been referred to by various names in the literature. In common language this feeling is sometimes called ‘‘intuition.’’ For William James (1891, first published, 1890) this peculiar state of consciousness was the intensely active ‘‘gap’’ which beckoned and plagued a person until he or she finally found the answer. For Polanyi (1962) it was the perception of a ‘‘hidden potentiality,’’ or, in the scientist gifted enough to have it, the capability of sensing the presence of as yet unrevealed logical relations or unknown solutions. This ability to see what cannot yet be articulated and proven, with a certainty that it will be correct once verified, is, in Polanyi’s studies of great scientific discoveries, one of the hallmarks of genius.
It is often called simply the tip-of-the-tongue (TOT) state, or sometimes the feeling of imminence (Smith, 1994), though recently Schwartz (submitted) has shown that when participants are asked to separate TOT states into those which indicate imminence, they can do so, and this latter feeling is a particularly strong predictor of impending resolution. “Theoretically, this correlation can be taken as evidence of direct access mechanisms of TOT etiology.” This correlation occurred in both an experimental study and in a diary study, with both showing that resolution was over 90% when people claimed to be in such a state. Other TOT-related feelings, like emotionality (which may indeed stem from an implicit inferential mechanism as described by Koriat and others) was positively correlated with resolution in Schwartz et al.’s (in press) experimental study but negatively correlated in Schwartz’ (submitted) diary study, and so it seems to be a poor predictor of retrieval success. Imminence may be specifically related to impending retrieval, not simply an implicit memory derived inference.

Funnell, Metcalfe, and Tsapkini’s (1996) studies of an anomic patient, H.W., convey something of the flavor of the state of imminence. When given sentence frames that specified unique nouns, verbs, and adjectives, for example, “One ______ a turkey by brushing butter on it while it is cooking,” H.W. was able to produce only one out of 300 so-specified words. However, he would frequently give a maximally high feeling of knowing rating indicating (correctly) that he would be able to choose the right answer; he told us that he knew the word; and, when given an eight-alternative forced choice test, he immediately chose the right answer. H.W. was nearly always right, when he expressed this kind of confidence (although he also noted that there were certain test conditions—like when alternatives had first letters in common that might cause his performance to deteriorate, and that there were certain words that he did not know, or could only guess at, and those he assigned a lower rating).

Taped interviews that Funnell conducted with H.W. also suggested that he had access to the lexical representation of many items when he was in this state, even though he could not say the words. Koriat argues that a person in this state does not know the answer—they have no direct access—but it seemed that H.W. did know the answer but was simply unable to articulate it. So, for example, when given the definition of a helicopter, H.W. could make the motions of a helicopter indicating very clearly what it did; he could tell us how it worked; he could even tell us that the last name of the person who had invented it started with an S (and showed pleasure when the interviewer knew that the inventor was Igor Sikorsky). Similarly, when describing what his company made, H.W. could use the words that he could articulate, as well as many gestures and air drawings, to get the interviewer to produce the term “oxygen” (at which point he exclaimed that yes! that was what he was after!) It was as if he was in the state of consciousness that you or I would be in if we were playing charades, except that he was unable to produce the word rather than disallowed from so doing by the rules of the game. Miozzo and Caramazza’s (1997) patient Dante, and other anomics, show a similar pattern.

**ii. Noetic Feeling Accompanying Recognition**

Koriat (1999) suggested that the quality that distinguishes explicit from implicit memory is that fact that with explicit memory—essentially by definition—people
not only know something, they also know that they know it. I suggest that it is the feeling state that accompanies explicit knowledge, more than even the retrieval of the time, place, circumstances, and sensory details of the original experience, that assures us that the event is a real memory of something that actually occurred to us. I have suggested (Metcalfe, 1994) that this feeling state (which, of course, may be bolstered by knowledge of the time, place, and circumstances) is itself the crucial component in our explicit certainty about and personal attachment to our memories. If this is a separable feeling state, then one might find cases of patients exhibiting overt disruptions in it. With ‘normal’ amnesia, though, both the content of the memories and the noetic feeling are usually impaired, so one cannot distinguish the feeling as a separate component. However, Stuss (1991) has reported a patient who seems to have relatively replete memorial information but an impairment in noetic feeling. If noetic feelings were completely derivative upon retrieved information, as in Koriat’s theory, then the dissociation shown by this patient should not be observable.

Stuss’s (1991) patient suffered from frontal lobe injury and was hospitalized for about 10 months. Upon dismissal from the hospital he was picked up by his wife. Although he had a great deal of memorial information about his wife and children, and, indeed, often noted how very similar the woman who picked him up from the hospital was to his wife—her looks, her height, hair color, age, etc., indeed, she was virtually identical to his wife—her actual presence at the end of his period of hospitalization apparently failed to evoke the appropriate noetic feeling. This feeling seems to provide the glue that fuses one episode with a person to another, allowing for the person’s perceived continuity of identity that we normally take for granted. So, for example, when one sees one’s child, after having been separated for summer camp, say, the noetic feeling that this person—though perhaps looking more in shape, and slightly older—is one’s child, allows us to weave events with that person into a coherent, seamless experience of the child as one person, seen repeatedly, rather than as multiple people who happen to look very similar to one another. Lacking this noetic feeling, Stuss’s patient came to believe that he had two wives and two families (eight children in all): the one that presented itself at the hospital and with whom he now lives, and the one that existed before and which was very very similar to his new family.

On the other hand, when a person experiences this noetic feeling—even if it is unjustified in the opinion of the world—he or she may tend to believe in the truth of the conclusion drawn. The feeling may indeed presage the person’s subjective belief, even, in some cases, when that belief flies in the face of what is objectively accepted as being true. A particularly poignant account in which the particular quality of noetic feeling was taken to indicate truth is given in a recent biography by Sylvia Nasar of John Nash—1994 Nobel Laureate in Economics. Nash is a brilliant mathematician, most famous for his work in cooperative game theory, who was able to solve problems that were intractable to lesser minds. Unfortunately, at around the age of 30, he succumbed to paranoid schizophrenia. In her prologue, Nasar describes a scene in which Nash, hospitalized, was visited by his Harvard professor friend and colleague George Mackey. ‘‘[T]he visitor sat upright, oppressed by the silence, acutely conscious that the doors were locked. Mackey finally could contain himself no longer. His voice was slightly querulous, but he strained to be gentle. ‘How could
you,’’ began Mackey, ‘‘how could you, a mathematician, a man devoted to reason
and logical proof . . . how could you believe that extraterrestrials are sending you
messages? How could you believe that you are being recruited by aliens from outer
space to save the world? How could you . . . ?’’

Nash looked up at last and fixed Mackey with an unblinking stare as cool and
dispassionate as that of any bird or snake. ‘‘Because,’’ Nash said slowly in his softly,
reasonable southern drawl, as if talking to himself, ‘‘the ideas I had about supernatural
beings came to me in the same way that my mathematical ideas did. So I took them
seriously.’’ (p. 11).

CONCLUSION

When I first proposed a model of the locus of feeling of knowing judgments (Metal-
calf, 1993) it seemed to me that these judgments were based on information, such
as cue-familiarity, that need not imply access to the target, a position compatible with
that outlined by Koriat. Certainly, irrelevant information and implicit information can
influence these metacognitive judgments, and by now it is beyond question that direct
access is not the only cause of feelings of knowing. Until recently, I thought that
when a person has just demonstrated that they could not recall the target, the notion
that they might, nevertheless, have direct access to the target trace and base their
metacognitive judgments on that access was a mystical hypothesis, and could be
dismissed out of hand. If people have just shown that they cannot retrieve the target
the judgments must be based on something else. What convinced me otherwise, was
talking with and testing an anomic patient (H.W), considering the very reasonable
and not at all mystical models of lexical access, considering the still scant phenome-
nological data on the state of imminence, and carefully considering the necessary
logical assumptions (which turn out to not involve magical assumptions about intu-
tive access to Ultimate Truths). I find it difficult to resist the conclusion, at least
with our own patient, H.W., that he really does have access to the semantic node
that he is seeking, but is unable to express the word—rather like a normal person
playing a game of charades, except that his lack of expression of the word is not
voluntary. If there is a direct access state—a true noetic state—that precedes actual
articulation, in some cases, and accompanies recall, in others, that state may have
special motivational status, may elicit certain kinds of beliefs in the person, may
have special neuropsychological consequences and markers, and may evoke special
feelings in the person, all of which warrant further intensive investigation.

ACKNOWLEDGMENTS

Preparation of this article was supported by NIMH Grant MH48066. I thank Bennett Schwartz, Steven
Smith, and Michele Miozzo for discussions of some of the issues raised in this Commentary.

REFERENCE


