In this rejoinder we clarify several issues raised by the commentators with the hope of resolving some disagreements. In particular, we address the distinction between information-based and experience-based metacognitive judgments and the idea that memory monitoring may be mediated by direct access to internal representations. We then examine the possibility of unconscious metacognitive processes and expand on the critical role that conscious metacognitive feelings play in mediating between unconscious activations and explicit-controlled action. Finally, several open questions are articulated for further scrutiny.

The aim of this rejoinder is twofold. First, to clarify several points, hoping to resolve some of the disagreements between the position presented in the target article and the arguments contained in the replies. Second, to address the arguments advanced by the commentators. Some of these arguments raise important theoretical and metatheoretical issues that invite further scrutiny.

INFORMATION-BASED VERSUS EXPERIENCE-BASED JUDGMENTS

The distinction outlined in the target article as well as in Koriat and Levy-Sadot (1999) between experience-based and information-based metacognitive judgments is slightly different from what Janet Metcalfe describes. This distinction is similar to that made in other contexts by Kelley and Jacoby (1996b), Strack (1992), Epstein and Pacini (1999), and others. A common theme in these discussions (see Schwarz, 1990; Schwarz and Clore, 1996) is that subjective feelings serve an informational function, sometimes providing the basis for judgments and actions. Thus, subjective feelings are no mere epiphenomena, but play a causal role in guiding decisions and behavior. These discussions contrast judgments based on subjective experience with judgments based on beliefs and memories. Along the same line, we proposed a distinction between metacognitive judgments that are based directly on subjective experience (an immediate “gut feeling”) and those that are based on the retrieval and weighing of information from memory. When a person is asked to indicate how confident she is that she knows the name of a certain acquaintance, she may base her judgment on a sheer gut feeling, as happens often in strong tip-of-the-tongue (TOT) states, or on an explicit deduction from a variety of pieces of information.
retrieved from memory. In the former case we talk about experience-based metacognitive judgments. When feeling-of-knowing (FOK) judgments are information-based, they rely on the content of the information retrieved, resulting in an educated inference about the plausibility that the solicited target will be recalled or recognized in the future. As Koriat (1993) noted, such deliberate inferences about one’s own knowledge may not be different from those underlying the prediction of future events in general (e.g., which football team is likely to win the game).

The argument presented in the target article is that the sheer, ‘‘direct’’ experience of knowing is based on an unconscious inference, much like the feeling of familiarity, as discussed by Jacoby and his associates (Kelley & Jacoby, 1998; Jacoby & Whitehouse, 1989). In fact, according to proponents of the indirect view of perception, such as Helmholtz, Brunswick, and Rock (see Palmer, 1999), even perceptual experience like the perception of depth is based on an unconscious inference from a variety of cues. Indeed, experimental evidence indicates that perceptual illusions can be produced by an unconscious inference in which the enhanced ease of processing produced by priming is misattributed to perceptual dimensions of the stimulus (e.g., Whittlesea, Jacoby, & Girard, 1990; Mandler, Nakamura, & Van Zandt, 1987).

Metcalfe states that ‘‘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 . . . and it follows that they are no longer feeling states once one knows their whys and wherefores.’’ This is not what we had in mind, and we apologize for not being more articulate on this matter. Consider an example that we have used previously (Koriat & Levy-Sadot, 1999): A person who does not like tuna fish may feel some repulsion toward a salad offered in a buffet when she learns that it contains tuna fish. Her choice to avoid the salad may then be based on the explicit information gained (information-based action) or on the immediate repulsive feeling (experience-based action). In contrast, consider a second example in which a person simply feels immediate repulsion toward the salad without being able to explain why. In that case, the decision to avoid the salad can only be based on sheer subjective experience.

Clearly, in the first example, the fact that the person learns that the salad contains tuna fish does not prevent her from feeling repulsion toward the salad. Furthermore, even though the feeling itself in that case is information-driven, judgments and behavior may still be based directly on the feeling itself rather than on the explicit information. Recent research in social and cognitive psychology has attempted to specify the conditions under which judgments are more likely to be based on information and those in which they are more likely to be based directly on subjective experience (Kelley & Jacoby, 1996a; Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka, & Simons, 1991; Strack, 1992).

The second example, in which behavior is based on an unexplained feeling, is analogous, in our view, to that in which a problem-solver has the intuitive feeling of approaching the solution (Metcalf & Wiebe, 1987), or that in which a person attempting to retrieve a solicited target from memory feels that the recall of the target is imminent (Schwartz, Travis, Castro, & Smith, in press). We chose to focus on
this kind of experience-based judgment because it discloses the unique aspects of metacognition, serving to interface between unconscious and conscious processes.

THE ISSUE OF DIRECT ACCESS

Metcalf argues that the 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.”

The observations that Metcalfe cites in favor of this argument are precisely those that have motivated a direct-access account of the TOT and FOK phenomena. The TOT and FOK states have attracted the attention of researchers precisely because of the discrepancy they instantiate between objective and subjective indexes of knowing. In the TOT state, for example, the person may know precisely the identity of the person whose name he has to retrieve (as in Yarmey, 1973, when participants are probed with a photograph of a celebrity). Yet he is (a) momentarily incapable of retrieving his name and (b), at the same time, quite confident that he knows the name. This is similar to the situation in which a person is capable of accessing a lexical representation (lemma) but is incapable of accessing the phonological representation. In both cases it is appealing to argue that the person has “direct access” to the lexical representation.

However, the issue is whether the subjective feelings associated with knowing derive from the direct monitoring of the presence of the solicited target or from the unconscious application of some heuristic. Phenomenological reports cannot help decide between the two possibilities because in both cases the subjective experience is expected to be that of direct perception. The findings accumulated in recent years (Benjamin & Bjork, 1996; Koriat, 1997; Metcalfe, Schwartz, & Joaquim, 1993; Reder & Schunn, 1996) strongly support the possibility that metacognitive feelings are heuristically based. These findings, however, do not automatically rule out the possibility that such feelings may also be based on direct access to characteristics of the solicited target (e.g., trace strength). In fact, we cannot envision an experimental procedure that would allow us to reject the direct-access hypothesis outright. In general, our inclination is to put aside the “direct access” view as long as we can do without it because we feel that this view does not provide an answer to the question of how we know that we know, but simply postulates that we “just know” (see Block, 1995). In a sense, it frees us from the burden of explanation.

We agree with Metcalfe that the phenomena associated with recognition make the idea of direct access quite appealing: There is no question that people “know” to some degree whether the answer retrieved or chosen is correct or not, as has been indicated in many studies of subjective confidence (e.g., Koriat & Goldsmith, 1996). Furthermore, when we search for a name in our memory and finally find it, the phenomenal experience suggests that we know a priori which candidate name fits the answer. Indeed, William James’ (1893) description of the TOT state nicely conveys this idea: “There is a gap therein; but no mere gap . . . If wrong names are proposed to us, this singularly definite gap acts immediately so as to negate them. They do not fit into its mould. And the gap of one word does not feel like the gap of another”
(p. 251). On intuitive grounds it would appear then that we must postulate a kind of ‘direct access’ process that operates during retrieval and allows a person to recognize the sought-after target once it is retrieved. Without such process, memory search could go on ad infinitum.

However, in recent years there has been growing support of the idea that recognition itself may be based on subtle inferential heuristics. In the context of episodic memory research, Whittlesea (1993), for example, observed that priming words before they appeared in a recognition test increased the likelihood of their being falsely recognized as old. A similar effect was obtained by enhancing the visual clarity of words during a memory test (Whittlesea et al., 1990). Similarly, in the context of semantic memory, confidence in the correctness of an answer to general-information questions has been found to be sensitive to manipulations that affect retrieval fluency (Kelley & Lindsay, 1993). These demonstrations, as well as the many recent studies which indicate that people can be readily fooled to identify with confidence a wrong answer (see Koriat, Goldsmith, & Pansky, 2000), suggest again that recognition itself may involve more complex processes than those implied by the trace-access model.

We agree with Metcalfe, however, that errors of metacognition, such as the illusion of knowing (Koriat, 1998) or illusory TOT (Schwartz, 1999), do not necessarily argue against the possibility that metacognitive judgments are mediated by direct access. Indeed, it is a weakness of most studies of FOK accuracy (which implicitly subscribe to the direct-access view) that they take for granted that the subject’s effective memory target is identical with the experimenter’s (‘‘correct’’) target. However, as noted by Koriat (1995), in some TOT studies (e.g., Brown & McNeill, 1966; Koriat & Lieblich, 1974) the accuracy of the information reported about the illusive target was evaluated against the subject’s declared target, whether or not that target was correct. The problem, however, is that it is not always simple to specify what constitutes the subject’s ‘‘effective’’ target, as illustrated by the following observation: In some cases FOK and TOT experiences appear to be misled by activations stemming from a memory entry other than the correct target (e.g., a neighboring target), suggesting that the person’s ‘‘effective’’ target differs from that intended by the experimenter. Yet once the person succeeds in retrieving that target, he sometimes immediately recognizes it as being the wrong target (see Koriat, 1994). Should we then conclude that TOT and recognition monitor two different ‘‘effective’’ targets?

Further complications arise from the study of brain-damaged and mentally disturbed patients. Many of the recent metacognition studies with normal participants have focused on demonstrating a dissociation between cognition and metacognition. However, the observations from mentally disturbed patients discussed by Graham and Neisser and those on brain-damaged patients discussed by Metcalfe (see Metcalfe, 1996) suggest the possibility of some additional subtle dissociations that may exist within metacognition. These dissociations present a challenge for current analyses of the bases of metacognitive judgments, including the idea of direct access, and deserve further experimental investigation.

THE POSSIBILITY OF UNCONSCIOUS METACOGNITIVE PROCESSES

Spehn and Reder question the attempt to link metacognition to consciousness, and argue that, in fact, much of the metacognitive control of behavior occurs implicitly
or automatically, without conscious awareness. Similarly, Graham and Neisser, who define metacognition as second-order beliefs, see no reason to deny the possibility of unconscious, self-reflexive metacognition.

To what extent is it necessary to tie metacognition to consciousness? Consider Graham and Neisser’s definition of metacognition in terms of reflexive, second-order beliefs and attitudes. Can such beliefs and attitudes be unconscious? On the one hand, there is ample evidence to suggest that people can hold first-order beliefs and attitudes unconsciously (e.g., Devine, 1989; Greenwald & Banaji, 1995). So why could not second-order beliefs and attitudes be also unconscious? On the other hand, however, the idea of unconscious reflexive beliefs appears intriguing. We can think of situations such as those mentioned by Graham and Neisser in which a person holds a conscious second-order belief about a conscious first-order belief. However, could a person hold an unconscious belief that black people are hostile and also hold an unconscious belief that that belief is wrong? Could one have an unconscious negative attitude toward a certain acquaintance and also hold an unconscious negative attitude about that attitude? The answers to these questions are not obvious. In any case, these questions deserve empirical investigation, and clearly they have important metatheoretical implications.

Spehn and Reder (see Reder, 1996) were more definite in their position that indeed metacognitive monitoring and control can go on outside of awareness, and they summarized results supporting that position. In general, the findings indicate that (1) advance priming of the components of a question can enhance FOK judgments, presumably through the mediation of increased cue familiarity, and that (2) such advance priming can also affect strategy selection.

These findings are not necessarily incompatible with our position. First, as we have argued, the implicit use of the cue-familiarity heuristic can result in a conscious FOK (see Schwartz & Metcalfe, 1992). If advance priming does enhance conscious FOK, then there is no reason why such FOK could not guide the deliberate, controlled choice of a particular strategy of answering a question.

Second, we certainly do not deny the possibility that unconscious processes can affect behavior (including strategy ‘‘selection’’) automatically and implicitly, without the mediation of conscious awareness. Work such as that of Bargh and his associates (see Bargh, 1997) indicates that rather complex behaviors can be automatically triggered through priming.

Where does Spehn and Reder’s position depart from that advanced in the target article? They ask: ‘‘Could there not be a monitoring of one’s knowledge that produces an unconscious output, which would in turn feed an unconscious control process?’’ This question implies the postulation of an unconscious feeling of knowing that mediates between the effects of priming and strategy selection. In our view, such unconscious monitoring alone cannot possibly support deliberate, self-initiated strategy selection. Indeed, on the basis of the available evidence collected so far, it does not seem likely that unconscious beliefs can influence controlled action (see, Posner & Snyder, 1975). However, should such unconscious feeling of knowing be postulated to mediate the automatic effects of priming on the strategy of question answering?

This question raises the intriguing question of what is it that one knows after having been exposed to priming manipulations such as those used in Reder’s studies. For
comparison purposes, consider the study of Bargh, Chen, and Burrows (1996) cited in the target article, in which a person exposed to words related to the elderly stereotype was found to walk more slowly across the corridor. What should such a person be assumed to “know” as a result of the experimental manipulation? Certainly, his behavior presupposes the existence of a complex preexperimentally acquired knowledge structure (e.g., the elderly stereotype) that mediates between priming and behavior. But what kind of knowledge is the person assumed to gain as a result of the transient priming event? Possibly, the entire sequence of events can be considered to be an expression of anoetic, procedural memory (Tulving, 1985).

In the same manner, it might be the case that an implicit, anoetic process underlies the effects of priming on strategy-selection in Reder’s experiments. Clearly, these effects presuppose some intelligent knowledge base that intervenes between the input and the output. However, it is not clear that we must postulate the emergence of some new unconscious belief that is produced by priming. Indeed, Spehn and Reder consider this possibility. They argue that the experimental evidence suggests that “strategy selection is either an unconscious metacognitive process; or, if metacognition implies awareness, it is an implicit process” (see also Reder & Schunn, 1996).

Where does this leave us with regard to the question of unconscious metacognition? There are two comments that we would like to make. First, as far as terminology is concerned, we would propose to use the term unconscious monitoring rather than unconscious feeling of knowing, which invokes the long debated issue of whether feelings, in general, can be unconscious (see Ekman & Davidson, 1994). The position that we opted for (Koriat & Levy-Sadot, 1999) is that although feelings can result from unconscious processes, they are themselves conscious, phenomenal states (see also Clore, 1994).

Second, as an empirical question, the possibility of unconscious monitoring of one’s own knowledge is an open question that deserves investigation. We agree with Graham and Neisser that if metacognitive monitoring is defined as knowledge about one’s own knowledge, there is no a priori reason for denying the possibility that such knowledge might also be implicit and unconscious. As we noted earlier, it is both interesting and important to determine whether people can indeed have unconscious beliefs about their own conscious or unconscious beliefs.

If we accept the idea of implicit metacognitive monitoring of one’s own knowledge, then the model outlined in the target article would have to be modified so as to include a monitoring node in each of the three modes of operation postulated. So far we have distinguished between information-based and experience-based monitoring, corresponding to the explicit-controlled mode, and crossover mode, respectively. The position advocated by Spehn and Reder implies the postulation of metacognitive monitoring in the implicit-automatic mode of operation as well.

THE ISSUE OF METACOGNITIVE CONTROL

Even if metacognition is defined to include implicit-automatic monitoring and control processes, we argue that the unique aspects of metacognition are brought to the fore in the crossover mode of operation. In this mode, a conscious feeling of knowing that is formed by implicit unconscious processes can give rise to deliberate controlled
action. The important function of experience-based metacognitive judgments, then, is that they enhance self control, that is, they allow some degree of voluntary control over processes that could otherwise flow automatically into behavior.

This thesis is based on the general assumption, which Graham and Neisser challenge, that “consciousness contribute(s) to behavior by augmenting control.” However, they misinterpret this claim as implying that consciousness enhances the subjective sense of control over one’s behavior, which then increases the tendency to behave in a particular way. In response, they develop the argument that this kind of controllability is actually confounded with the “subjective sense of predictability.”

However, this is not the meaning of control that we had in mind. Graham and Neisser’s analysis focuses on the meaning that the term has in the context of the locus-of-control framework, as a set of metacognitive beliefs or feelings about the extent to which the outcome of one’s actions depends on oneself rather than on external factors. This meaning is consistent with Graham and Neisser’s emphasis on metacognition as second-order beliefs. In fact, control in this sense corresponds to the “monitoring” (rather than the “control”) component in the “monitoring-affects-control” equation (see Nelson, 1996; Nelson & Leonesio, 1988).

Rather, what we had in mind was precisely what Graham and Neisser described as the two possible contributions of consciousness to behavior (mss. p. 5): (1) to increase “the overall intelligence or rational appropriateness of behavior,” and (2) to contribute to the “ability of persons to control their own behavior and exert personal autonomy over bodily movements.” Beginning with the second mentioned contribution, we surmise that consciousness provides the person with some degree of voluntary control over his actions. Thus, while the processes described by Reder (like those described by Bargh and others), are peremptory in nature, prescribing a particular course of action (see Tulving, 1985), a metacognitive feeling leaves the person some room for negotiating action and judgment. Thus, he has, to a certain extent, the option not to act on that judgment, when he realizes, for example, that subjective experience has been contaminated by biasing influences (e.g., Kelley & Jacoby, 1996a). Indeed, several discussions in metacognition literature share this emphasis on the contribution of metacognitive monitoring to controlled, self-directed action (Koriat & Goldsmith, 1998; Nelson & Narens, 1994; Schwartz, Benjamin, & Bjork, 1997).

Related to this contribution of consciousness is the first contribution mentioned by Graham and Neisser. The position advanced in the target article is that when unconscious influences give rise to a subjective, conscious feeling, that feeling can be integrated with the goals and other beliefs of the person, and modulated by them in guiding judgment and action. The result is that judgments and actions are more intelligent and adaptive than when they represent the product of unopposed, unconscious, and automatic influences (see Hay & Jacoby, 1996). As Block (1995) argued, without consciousness one loses the “rational control of action.” In short, we concur with Feshbach’s (1999) conclusion that “consciousness makes a difference” (p. 207).

In concluding this rejoinder, we propose that the position presented in the target article should best be seen as a working hypothesis that is worth entertaining, although some of its tenets are still open to debate and investigation. Certainly, the possibility raised by Metcalfe, that metacognitive judgment sometimes can be based
on direct access to internal representation deserves further examination. So is the idea of unconscious metacognitive monitoring implied by Spehn and Reder. The possibility that metacognitive knowledge may mediate the implicit and automatic effects of priming on behavior is important for the evaluation and specification of the postulated distinction between the implicit-automatic mode of operation and the crossover mode of operation. More generally, the hypothesis raised by Graham and Neisser’s discussion, that second-order, self-reflexive beliefs can be unconscious (even when they apply to first-order unconscious beliefs) deserves to be tested because its rejection can have important metatheoretical implications. Finally, the target article was predicated on the working hypothesis that consciousness does exert a causal role on behavior, although, as Graham and Neisser stress, the observations in support of this hypothesis are open to alternative interpretations.

REFERENCES


Koriat, A., & Levy-Sadot, R. (submitted). The combined contributions of the cue familiarity and the accessibility heuristics to the feeling of knowing.


