Consciousness and Cognition **10,** 294–340 (2001)
doi:10.1006/ccog.2000.0494, available online at http://www.idealibrary.com on IDEAL®

Confidence and Accuracy of Near-Threshold Discrimination Responses

Craig Kunimoto

University of California, San Diego, La Jolla, California 92093

Jeff Miller

University of Otago, Dunedin, New Zealand

and

Harold Pashler

University of California, San Diego, La Jolla, California 92093

This article reports four subliminal perception experiments using the relationship between confidence and accuracy to assess awareness. Subjects discriminated among stimuli and indicated their confidence in each discrimination response. Subjects were classified as being aware of the stimuli if their confidence judgments predicted accuracy and as being unaware if they did not. In the first experiment, confidence predicted accuracy even at stimulus durations so brief that subjects claimed to be performing at chance. This finding indicates that subjects's claims that they are ''just guessing'' should not be accepted as sufficient evidence that they are completely unaware of the stimuli. Experiments 2–4 tested directly for subliminal perception by comparing the minimum exposure duration needed for better than chance discrimination performance against the minimum needed for confidence to predict accuracy. The latter durations were slightly but significantly longer, suggesting that under certain circumstances people can make perceptual discriminations even though the information that was used to make those discriminations is not consciously available. © 2001 Academic Press

For more than a century, psychologists have attempted to determine whether impoverished visual stimuli can be perceived without awareness (e.g., Eriksen, 1960; Holender, 1986; Sidis, 1898). In the 1950s, this area of research came to the forefront of psychology and to the attention of the lay public when an attempt was made to induce moviegoers to buy popcorn and other refreshments through the use of briefly presented, ''subliminal'' messages (e.g., McConnell, Cutler, & McNeil, 1958; Vo-

This research was carried out as a Ph.D. thesis (Kunimoto, 1994) by the first author under the supervision of the second and third authors, and it was supported by National Institute of Mental Health Grant PHS-40733 to Jeff Miller. We thank Anthony Greenwald, Steve Hillyard, Philip Merikle, Toby Mordkoff, Yoshi Nakamura, David Phillips, Eric Ruthruff, and Joan Stiles for helpful comments and discussion on earlier versions of this article. We also thank Tina Mirrashidi for help running the experiments.

Address correspondence and reprint requests to Craig Kunimoto, Institute for Theoretical Physics, UCSB, Santa Barbara, CA 93106; Jeff Miller, Department of Psychology, University of Otago, Dunedin, New Zealand; or Hal Pashler, Department of Psychology, UCSD, La Jolla, CA 92093–0109. E-mail: ckunimoto@ucsd.edu, miller@otago.ac.nz, or hpashler@ucsd.edu.

1053-8100/01 $35.00

key & Read, 1985). Despite intensive research over the past 5 decades, it is still not clear exactly how to study the relation between perception and awareness.

One major problem has been to define awareness itself. It might seem that a definition of awareness is a prerequisite for the study of perception without awareness (i.e., ‘‘subliminal perception’’), but a survey of the literature suggests otherwise. Instead, researchers who wish to test for subliminal perception often propose operational definitions of awareness, use them to find out whether perception can occur without awareness, and then argue after the fact about whether the definition was good (e.g., see Holender, 1986, and its commentaries: e.g., Bisiach, 1986; Fowler, 1986; Latto & Campion, 1986; Lupker, 1986; Merikle & Cheesman, 1986; Morton, 1986; Navon, 1986; Paap, 1986; Wolford, 1986; see also, Cheesman & Merikle, 1984, 1986; Henley, 1984; and Merikle, 1984).

This article examines a method for the operational definition of awareness involving confidence judgments (e.g., Peirce & Jastrow, 1884). The definition and methodology are introduced below after a brief discussion of other historically influential definitions of awareness. It is argued that the present confidence-based definition not only has the most important virtues of previous definitions, but also avoids the most serious of their problems. After this method is introduced, four experiments assessing its properties are presented. In the first experiment, the present method is contrasted with another recent method for measuring awareness. In the second experiment, the present method is used to test directly for subliminal perception. In the past, even when researchers agreed upon a definition of awareness, they were often unhappy with its implementation in particular experiments. The definition introduced here is no exception, and in the subsequent experiments, the proposed methodology for testing perception without awareness is refined.

Subjective Definitions of Awareness

The most intuitively appealing definitions of awareness are based on people’s introspective reports of their inner states (cf. James, 1890). Only the subject has access to his or her internal states of awareness, so if the subject claims to perceive (or not to perceive) a stimulus, then the awareness measure should not suggest otherwise. This intuitive grounding led many early researchers investigating subliminal perception to employ subjective definitions of awareness (e.g., Peirce & Jastrow, 1884; Sidis, 1898; Stroh, Shaw, & Washburn, 1908).

Sidis (1898), for example, had subjects discriminate between alphanumeric characters presented at a large-enough distance that subjects reported that they only saw a faint spot where the character should be. According to Sidis, this distance corresponded to the distance at which the subject was not quite aware of the stimuli. Despite subjects’ lack of awareness of the stimuli, Sidis found that they were still able to discriminate among stimuli at levels above chance performance. From this finding, Sidis concluded that ‘‘there is within us a secondary, subwaking, reflex consciousness’’ which influences discrimination responses. Similar conclusions were reached by Stroh, Shaw, and Washburn (1908), who found that subjects could discriminate among whispered letters even though they claimed not to hear any sounds.

Another type of subjective definition is based on confidence. It is plausible to assume that those who are confident of their perceptions have conscious awareness of them, whereas those who have no confidence have no awareness. For example, Peirce and Jastrow (1884) defined awareness in these terms in their studies of perception of small differences in tactile sensations. The subject would hold a weight on his or her finger, and the experimenter then either increased and decreased the weight or decreased and increased the weight. The subjects' task was to discriminate the direction of the weight change (i.e., to say whether the weight initially increased or initially decreased). In addition, subjects made a confidence response on a 4-point scale, ranging in confidence from . . . *it seemed nonsensical to answer at all* to *as strong a confidence as one would have about such sensations.* The results showed that subjects could discriminate the direction of weight change even when they had the lowest level of confidence in their discrimination responses. Peirce and Jastrow concluded that we may obtain information ''from sensations so faint that we are not fairly aware of having them . . . .''

Despite the phenomenological appeal of definitions based on claims of awareness or confidence, there is one major criticism of studies using such definitions: Subjective reports may be influenced by response biases. Thus, subjects may systematically claim not to see stimuli that they have partially or even fully seen (e.g., Erdelyi, 1974; Eriksen, 1960; Goldiamond, 1958; Postman, 1953). Indeed, evidence that subjects are systematically underconfident in sensory discrimination tasks (Björkman, Juslin, & Winman, 1993) suggests that this is a very real possibility. Thus, experimenters cannot be sure that a stimulus was subliminal when the subject reports that he or she was not aware of it, and it remains possible that evidence of perceptual processing in these experiments really represents perception with partial or complete awareness rather than subliminal perception as intended.

It should be emphasized that response biases need not be limited to uncooperative or untrained subjects. Awareness is not necessarily one of two disjoint states (unaware and aware), but may instead lie on a continuum of states ranging from unaware, through an infinite number of partially aware states, to complete awareness. Subjects in subliminal perception studies have to determine which states on this continuum correspond to the discrete labels of ''aware'' and ''unaware'' imposed on them by the experimenter. Those who are conservative may not report awareness unless absolutely sure, whereas those who are liberal may report awareness if anything at all is seen. The former subjects will have set a criterion for awareness closer to the *complete awareness* end of the continuum, whereas the liberal subjects will have set a criterion closer to the *unaware* end of the continuum. Furthermore, because experiments testing for subliminal perception virtually always involve difficult tasks with difficult awareness judgments, even highly trained subjects could well vary on the placement of their criteria. Thus, without a direct measure of subjects' criteria, it is impossible to determine whether the stimuli were actually ''subliminal'' in the sense that subjects had no awareness of them at all.

The problem of potential response bias contaminates each of the studies discussed above. In the Sidis (1898) study, for example, subjects' claims to see only faint spots did not truly establish that they were completely unaware of the stimulus letters; indeed, that they claimed to see any spots at all suggests some meager awareness. One could imagine, for example, that the faint spot produced by an *E* would look a little wider than a faint spot produced by an *I*. In short, even meager awareness of the spots may have been sufficient for subjects to discriminate at levels above chance. A precisely analogous criticism can be made about the auditory stimuli in the Stroh et al. (1908) study. Similarly, in the Peirce and Jastrow (1884) study, a ''nonsensical'' response was accepted as evidence that the subject had no awareness of the stimuli, but it is impossible to know exactly what ''nonsensical'' meant to each of the subjects.

In summary, it is impossible to define awareness simply in terms of subjects' reports that they were or were not aware of certain stimuli because their criteria for making such reports are simply unknowable. Clearly, then, despite the inherent appeal of using subjective report as an index of awareness, no certain conclusions concerning the hypothesis of perception without awareness can be drawn from studies based solely on such reports.

Objective Definitions Based on Correct and Incorrect Identifications

Objective definitions of awareness were introduced to avoid the problem of response bias in subjective reports, and they effectively transfer the definition of awareness from the hands of the subjects to those of the experimenter. According to one objective definition, for example, awareness can be inferred from the correct identification of a stimulus, and unawareness can be inferred from an incorrect identification (e.g., Lazarus & McCleary, 1951; McGinnies, 1949; see also Allport, 1977; Ellis & Marshall, 1978; Fowler, Wolford, Slade, & Tassinary, 1981; and Williams & Parkin, 1980, for variations of this definition). Using this definition, experimenters no longer have to worry about where subjects set their criteria for awareness, but can instead infer awareness directly from subjects' discriminative responses. Clearly, this definition has the advantage that any evidence for or against subliminal perception is independent of subjects' individual definitions of awareness.

One study that used this definition was conducted by Lazarus and McCleary (1951). On each trial, they presented subjects with 1 of 10 nonsense syllables. Five of these syllables had previously been conditioned to be associated with an electric shock, and the other 5 had not. All 10 nonsense syllables were presented tachistoscopically, and Lazarus and McCleary looked for galvanic skin responses (GSRs) of the sort typically elicited by shock-associated words. They found larger GSRs to the shock conditioned stimuli than to the nonconditioned stimuli, even on trials on which the subject was unaware of the stimulus—as indicated by an incorrect identification response. Lazarus and McCleary concluded that subjects were able to discriminate the stimuli (as measured by GSR) even in the absence of conscious recognition.

Lazarus and McCleary's (1951) definition successfully circumvents the response bias criticism that was raised against subjective definitions of awareness because it is difficult to argue that subjects were biased to respond with incorrect identifications of conditioned stimuli. Their definition still implicitly assumes, however, that awareness is dichotomous (i.e., people are either sufficiently aware to give a correct response or else completely unaware). As discussed previously, subjects might also experience states of partial awareness, in which they incorrectly identify a stimulus but still have some meager awareness of it. If the subjects were partially aware of some misidentified nonsense syllables, of course, this partial awareness could have been responsible for the fact that GSRs were larger for shock-associated stimuli than for non-shock-associated stimuli, so the effect would not indicate subliminal perception.

In fact, this alternative account of Lazarus and McCleary's (1951) findings was supported by Bricker and Chapanis (1953). Bricker and Chapanis gave subjects a list of nonsense syllables, and on each trial asked them to identify which syllable had been tachistoscopically presented. If a subject incorrectly identified a stimulus, the subject was asked to try again and continue trying until he was correct. Bricker and Chapanis examined the trials to which the initial response was wrong and found that subjects required significantly fewer additional guesses to identify the stimulus than would be expected by chance. They concluded that even when subjects incorrectly identify a stimulus, they must still have enough partial information about the stimulus to facilitate a correct guess on subsequent identification attempts. This view received additional support from the experiments of Wiener and Schiller (1960), who also found evidence of partial information in three different subliminal perception paradigms (see also, Eriksen, 1956; and Howes, 1954).

In summary, even though an objective definition of awareness based on correct versus incorrect identifications successfully avoids the problem of response bias in subjective reports, it is still problematic because of its underlying assumption that an incorrect identification implies no awareness. In fact, subjects who make an incorrect identification may still have some awareness of the stimulus (Bricker & Chapanis, 1953), so evidence of perceptual processing in conjunction with an incorrect identification cannot be taken as evidence of perception without awareness.

Objective Definitions Based on Chance and Greater Than Chance Performance

To avoid the problem of partial information described in the previous section, many subsequent researchers have used objective definitions based on subjects' overall performance in a presence–absence judgment task or a stimulus discrimination task. Subjects are said to be aware of the stimuli if their performance is above chance and unaware if their performance is at chance. Like the objective definition based on the correct and incorrect identification of a stimulus, experimenters infer subjects' awareness based on the subjects' responses to the stimuli and not on subjects' introspections; thus, the measure of awareness is not contaminated by biases in subjects' assessments of their own awareness. Furthermore, the definition is not susceptible to the partial information criticism because partial information about a stimulus should result in above-chance performance over a sufficiently long series of trials.

For example, Cheesman and Merikle (1984) used discriminative responses to assess awareness. They determined separate awareness thresholds for four color words by decreasing the exposure durations of the words until subjects discriminated the words consistently at chance. When Cheesman and Merikle presented these words at the threshold durations as primes in a color naming (Stroop) task, they found that, unlike consciously presented primes, primes presented at the awareness threshold had no effect. That is, the time to name the color of a color patch was the same with consistent prime words as with inconsistent primes. Based on their assumption that the discrimination threshold was equal to the awareness threshold, Cheesman and Merikle concluded that perception does not occur without awareness (see also Nolan & Caramazza, 1982).

Based on negative results of a number of studies using performance-based definitions of awareness, a number of researchers have concluded that no perception, conscious or otherwise, occurs when stimuli are presented below this threshold (e.g., Holender, 1986; Kunimoto, 1994; Merikle, 1992). Some studies have contradicted this conclusion by reporting evidence of subliminal perception with stimuli presented below performance-based thresholds (e.g., Balota, 1983; Erdley & D'Agostino, 1988; Fowler, Wolford, Slade, & Tassinary, 1981, Experiments 5 and 6; Greenwald, Klinger, & Liu, 1989; Greenwald, Klinger, & Schuh, 1995; Marcel, 1983; Niedenthal, 1990; Snodgrass, Shevrin, & Kopka, 1993), but the evidence from such studies has generally been disputed because of demonstrated or suspected methodological problems—for example, insufficient trials used for threshold determination (e.g., Dosher, 1998; Duncan, 1985; Holender, 1986; Kunimoto, 1994; Merikle, 1982, 1992; Miller, 2000; Nolan & Caramazza, 1982; Van Selst & Merikle, 1993). Thus, it is still arguable which paradigms are capable of determining whether perception occurs when stimuli are presented below the threshold for a discriminative or presence–absence response.

Ultimately, studies of whether perception occurs when performance is at chance may not help resolve the issue of whether perception can occur without awareness. Performance-based definitions of awareness are completely insensitive to phenomenal experience, and it seems intuitively clear that any truly satisfactory operational definition of awareness must take phenomenology into account somehow (Bisiach, 1986; Cheesman & Merikle, 1984, 1986; Fowler, 1986; Latto & Campion, 1986; Lupker, 1986; Merikle & Cheesman, 1986; Morton, 1986; Navon, 1986; Paap, 1986; Wolford, 1986). As Paap (1986) pointed out: Even though a thermostat may make adequate temperature discriminations, it clearly is not aware of temperature. Similarly, subjects may adequately discriminate or make presence–absence judgments about stimuli of which they are not aware. It is certainly possible, then, that reducing exposure durations to the level of chance performance (e.g., Cheesman & Merikle, 1984) eliminates all forms of perception, both with and without awareness. In that case, it would not be surprising that many studies have failed to find evidence of subliminal perception below the performance-based threshold.

Subjective Definitions with a Converging Measure

Cheesman and Merikle (1986) rejected objective definitions of awareness because of their insensitivity to phenomenological experience, and they argued that, by definition, awareness could only be measured subjectively. They acknowledged that subjective definitions might be contaminated by response biases, and suggested that experimental demonstration of qualitative differences between perception of ''subliminal'' and ''supraliminal'' stimuli could be used to confirm that the subliminal stimuli had indeed been perceived without awareness. In essence, they argued that qualitative differences in processing of weak and strong stimuli would confirm that the former had really been processed subliminally (cf. Jacoby & Whitehouse, 1989).

Cheesman and Merikle (1986) applied this idea in the following way. In the first part of an experimental session, they estimated subjects' awareness thresholds for four stimulus words (*Blue*, *Green*, *Red*, and *Yellow*). These words were pattern-masked, and the subjects' task was to identify which of the four stimuli was presented on each trial. In addition to this response, at the end of each 24-trial block, subjects were asked to report the total number of trials in which they were confident that they were correct. Awareness thresholds were established by systematically decreasing the prime-mask stimulus onset asynchrony (SOA), until subjects consistently indicated over three consecutive blocks that fewer than 3 of the 24 decisions had been made with confidence.[1] The resulting SOA was taken to be the subjects' awareness threshold.

In the second part of the experiment, the subjects' task was to name, as quickly as possible, the color of a color patch. The words from the first part of the experiment were presented as primes prior to the color patches. In half of the trials, the primes were presented at the subjects' awareness thresholds; in the other half, the primes were presented at an exposure duration easily perceived by the subjects. In addition, the prime–target relationship was also varied: In congruent trials the prime and target had the same name, whereas in incongruent trials they had different names. With supraliminal presentation, not surprisingly, naming responses were faster when the prime–target relationship was congruent than when it was incongruent. More importantly, this effect was also obtained with primes presented just below the awareness threshold, suggesting the presence of subliminal perception. Crucially, however, there is a possible response-bias-based alternative explanation of these results. Specifically, subjects may have actually had some awareness even at their awareness thresholds (i.e., at the durations for which they reported 3 or fewer confident responses per block), but they may have been biased to report that they had made relatively few correct responses. After all, it is impossible to be sure what criterion a subject is using to decide whether a response was likely to have been correct, just as it was impossible to be sure what criterion a subject uses in judging whether he or she is aware of a stimulus.

To buttress their evidence of perception without awareness against this alternative explanation, Cheesman and Merikle (1986) manipulated the proportions of congruent and incongruent prime–target relationships. In one half of the experimental session, congruent trials occurred in 25% of the trials and incongruent trials occurred in the other 75%; in the other half, these proportions were reversed. Several studies (e.g., Glaser & Glaser, 1982; Taylor, 1977) have shown that the Stroop priming effect (Stroop, 1935) increases with the proportion of congruent prime–target combinations. Cheesman and Merikle (1986) suggested that the Stroop effect depends on the proportion of congruent trials when subjects are aware of the proportion manipulation and voluntarily adopt a strategy of using the predictive prime information (e.g., Neely, 1977; Posner & Snyder, 1975; Taylor, 1977). On this view, the effect of proportion

[1] In previous experiments Cheesman and Merikle asked subjects for a percentage correct estimate of their performance, decreasing prime-mask SOAs until subjects consistently indicated over three consecutive blocks that their performance was at chance. Cheesman and Merikle found comparable thresholds for the two procedures.

should disappear if subjects are not aware of the primes because they would not be aware of the proportion manipulation. Consistent with this analysis, Cheesman and Merikle (1986) found that the proportion manipulation influenced the priming effect when primes were presented supraliminally but not when they were presented at the estimated awareness thresholds. Based on this qualitative difference, Cheesman and Merikle argued that subjects really had not had any awareness of stimuli presented at the estimated awareness thresholds and that the Stroop priming produced by these stimuli was therefore strong evidence of subliminal perception.

Although the converging operations methodology proposed by Cheesman and Merikle (1986) is a clear improvement over complete reliance on subjective thresholds, we nonetheless regard it as a suboptimal methodological approach for two reasons. First, it depends ultimately on the assumption that qualitatively different results occur *only* because of a change from processing stimuli with awareness to processing stimuli without awareness. Unfortunately, this is a rather doubtful assumption. Logically, it is quite possible that qualitative differences could occur for a variety of reasons, even between supraliminal conditions. In Cheesman and Merikle's experiment, for example, subjects might have failed to use the predictive information from a brief prime because they were not confident about the prime's identity, even if they had some conscious percept of the prime. That is, people may be fairly conservative about bothering to use a prime just as they may be rather conservative about claiming to have seen a prime. There is no obvious reason to believe that the exposure duration at which people start to use a prime is the same as the exposure duration at which they start to become aware of the prime. If the former exposure duration is slightly larger, then Cheesman and Merikle's qualitative difference does not demonstrate that subjects were unaware of the primes, but simply reinforces the fact that subjects were not confident about them. Similar criticisms can be made of other studies using this same converging operations methodology (cf. Joordens & Merikle, 1992; Kunimoto, 1994). Second, the converging operations methodology is necessarily indirect. Rather than studying the relation between perception and awareness directly, experimenters are forced to find experimental manipulations that will have qualitatively different effects on subliminal versus supraliminal processes. Yet, it is not clear that such experimental manipulations must always be available.

A Bias-Free Measure of Awareness

We turn now to a measure that seems to satisfy the two major requirements of an awareness measure: (a) It appears to be phenomenologically valid because it depends crucially on the subject's own introspective analysis of his or her performance and (b) it appears to be bias-free because it can be defined in such a way that it is essentially unaffected by the subject's own criterion for awareness.

The approach presented here follows the tradition of using introspective reports to assess awareness because conscious experience is the *sine qua non* of awareness. We use subjective reports somewhat differently than those researchers described above, however, in order to eliminate the response biases to which introspective reports are naturally prone.

To illustrate the method, consider the following procedure: Suppose that after making a set of discrimination responses (at a given exposure duration), subjects are asked to rank those responses based on their confidence that the response was correct. If subjects have no awareness of the stimuli, then there should be no relationship between subjects' confidence rankings and performance. On the other hand, if subjects do have some awareness of the stimuli, then higher confidence rankings should be associated with more correct discrimination responses than lower confidence rankings. Although subjects are not asked directly about their awareness, their awareness can be assessed by the relationship between their confidence judgments and accuracy, under the plausible assumption that their confidence cannot reflect their accuracy unless they are at least partially aware of the information on which they based their discriminative responses. As the exposure duration of the stimuli is decreased to the true awareness threshold, subjects' ability to rank the discrimination trials based on their confidence should also decrease. When stimuli are presented so briefly that subjects have no conscious awareness of them, their confidence rankings should be unrelated to their performance.

Unfortunately, this illustrative procedure would not be practical because the subjects could not reasonably be expected to keep track of their responses and confidence levels across a long series of trials. A simpler method is to let subjects assign one of only two confidence judgments to their performance on each trial—high or low confidence—and then note how well these confidence judgments predict performance. Even with this simplified confidence rating task, the above reasoning holds. Subjects should randomly assign confidence judgments to discrimination responses if and only if they are completely unaware of the stimulus information used to make those responses, assuming that confidence judgments reflect their awareness of the consciously available stimulus information. Alternatively, subjects should assign confidence judgments systematically (i.e., report high confidence more often for correct trials than incorrect trials) if and only if they are aware of the stimuli because states of confidence should be related to accuracy only for subjects who are at least partially aware of the stimuli they are discriminating.

This trial-by-trial categorization of correct versus incorrect discrimination responses with high versus low confidence responses can be analyzed using signal detection theory (SDT)—a model well-known for its ability to separate biases from sensitivity (Green & Swets, 1966). Table 1 shows how the trials can be categorized in terms of SDT: A ''hit'' corresponds to a correct identification with high confidence; a ''miss,'' a correct identification with low confidence; a ''false alarm,'' an incorrect identification with high confidence; and a ''correct rejection,'' an incorrect identifi-

TABLE 1
Categorization of Subjects' Confidence Responses for a Signal Detection Analysis

Discrimination accuracy	Confidence	
	High	Low
Correct	Hit	Miss
Incorrect	False alarm	Correct rejection

cation with low confidence. SDT's measure of sensitivity (*d′*) is easily estimated using data categorized in this manner.

But what does *d′* represent here? Figure 1 shows the frequency distributions representing correct and incorrect responses, which can be constructed from Table 1. As seen in the figure, *d′* is the distance between the means of the distributions representing correct responses and incorrect responses. Therefore, *d′* measures subjects' ability to discriminate between their correct and incorrect trials; and, thus, it measures subjects' awareness of their own performance. This is exactly the conclusion and interpretation given for the ''ranking'' of two confidence responses described above.

A crucial advantage of this approach is that the application of SDT eliminates the effects of response biases from subjects' confidence judgments. As seen in Fig. 1, the placement of the criterion to report ''high'' or ''low'' confidence reflects the amount and direction of bias for each subject. Thus, some subjects may be biased to report low confidence more often than high confidence (or vice versa) due to differences in the placement of their criteria. But the placement of the criterion is unimportant. It allows subjects to decide what response to make, but it does not obscure the subjects' ability to discriminate between correct and incorrect responses. It is the ability of subjects to discriminate between correct responses and errors, unaffected by criterion placement, which we are interested in, and this is exactly what *d′* measures.

In summary, the proposal here is to measure awareness in terms of the subjects' ability to discriminate between correct and incorrect responses, using the metric of *d′* provided by SDT. We conclude that subjects are aware of the stimuli if and only

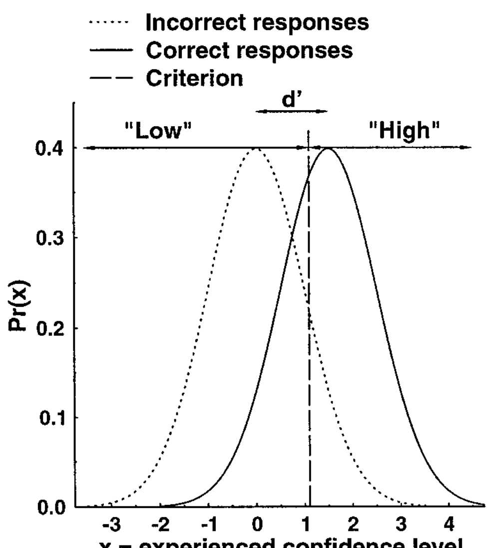

FIG. 1. Assessing awareness (*d′*) with Signal Detection Theory. The incorrect-response frequency distribution is on the left, whereas the correct-response frequency distribution is on the right. The criterion for reporting confidence is indicated by the vertical line. *d′* is defined as the distance between the means of the distributions (in standard deviation units) and indicates the sensitivity of the subject in discriminating between correct and incorrect responses. A *d′* of zero indicates no awareness, whereas a *d′* greater than zero indicates awareness.

if confidence is related to accuracy (i.e., $d' > 0$). Thus, this approach supplements the tradition of using subjective reports (i.e., confidence judgments) for assessing awareness by analyzing confidence reports with techniques developed in SDT in order to eliminate response biases. The resulting awareness measure is not only phenomenologically valid, but also bias-free. In the terms of Reingold and Merikle (1988), this is a ''direct'' measure of perceptual awareness, because choosing a confidence level (i.e., monitoring one's own accuracy level) is one of the task requirements.

A closely related approach has been used by Kolb and Braun (1995) in a study with somewhat complicated psychophysical tasks involving motion detection and binocular rivalry. After each display, they asked subjects to give both an identification response and a 1–10 confidence rating. Using an analysis based on receiver operating characteristics, they found that the accuracy of the identification response was unrelated to the level of confidence, and on that basis they concluded that subjects were not aware of perception underlying their correct identifications. Our approach differs in three main ways from that taken by Kolb and Braun. First, we studied word identification tasks rather than psychophysical discrimination. Both types of tasks are important, of course, but the word identification tasks have probably been more common in the literature on subliminal perception. Second, in Experiment 1, we compared our awareness measure to one of the primary awareness measures employed previously. Kolb and Braun, in contrast, did not examine other awareness measures because they were primarily interested in investigating the mechanisms underlying motion perception and rivalry. Third, in Experiments 2–4 we adjusted display parameters based on the new awareness measure to estimate the threshold for awareness directly. Kolb and Braun did not adjust their display parameters based on awareness measurements; rather than estimating the threshold for awareness, they showed that a fixed set of display parameters was below this threshold.

One conceptual line of criticism must be addressed before proceeding. It could be argued that confidence judgments might be made without awareness just as discrimination judgments can be made without awareness. If so, then confidence judgments could predict response accuracy even if there is no awareness of the information on which those responses were based, which would invalidate our proposed measure of awareness as well as that of Kolb and Braun (1995). This could occur if subjects made their confidence judgments using specific strategies not based on awareness. For example, sophisticated subjects might realize that in difficult psychophysical tasks fast responses are more likely to be correct than slow ones (e.g., Kellogg, 1931; Swensson, 1972; Vickers, Burt, Smith, & Brown, 1985). If these subjects could estimate their response times, then they might be able to discriminate correct responses from errors without any direct awareness of the stimulus information used to make the response. Analogous ''cross-cuing'' effects have been observed in split-brain patients, where one hemisphere can gather information from another by monitoring an external cue like direction of gaze (e.g., Gazzaniga, 1970).

There are two responses to this conceptual criticism. First, like the phenomenology problem associated with discriminative reports, the cross-cuing criticism is only problematic for experiments which do not find evidence for subliminal perception. In the priming paradigm, for example, use of information other than awareness would drive the estimated threshold of awareness below the actual threshold and thus would bias the results against finding significant priming. As is shown, the present experiments provide evidence in favor of subliminal perception, so they are not subject to this criticism. Second, this same argument could be made about any behavioral measure, so if it is accepted researchers would be left with no measure of awareness. For example, suppose subjects are asked to say ''I saw it'' or ''I did not see it'' when presented with a target stimulus. Subjects could potentially make these responses based on cross-cuing information (e.g., changes in display flicker) rather than on any phenomenological experience of the target stimulus. There is no evidence, however, that such cross-cues are typically available in subliminal perception experiments or that normal subjects do use them. Thus, in the absence of any such positive evidence for cross-cuing, it seems reasonable to assume that subjects' confidence ratings simply reflect their awareness of the information on which their responses are based.

Applying the Definition

In Experiment 1, the present definition of awareness was compared with a more established definition in order to estimate the effect of response bias on subjective measures of awareness. Specifically, awareness thresholds estimated using the SDT procedure discussed above were compared with awareness thresholds estimated using Cheesman and Merikle's (1984, 1986) block-estimation procedure, which is potentially contaminated by response biases. Their procedure was chosen for comparison because it is the most rigorous of the procedures previously proposed for measuring an awareness threshold using introspective reports. Experiment 1 shows, however, that Cheesman and Merikle's procedure yields substantially larger estimated thresholds than the new procedure proposed here, from which we conclude that response biases are an important contaminating factor that must be taken into account in that procedure.

Experiments 2–4 used the proposed procedure in direct tests of the hypothesis of perception without awareness. Perception was assumed to have occurred, to at least some degree, if and only if discrimination performance was better than chance and subjects' thresholds for perception—i.e., better than chance discrimination performance—were found by manipulating stimulus exposure duration.[2] Similarly, subjects' thresholds for better than chance discrimination between their correct responses and errors—i.e., $d' > 0$, the awareness threshold—can also be found. A simple comparison of the two thresholds directly answers the question of whether people can perceive briefly presented stimuli without awareness. If the two thresholds are the same (i.e., if, as exposure duration decreases, subjects lose their ability to discriminate correct responses from errors at the same exposure duration at which they lose their

[2] The psychophysical *threshold* has traditionally been defined as the point at which subjects' performance is 50% (corrected for guessing). In studies exploring perception without awareness, however, the threshold is defined as that point at which subjects' performance is 0%. As numerous researchers have pointed out, if the threshold is defined as 50%, then stimuli presented below the threshold ''might well be perceivable as much as 49% of the time'' (McConnell, Cutler, & McNeil, 1958). In the experiments presented here, then, the *threshold* refers to the point at which performance is 0% after correction for guessing.

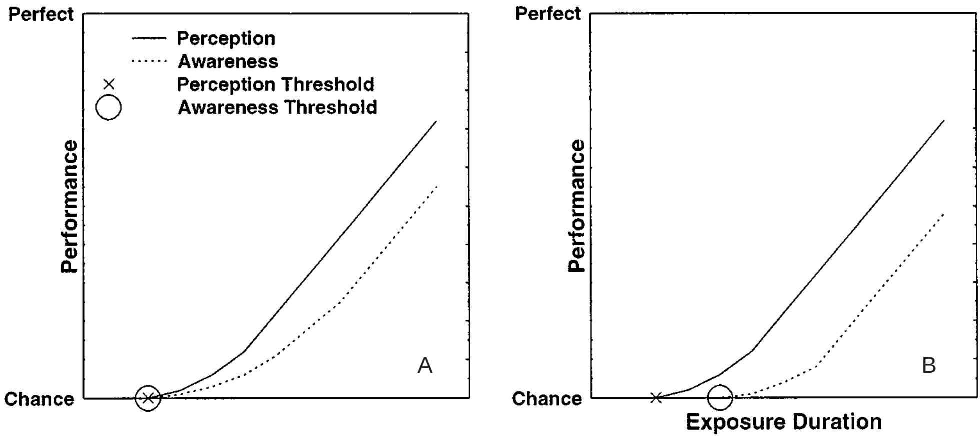

FIG. 2. A diagnostic for testing the hypothesis of perception without awareness. Hypothetical performance functions for perception and awareness are shown as a function of exposure duration. The diagnostic is to compare the minimum exposure duration needed to perceive with the minimum exposure duration needed to be aware. If the durations are the same, as seen in A, then the subjects are aware of what they are perceiving. However, if the minimum duration for perception is lower than the minimum duration for awareness, as seen in B, then the subjects are not aware always aware of what they are perceiving. When subjects have no awareness of the stimuli (as indicated by the awareness threshold), subjects are clearly perceiving at above-chance levels.

ability to discriminate among the stimuli), then we can conclude that there is no perception without awareness. This possible outcome is illustrated in Fig. 2A. If thresholds for awareness are higher than those for perception (i.e., if, as exposure duration decreases, subjects lose their ability to discriminate correct responses from errors before they lose the ability to discriminate among stimuli), however, then there must be perception without awareness, at least at certain intermediate thresholds. This possibility is illustrated in Fig. 2B. Thus, a direct diagnostic for determining whether people can perceive without awareness is a simple comparison of thresholds, and this diagnostic was used in Experiments 2–4. Experiment 2 provided an initial comparison of awareness thresholds with perception thresholds, and evidence of perception without awareness was found. Experiments 3 and 4 addressed several potential artifactual explanations of the evidence obtained in Experiment 2.

EXPERIMENT 1

Because the proposed method for assessing awareness thresholds removes response biases, it is potentially an advance over other procedures which use subjective definitions of awareness (e.g., Cheesman & Merikle, 1986). Whether this advance is significant or not in practice, however, depends on the extent to which response biases are present in the other procedures. As noted previously, Cheesman and Merikle's (1984, 1986) procedure is perhaps the most rigorous of the currently available methods for determining subjective thresholds. Therefore, in this experiment, awareness as measured with the proposed SDT procedure was contrasted with awareness as measured by Cheesman and Merikle's (C&M) procedure. If the C&M procedure is not contaminated by response biases, then the two procedures should yield comparable awareness thresholds. But if the C&M procedure is contaminated, then the two procedures should yield different awareness thresholds.

Separate groups of subjects were tested in the two procedures in order to avoid possible carryover effects. This between-subjects design threatened the statistical power of the experiment, however, because of individual differences in sensitivity to brief, masked displays. Worse yet, the two different methods for measuring awareness require slightly different events on each trial, and these trial-to-trial differences could also cause actual changes in sensitivity. However, differences in awareness resulting from sensitivity differences are not of interest, but only those differences resulting from effects of response bias. Therefore, to increase power and control for any effects of display sequence on sensitivity, discrimination performance was measured concurrently with awareness at various levels of performance approaching the discrimination threshold. In essence, the question was whether the two measures of awareness behaved similarly as discrimination performance approached threshold.

Method

Apparatus and stimuli. Subjects sat approximately 60 cm from a Tektronix 602 display unit equipped with a P15 fast-decay phosphor. The stimuli were the words *Blue, Green, Yellow,* and *Orange* presented in the middle of the screen. Each letter subtended 0.53° (5.5 mm) vertically × 0.38° (4 mm) horizontally, and the letters were separated by 0.48° (5 mm) of visual angle center to center. Each word was preceded by a fixation box in the center of the screen, subtending 0.53° (5.5 mm) vertically × 0.76° (8 mm) horizontally, and was followed after variable stimulus onset asynchronies (SOAs) by a pattern mask, consisting of eight pound signs (#). Each pound sign subtended the same visual angle as each letter. The shortest word subtended approximately 1.91° (20 mm), and the longest word subtended approximately 2.87° (30 mm); the mask subtended approximately 3.34° (35 mm). The fixation box, stimuli, and mask were composed of dots such that the grain of a horizontal or vertical line was approximately 1 dot/mm. The number of dots required for the fixation box, word stimuli, and mask differed, so they took different amounts of time to plot on the point-plot display: the fixation box took approximately 3 ms, the word stimuli approximately 11 ms, and the mask approximately 16 ms.

An IBM PC compatible computer equipped with a Data Translation 2801 digital-to-analog converter was used to control the experiment and present the stimuli. Responses were collected with an AT-style keyboard. The keys *Z, X,* >, and *?,* corresponding to the four stimulus words, *Blue, Green, Yellow,* and *Orange,* respectively, were labeled with stickers indicating which word each key represented. The keys *1* and = (also labeled; see below) were used to indicate low and high confidence responses, respectively, for the new procedure (SDT procedure; described below). The number keys were used to enter the estimated percentage of correct discriminations for Cheesman and Merikle's procedure (C&M procedure; described below). During the course of the experiment, a CRT next to the Tektronix 602 presented on-line instructions to subjects.

Subjects and procedure. Forty undergraduates participated in the experiment to partially fulfill a course requirement. Half the subjects were tested with the SDT procedure and the other half were tested with the C&M procedure. Each subject was tested individually, in a single session averaging approximately 45 min.

To determine discrimination thresholds, a procedure modeled after Cheesman and Merikle (1984) was used. Stimuli were presented in blocks of 48 trials, with each stimulus appearing 12 times within each block. A new random order was generated for each block. The stimulus-to-mask SOA was constant within each block of 48 trials, and it was adjusted from block to block dependent on subjects' performance. The first block for each subject had an SOA of 100 ms. If discrimination performance was below 85% correct, the SOA was raised 15 ms and raised 15 ms for each subsequent block until the subject exceeded 85% correct for a block. Once the subject achieved at least 85% correct for a block, the SOA was lowered 20 ms for each block in which discrimination accuracy fell between 85 and 100%, lowered 10 ms for each block in which accuracy was between 50 and 85%, and lowered 5 ms for each block in which accuracy was between 30 and 50%. Discrimination threshold was defined as the SOA for which discrimination performance was below 30% for three consecutive blocks. If discrimination performance rose above 30%, then the SOA was again lowered 5 ms and the subject was again required to perform below 30% correct for three consecutive blocks. The session was completed when discrimination threshold was reached (i.e., three consecutive blocks at or below 30% correct). The session was also terminated if the subject ran through three consecutive blocks at the lower presentation limit for the stimulus words (11 ms).

The two procedures for measuring awareness (C&M versus SDT) were tested on different groups of subjects. Half of the subjects estimated their performance after each block, following Cheesman and Merikle's (1984) procedure (C&M procedure); the other half made low/high confidence judgments after each trial, for later analysis with signal detection theory (SDT procedure).

The trials for the two groups were very similar. First, a fixation box appeared for 1 s. Then a stimulus word was presented, followed for 300 ms by the pattern mask (the SOA varied according to the scheme described above). Following the mask, the subject made a keypress to indicate which stimulus had been presented.

For subjects in the C&M group, the fixation box for the next trial was displayed 750 ms after each discrimination response. At the end of the block, subjects typed in their accuracy estimates (in Percent Correct) for the discrimination trials. Subsequent blocks were started by pressing the space bar.

For the SDT group, the first three blocks were the same as for the C&M group with the exception that no accuracy estimates were made. Upon completing the third block, however, the experimenter introduced the trial-to-trial confidence judgments by using a monetary incentive. For each block, subjects were allocated 24 green chips to bet when they thought they had discriminated correctly (high confidence) and 24 red chips to bet when they thought they had discriminated incorrectly (low confidence). Subjects bet a single chip per trial and were given equal numbers of the two types of chips in order to make sure that both confidence levels were used (a necessity for our measurement of d'). The number of chips of each type were not physically present, but instead were displayed on the CRT.

Subjects in the SDT group were told that on each trial they should bet one chip according to their confidence on that trial. The relation of chip color to confidence was made concrete in terms of the monetary payoffs. If subjects discriminated correctly and bet a green chip (a hit), they won \$0.05; if they discriminated correctly but bet a red chip (a miss) they won \$0.03; if they discriminated incorrectly and bet a green chip (a false alarm) they lost \$0.03; and if they discriminated incorrectly and bet a red chip (a correct rejection) they lost \$0.01. A card displaying these payoffs was placed next to the keyboard. The experimenter emphasized that it was always better to discriminate correctly than incorrectly, and it was always better to bet a green chip if the discrimination response was correct and a red chip if the discrimination was incorrect.

Thus, for the SDT group, trials subsequent to the end of the third block proceeded as for the C&M group up through the discrimination response. After each discrimination response, however, subjects were prompted to bet a chip of one color or the other. The CRT indicated the number of chips of each type remaining (along with a reminder of what each chip represented). After this second response, the CRT acknowledged the color of the bet chip for 1 s. Then, following a 1-s blank interval, the fixation box for the next trial appeared.

If subjects tried to bet chips they did not have, the CRT indicated that no more of the chosen chip remained, then redisplayed the number of chips of each type, along with a prompt for a new bet. Subjects made bets by pressing the *1* key for ‘‘red chip’’ and the = key for ‘‘green chip’’ (both these keys were labeled with a sticker, indicating the chip color).

At the end of each block, the CRT displayed subjects’ current balance (in cents) for subjects in the SDT group. Subsequent blocks were started by pressing the space bar. Subjects did not have to pay the experimenter if the ending balance was negative, and the experimenter never payed more than a maximum of \$2.50. Both of these details were explained to the subject beforehand. Subjects’ motivation was maintained in the latter case by explaining that the trials would get more difficult as the experiment progressed, so the potential to lose money would increase.

Results and Discussion

The average number of blocks needed to complete a session was nine. Fifteen subjects completed the session by reaching the criterion for a discrimination threshold, and 25 subjects completed the session by reaching the lower limit of the apparatus on word presentation time. Even though these subjects did not reach discrimination threshold, they did approach it closely. Thus, as is shown, their data were still useful in investigating the relation between awareness and nearthreshold discrimination.

For each subject, each block of trials yielded measures of both discrimination accuracy and awareness. Figure 3 shows the relationship between these measures for both the C&M group (Fig. 3A) and the SDT group (Fig. 3B). Discrimination performance in percentage correct (PC) is plotted on the abscissa, while awareness (in estimated PC for the C&M procedure and d' for the SDT procedure) is plotted along the ordinate. Each point on the figure represents one block of trials.

When subjects were asked to estimate their performance at the end of each block

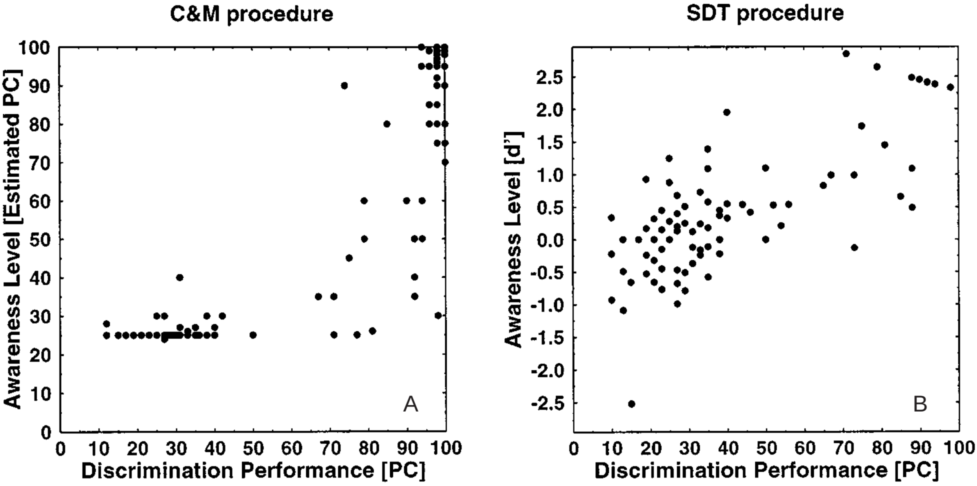

FIG. 3. Awareness as a function of discrimination performance for the (A) C&M procedure and (B) SDT procedure. (Each point represents one block of trials. The first three blocks for each subject are not plotted; blocks in which subjects' discrimination performance was 100% are also not plotted.)

(Fig. 3A), it is clear that most estimated they were performing at chance (25% correct) even when they were actually correctly identifying the stimulus as much as approximately 60% of the time. This result is consistent with the general pattern of underconfidence shown by subjects in sensory discrimination tasks (Björkman et al., 1993), and it replicates the pattern obtained by Cheesman and Merikle (1984, Experiment 2) very well. Thus, if one accepts subjects' estimates of chance performance as the operational definition of awareness, it is clear that there is substantial subliminal perception in this group. As discussed above, however, another possibility is that response biases lower subjects' estimates of their own discrimination accuracy. Specifically, subjects may report that they are guessing when not absolutely sure about what they have seen, even if they were aware of having extracted partial information about the stimulus from the display. If that is the case, then subliminal perception cannot be inferred from the finding of better than chance discrimination performance in blocks for which subjects estimated their discrimination accuracy to be at chance.

The results were quite different when subjects were asked to indicate their relative confidence in their response for each trial (Fig. 3B). In this procedure, subjects generally showed evidence of awareness (i.e., they could to some extent discriminate correct responses from errors) in all blocks of trials for which discrimination accuracy was above chance. From these results, there is no clear evidence of discrimination without awareness.

To quantify the between-groups differences suggested by Fig. 3, discrimination performance was categorized into four ranges or ‘‘bins’’: 0–30%, 30–60%, 60–90%, and 90–100%. Awareness assessments for each block were put into the bin corresponding to discrimination performance for that block. If several blocks for a given subject fell within a bin, the awareness assessments for those blocks were averaged

TABLE 2
Mean Awareness Level as a Function of Discrimination Performance and Awareness Assessment Procedure

		Discrimination performance			
Awareness assessment		0–30%	30–60%	60–90%	90–100%
C&M procedure (PC)	Mean	25.20	26.42	42.61	91.12
	N	20	18	9	20
	t	1.60	1.71	3.48	32.36
	p	*ns*	*ns*	$<.01$	$<.001$
SDT procedure (d')	Mean	−0.093	0.241	1.392	2.357
	N	20	15	12	9
	t	−1.14	2.49	5.35	211.36
	p	*ns*	$<.05$	$<.001$	$<.001$

Note. The *t*-tests and associated *p* levels test the null hypothesis of chance performance (i.e., that $d' = 0$ for the SDT procedure or that PC = 25 for the C&M procedure).

together, so that each subject would contribute only a single observation to each bin and the various observations within a bin would be statistically independent (i.e, one per subject). For example, if a subject produced four blocks with discrimination accuracy under 30%, three blocks in the range 30–60%, zero blocks in the range 60–90%, and 1 block in the range 90–100%, then the four blocks under 30% would be averaged together, the three blocks from 30–60% would be averaged together, the subject would not contribute to the 60–90% bin, and only one estimate would contribute to the 90–100% bin.

Table 2 shows for each group the mean awareness assessment within each bin, the number of subjects who contributed to each bin, and the results of *t*-tests of the null hypothesis of no awareness (i.e., PC = 25% for the C&M procedure and $d' = 0$ for the SDT procedure). The C&M and SDT procedures clearly do not give the same awareness assessments. When subjects' objective discriminations are between 30 and 60% correct, the C&M procedure suggests that the discriminations are made without awareness, whereas the SDT procedure suggests that awareness is present.[3] The SDT procedure shows no significant awareness for the 0–30% bin ($p > .20$), but does show clear awareness for all bins greater than 30% ($p < .05$, $p < .001$, and $p < .001$, respectively). The C&M procedure, on the other hand, shows no awareness for either the 0–30% bin or the 30–60% bin ($p > .10$ for both), but clear awareness for the 60–90% and 90–100% bins ($p < .01$ and $p < .001$, respectively).

This comparison between the two groups makes it clear that subjects can tell the difference between their correct responses and their errors, as assessed by the SDT procedure, even when they report that they are guessing in the C&M procedure. This

[3] From Table 2 and Fig. 3, it should be clear why it does not matter that most subjects completed the experiment by reaching the lower presentation limit. A threshold of discrimination is not necessary for the question we wish to answer, namely whether awareness assessments for the two procedures vary differently as a function of discriminative performance.

finding suggests that subjects will report that they are guessing even when they have some awareness and thereby supports the claim that the C&M procedure is susceptible to response biases (i.e., subjects in the C&M group systematically underestimate their discrimination performance). More importantly, this causes the two groups to produce quite different estimates of the relation between awareness and discrimination performance: When response biases are not controlled (C&M group), subjects appear to be aware of discriminative information only when discrimination performance is above 60% correct. However, when biases are controlled (SDT group), subjects appear to be aware even as discrimination performance approaches chance (25% correct).

In summary, the results of the first experiment indicate that response biases lower subjects' estimates of their own discrimination performance in the C&M procedure, consistent with repeated findings that subjects are underconfident in sensory discrimination tasks (Björkman et al., 1993). Subjects clearly have some ability to discriminate between correct responses and errors even when they claim to be performing at chance, indicating that such claims are contaminated by response bias. Furthermore, the presence of such bias obviously challenges the interpretation of qualitative differences in processing of strong versus weak prime stimuli in terms of processing with and without awareness (Cheesman & Merikle, 1986). Perhaps most importantly, this experiment illustrates how one can conveniently eliminate response biases from subjective awareness measures (i.e., confidence) by using signal detection techniques.

It should also be noted that the results from the SDT procedure provide little evidence one way or the other regarding the hypothesis of subliminal perception. In Table 2, the bin with discrimination performance at chance (i.e., 0–30%) shows no evidence of awareness (i.e., $d' \approx 0$), whereas the bins with discrimination performance above chance (30–60% and above) show clear evidence of awareness ($d' > 0$). On the surface, this pattern suggests that there is no discrimination without awareness—i.e., no subliminal perception. The experiment does not provide a fair test of the hypothesis, however, because it was designed to compare the SDT and C&M procedures. In particular, the bin widths are too large to give a clear picture of the relationship between awareness and discrimination performance at threshold, and there were insufficient data to support much narrower bins. Thus, the question of whether there is subliminal perception by the current definition was addressed with a slightly different procedure in Experiments 2–4.

EXPERIMENT 2

This experiment was designed to test directly for subliminal perception by comparing perception and awareness thresholds that were estimated separately for each of the four stimuli used in Experiment 1. Perceptual thresholds were estimated as the minimum SOAs needed to discriminate among stimuli, and awareness thresholds were estimated as the minimum SOAs needed to discriminate between correct and incorrect responses. If people can perceive without awareness, awareness thresholds should be higher than perception thresholds; if not, the two types of thresholds should be the same.

Creelman and Kaplan's (1973) Multiple PEST (Parameter Estimation by Sequential Testing) procedure was used to estimate thresholds.[4] PEST was selected for a number of reasons. First, it is efficient because it concentrates data collection around the point of most interest. This is particularly important in the present experiment because, as described further below, eight different threshold estimates were needed for each subject. Second, PEST was selected because of its ability to estimate arbitrary values along the psychometric function, including the point at which subjects just begin to perform at a level above chance. Not all methods have such flexibility (e.g., the two up–one down procedure described in Levitt, 1970). Third, PEST has well-defined rules for implementation (e.g., see Pollack, 1968) and produces statistically based estimates (cf. Harvey, 1986).

A potential problem is that PEST gives biased estimates under certain circumstances (Kunimoto, 1994; Taylor & Creelman, 1967), although it is less susceptible to estimation biases than other competing procedures (e.g., Emerson, 1986; Harvey, 1986; O'Regan & Humbert, 1989). The present experiment was designed to minimize the importance of such biases, however, by using maximally comparable procedures to estimate the thresholds for perception and awareness. With comparable procedures, PEST should introduce the same biases into both types of thresholds, so the difference between thresholds should be independent of bias (see Kunimoto, 1994, for computer simulations indicating nearly identical bias in estimated thresholds for perception and awareness using the procedures of the present Experiments 2–4). Thus, despite possible estimation biases, perception and awareness thresholds can be compared directly to test the hypothesis of subliminal perception.

Method

Subjects. Eighteen undergraduates completed the experiment for partial class credit and payment. For an additional five subjects, the experimental session was stopped because they did not finish after 3 h, and the data from these subjects were excluded from the analyses. The average session time for the 18 subjects who completed the experiment was 120 min. The shortest session was approximately 40 min, whereas the longest was 3 h.

Apparatus and stimuli. The equipment and stimuli from Experiment 1 were used in Experiment 2. The stimuli were plotted with fewer points (1 dot/1.43 mm), however, to reduce the problem of reaching the minimum presentation time of the equipment, as encountered in Experiment 1. This increase in grain size made stimulus identification more difficult and also decreased the minimum presentation time for a stimulus to 9 ms. The change in grain was not large enough to change the time needed to plot the fixation box; it remained at 3 ms. The mask was changed to a row of six pound signs, each measuring 0.62° (6.5 mm) vertically by 0.43° (4.5 mm) horizontally. The time to plot the mask was shortened to 11 ms.

[4] Multiple PEST is a variant of Taylor and Creelman's (1967) PEST procedure. The only difference between the two is that Multiple PEST allows for the estimation of several thresholds or points on the psychometric function simultaneously: the procedures are essentially the same. Throughout this article, references to *PEST* also apply to *Multiple PEST*, unless noted otherwise. Details of the present implementation of PEST are given in Appendix A.

Procedure. The actual trials were similar to the SDT trials of Experiment 1. A fixation box was displayed for 1 s, the stimulus was presented for a variable duration, and then the mask was displayed for 200 ms. After the subject made a keypress to indicate which stimulus was presented, the CRT prompted the subject to bet a chip (i.e., make a confidence response). After this second response, the CRT confirmed the chip bet for 250 ms and then presented a blank screen for 250 ms. The display unit then presented the fixation box for the next trial.

As in Experiment 1, subjects bet hypothetical chips to indicate their confidence. In this experiment, however, the number of chips allocated to subjects did not correspond to the number of trials in each block, and the time course of their replenishment was changed. In particular, two ‘‘stacks’’ of 20 chips were replenished whenever the total number of chips fell to 5. This allowed subjects some freedom of response near the end of each block. In the previous experiment, subjects were forced to choose ‘‘I was correct’’ (or ‘‘I was incorrect’’) for the last confidence response independent of how they actually felt about their response for that trial because the number of chips exactly matched the number of trials in each block. On the other hand, the new chip allocation scheme still forced subjects to make approximately 50% high confidence (‘‘I was correct’’) and low confidence (‘‘I was incorrect’’) responses, which is convenient for the signal detection analysis. Other minor changes included the following: black chips rather than green chips indicated ‘‘I was correct’’; hits were allocated 3 cents, misses 2 cents, false alarms −2 cents, and correct rejections −1 cent; the upper cash limit was raised to $4.00; and last, subjects took a minimum of 2 minutes’ break after the first three blocks and 2 minutes’ break after every four subsequent blocks to help reduce fatigue.

For each subject and each of the four stimulus words, two thresholds (SOAs in milliseconds) were estimated: one for discriminating among the stimuli (perception threshold) and one for discriminating correct responses from errors (awareness threshold). Trials were grouped into blocks of 48; 28 trials were used to estimate the awareness thresholds (7 trials for each stimulus word) and 20 trials were used to estimate the perception thresholds (5 trials for each stimulus word). These trials were presented in random order. The first block was considered practice and was not used to estimate thresholds. The trials used to estimate the perception and awareness thresholds were identical, apart from possible differences in presentation time.

The adaptive procedure used to determine perception and awareness thresholds for each of the four stimulus words was a version of Creelman and Kaplan’s (1973) Multiple PEST procedure, which is described in more detail in Appendix A. Multiple PEST uses the results from each trial to update a running estimate of performance for the stimulus presented and raises, lowers, or keeps constant the SOA for that stimulus to achieve or maintain the desired performance level. For trials used to estimate perception thresholds, the correctness of the subject’s discrimination response was input into Multiple PEST. The Multiple PEST procedure then updated the subject’s performance (in percentage correct) for the current run of trials (as demarked by the last change in SOA) and then adjusted the SOA (if needed) for the next trial. On the trials used to estimate awareness thresholds, extra steps were required in the process because Multiple PEST requires percentage correct (PC) as a measure of performance rather than d' (see Appendix A). Thus, for trials used to estimate awareness thresholds, the correctness of the subject's discrimination response was recorded, along with the subject's confidence. An experimental session was completed when Multiple PEST converged on all the thresholds.

Results and Discussion

Table 3 shows perception and awareness thresholds for each of the four words, averaged across subjects. An analysis of variance indicated that Threshold, the effect of most theoretical interest, was highly reliable [$F(1, 17) = 15.06$, $p < .001$]. Across words, perception thresholds ($M = 19.16$ ms, $SD = 3.36$ ms) were about 4 ms lower than awareness thresholds ($M = 23.39$ ms, $SD = 4.87$ ms), thus supporting the hypothesis of perception without awareness.

The main effect of Word was also significant [$F(3, 51) = 12.01$, $p < .001$], indicating that the thresholds for the words were different. This result is not surprising because the words differ along a number of physical dimensions, such as length and letter composition.

The interaction between Threshold and Word was also significant [$F(3, 51) = 4.83$, $p < .005$]. Simple main effects analyses indicate that the perception and awareness thresholds were significantly different for the stimuli *Blue* and *Green* [$F(1, 17) = 8.28$, $p < .01$ and $F(1, 17) = 9.45$, $p < .01$, respectively], but were not significantly different for *Yellow* or *Orange* [$F(1, 17) = 0.53$ and $F(1, 17) = 0.10$, both *ns*, respectively].

The most important finding in this experiment is the significant main effect of Threshold. A significant difference between perceptual and awareness thresholds provides strong support for the hypothesis of perception without awareness. Clearly, however, the effect is quite small because there is, on average, only a 4-ms range of SOAs in which there is evidence of discrimination without awareness (i.e., discrimination is above chance yet confidence is unrelated to accuracy).

Two methodological points concerning this comparison deserve emphasis. First, as we have emphasized, the present technique for measuring awareness is both phenomenologically valid and insensitive to response bias in confidence judgment (i.e., underconfidence). Second, perception and awareness thresholds were estimated si-

TABLE 3

Experiment 2: Mean Threshold (SOA in Milliseconds) as a Function of Threshold Measured (Perception and Awareness) and Word (*Blue, Green, Yellow,* and *Orange*)

	Threshold	
Word	Perception	Awareness
Blue	22.4	31.7
Green	17.2	23.9
Yellow	16.4	18.0
Orange	20.7	20.0
Average	19.2	23.4

multaneously and compared directly against one another instead of using the more common procedure of estimating thresholds in one phase of an experiment and looking for indirect effects of subthreshold stimuli in a later phase. This direct comparison has the obvious advantages that both perception and awareness are measured on the same scale and that extraneous variables (e.g., practice, and light adaptation) are equated across conditions.

The significant interaction between Threshold and Word implies that there is a qualitative difference between stimulus words, with the means suggesting that *Blue* and *Green* can be perceived without awareness but *Yellow* and *Orange* cannot. This would be a reasonable conclusion if there were a clear reason why certain words would be perceived without awareness and others would not. For example, McGinnies (1949) proposed that taboo words are suppressed by the perceptual system but ordinary words are not, which could cause subliminal perception of the former words but not the latter. We know of no obvious reason to expect subliminal perception of *Blue* and *Green* but not *Yellow* and *Orange,* however, and therefore suspect that this interaction arose from another source.

An alternative explanation for the apparent qualitative difference between words is that there were biases in discrimination responses. Subjects were not compelled to choose an equal number of each of the four possible responses, and they may have favored certain responses when guessing (e.g., a tendency to guess ''Green'' more than 25% of the time). Such response biases would increase the proportion of correct responses for a stimulus toward which the subject is biased and correspondingly decrease the proportion correct for other stimuli. As a consequence, response biases would reduce perception thresholds for favored stimuli and inflate perception thresholds for unfavored stimuli. Because the subliminal effect is indexed by the difference between perception and awareness thresholds, this would increase the subliminal effect for favored stimuli and decrease the effect for unfavored stimuli. Thus, it is possible that subjects were biased to respond ''Blue'' and ''Green'' (see Table 3), inflating the apparent subliminal effect for these two stimuli and eliminating a small, true subliminal effect for *Yellow* and *Orange.*

To check for the suspected response biases, we counted the number of times each of the four responses was used by each subject in the last 6 blocks of trials. If selected equally often, each response should have been selected 30 times (i.e., 5 discrimination trials per response for each block multiplied by 6 blocks). On average, however, subjects responded ''Blue'' 27.9 ($SD = 9.1$) times, ''Green'' 37.6 ($SD = 8.0$) times, ''Yellow'' 27.6 ($SD = 8.6$) times, and ''Orange'' 26.3 ($SD = 10.0$) times. A one-way ANOVA on these data indicated that significant response biases were present [$F(3, 51) = 4.6$, $p < .01$]. Clearly, the high observed frequency of ''Green'' responses suggests that the perception threshold was depressed for the stimulus *Green* and inflated for the other three stimuli, and it is possible that such inflation masked a true subliminal effect for *Yellow* and *Orange.*

This post hoc analysis of biases in discrimination responses raises doubt about the interpretation of the Word by Threshold interaction, but it clearly is not a completely satisfactory resolution of the issue. A better solution to the problem of discrimination response biases would be to eliminate their effects altogether, and this solution is implemented in Experiment 3. In addition, Experiment 3 addresses possible artifactual explanations, discussed next, of our main evidence for subliminal perception—that is, the finding that perception thresholds are lower, on average, than awareness thresholds.

Alternative Explanations for the Difference in Thresholds

In this section we consider three alternative explanations of Experiment 2's key finding, i.e., that thresholds for perception were lower than thresholds for awareness. Under certain circumstances, discussed below, this difference could have been created artifactually by some combination of (a) differences in the pairwise discriminabilities of the stimuli, (b) differences in the slopes of the psychometric functions for perception and awareness, and (c) judgment difficulty effects.

Differential pairwise discriminabilities. One alternative explanation for the evidence favoring perception without awareness is that there might have been unequal pairwise discriminabilities—or confusabilities—of the different stimuli used in the experiment. By way of example, we illustrate the effects of unequal pairwise discriminabilities in terms of an imaginary experiment employing the stimulus letters *E, F, O,* and *Q.* Note that not only are the pairs *E* and *F* similar to one another and *O* and *Q* similar to one another, but also, between one another, the two pairs are very dissimilar. If the letter *E* is presented on an experimental trial at an exposure duration very close to the subject's threshold, then one could imagine the following scenario unfolding (see Table 4).

For example, suppose that in each trial the subject was aware of seeing only the upper left-hand corner of the stimulus. This information would be sufficient to eliminate *O* and *Q* but insufficient to distinguish between *E* and *F.* Therefore, the subject

TABLE 4
Possible Example of a Stimulus Discriminability Artifact (Presenting *E* as the Stimulus near Threshold)

Discrimination response	Confidence ''Low''	Confidence ''High''	Total trials
''E'' (Correct)	25	25	50
''F'' (Incorrect)	25	25	50
''O'' (Incorrect)	0	0	0
''Q'' (Incorrect)	0	0	0

Note. The table shows the number of trials a subject might respond with each combination of identification response and ''Low'' and ''High'' confidence of 100 trials in which the stimulus letter *E* was presented. The subject's discrimination performance is above chance (50%) because the *E* is only confused with the *F,* never with the *O* or *Q.* The subject's confidence level does not predict accuracy (i.e., $d' = 0$), however. Thus, even though the subject might be fully aware that only *E* or *F* was presented, nevertheless, the measure of the subject's confidence (d') might not indicate that awareness.

would choose randomly between two acceptable candidates (e.g., ‘‘E’’ or ‘‘F’’) for the discrimination response and, in the long run, perform at 50% correct. Because this value is well above chance performance (25%) for the discrimination response, we would rightly conclude that some perception had taken place. Because the subject had to choose randomly between the two acceptable candidates, however, confidence would be unrelated to accuracy (i.e., both high and low confidence responses would be 50% correct). According to the proposed definition of awareness, the subject's inability to distinguish between correct and incorrect discrimination responses ($d' = 0$) would be taken as evidence that the subject was unaware of the stimuli. But this would clearly be an invalid conclusion, given our assumption that the subject *was* aware that the stimulus was not *O* or *Q*. In sum, this scenario illustrates that when stimuli are not equally discriminable, subjects' might have some awareness of the stimuli and yet be unable to discriminate between correct responses and errors. Within the proposed methodology, this situation could yield spurious evidence for perception without awareness.

Unfortunately, it is impossible to know for sure whether differences in stimulus discriminability caused the differences in the estimates of perceptual and awareness thresholds found in Experiment 2 because the pairwise discriminabilities of the stimuli were not equated. One of the purposes of Experiment 3, then, was to eliminate this potential artifactual explanation of the results. As discussed in the introduction to that experiment, this was accomplished by using a same–different task.[5]

Differential slopes of psychometric functions. A second possible alternative explanation of different thresholds for perception and awareness is that the psychometric function for perceptual judgments is much steeper than the psychometric function for awareness judgments. To take an extreme example, imagine that the threshold for perception is located at an SOA of 5 ms and that the psychometric function rises so steeply that performance is essentially perfect for SOAs of 6 ms or more. Imagine further that the threshold for awareness is also located at an SOA of 5 ms, but that the psychometric function rises so slowly that performance is very poor ($d' = 0.01$) at an SOA of 100 ms. Any threshold estimation procedure will virtually always estimate the threshold to be less than 6 ms for the steep perception function, even though the threshold estimate will vary slightly due to randomness inherent in the subject's responses. On the other hand, the estimation procedure will estimate the threshold to be much larger than 5 ms for the shallow awareness function, due to a combination of the truly low values of d' at SOAs of 6–100 ms, and the inherent randomness in responses. Kunimoto (1994) examined this potential source of bias using simulations and found that PEST does indeed give substantially greater threshold estimates, on average, when the underlying psychometric function is shallow than when it is steep.

[5] It might seem appropriate to eliminate the artifact by equating pairwise discriminabitilities, through selection or construction of stimuli that are all equally confusable with one another on the average across trials. Unfortunately, equating average pairwise discriminability would not be sufficient to eliminate the problem. The artifact arises when the subject has to guess among a subset of the possible responses, even if a given stimulus gives rise to different subsets on different trials (e.g., due to variations in the exact location of the subject's fixation or the momentary sensitivity of individual neurons in the visual cortex). Thus, the stimuli could be equally discriminable from one another on the average, but unequally discriminable on individual trials, and the artifact might still contaminate the results.

It is unclear, however, whether this artifact actually operated in Experiment 2 because the slopes of the psychometric functions for perception and awareness were not measured. The slopes were measured in Experiment 3, so that this possible artifact could be evaluated. The logic was this: If slope measurements indicate that the psychometric function is much steeper for perception than for awareness, then it is difficult to know whether an observed difference in threshold estimates is real or an artifact of the slope difference. But if slope measurements indicate that the two psychometric functions are approximately equally steep or that the awareness function is steeper, then it is possible to reject this artifactual explanation of the finding that the threshold for perception is lower than the threshold for awareness.

Judgment difficulty effects. A third alternative explanation is that the perception–awareness difference is a natural consequence of the different types of information indexed by the discrimination and confidence judgments, not a reflection of any dissociation between conscious and nonconscious perceptual processes. As discussed in detail in Appendix B, discriminating among stimuli is not the same as discriminating correct responses from errors, even to an ideal observer using precisely the same information to make both judgments (Galvin, 1988; Galvin, Podd, Drga, & Whitmore, 2000). In fact, there is a very real sense in which discriminating among stimuli is the easier of the two judgments. Thus, using the present methods we might well expect that under some circumstances estimates of the perceptual threshold would be lower than estimates of the awareness threshold even for a completely mechanical observer for whom there is really no possible distinction between conscious and unconscious perceptual processing.

As it turns out, an ideal observer would produce evidence for perception without awareness only when the psychometric function relating discrimination ability to SOA is fairly shallow (see Appendix B for details). Thus, estimation of psychometric functions, which is included in Experiment 3 in order to evaluate the alternative explanation in terms of slope differences, can also be used to address the alternative explanation in terms of judgment difficulty effects.

EXPERIMENT 3

This experiment was conducted primarily to address the three methodological issues raised in connection with Experiment 2. Specifically, the evidence for perception without awareness found in that experiment could have been artifactual, resulting from differences in stimulus discriminability, differences in the slopes of the psychometric functions for perception and awareness, or judgment difficulty effects. In addition, the possibility of biases in the discrimination response may have led to the false conclusion that perception without awareness can be elicited by some words and not others. The experiment presented here incorporates solutions to all of these problems.

The differential stimulus discriminability problem was solved by changing the 4-AFC identification task to a same–different task. On each trial, a target stimulus was presented briefly and then masked. Then, a comparison word was presented, and subjects decided whether the stimulus and comparison were the same. If same and different comparison stimuli are presented equally often, then chance performance on this same–different task should be 50%. Performance above 50% would indicate perception of the stimuli.

The same–different task avoids the differential stimulus discriminability problem. Recall that this problem occurs in the *N*-AFC task because subjects do not have the opportunity to give high confidence responses about stimulus alternatives that they have consciously rejected. In the same–different task, however, a subject can give a high confidence ''different'' response when the randomly selected comparison is a stimulus that the subject knows was not presented. For example, consider the stimulus set *E, F, O, Q* (discussed previously), in which *E* is presented at a duration for which *O* and *Q* can be ruled out, but not *F*. When *O* or *Q* are presented as comparison stimuli, the subject can respond ''different'' correctly and with ''high confidence.'' In the 4-AFC task, however, the subject would never select *O* or *Q* as stimulus alternatives and thus would never have an opportunity to show awareness that the stimulus presented was *not O* or *Q*. Thus, unlike the subject in the 4-AFC task, the subject in the same–different task can indicate awareness of the stimulus via his or her confidence judgments even if the stimuli are unequally discriminable.

In this experiment the slopes of the near-threshold psychometric functions for perception and awareness were also estimated to evaluate the contributions of differential slope artifacts and judgment difficulty effects. Slopes were found by estimating SOAs corresponding to performance levels of 10 and 20% correct, adjusted for guessing, from which slopes of the psychometric functions can be estimated. Because Multiple PEST is relatively unbiased when estimating exposure durations for above-chance performance levels, it should be possible to estimate these SOAs fairly accurately, and slopes can then be estimated from the corresponding SOAs. For example, if a subject requires 25 ms to perform at 10% correct and 35 ms to perform at 20% correct, then the slope of the psychometric function between these two points is $(20 - 10)/(35 - 25) = 1.0\%/\text{ms}$.

Estimates of the near-threshold slopes of the psychometric functions for perception and awareness can address the potential slope difference artifact directly. The artifact arises because threshold estimates tend to be inflated for shallow psychometric functions. Thus, if the psychometric function for perception is at least as shallow as that for awareness, the thresholds for the two functions can be directly compared, without worrying about the differential slope artifact. If the psychometric function for perception is steeper, however, the effect of the slope difference on threshold estimates will have to be assessed further (e.g., by computer simulation) to see whether the effect is large enough to account for the results.

Slope estimates can also be used to evaluate the possibility of judgment difficulty effects. As described in Appendix B, an ideal observer could produce higher thresholds for awareness than for perception, using the current operational definitions, if perceptual discriminability (i.e., d') increased slowly with SOA. Clearly, this possibility can be evaluated at least informally by examining the rate of growth in discrimination accuracy as a function of SOA.

The values of 10 and 20% correct were chosen as the target performance levels to use in obtaining slope estimates because these values are themselves quite close to threshold. Although one could obtain more stable slope estimates by using more widely spaced target performance levels (e.g., 10% versus 90%), narrowly spaced target levels were nonetheless used for two reasons. First, it is only the slope near threshold that is relevant in the present case because this is where the potential artifacts would arise. Second, the slope is likely to vary across such a wide performance range (i.e., the psychometric function is probably not linear), so a slope estimate obtained across a wide range of performance levels would not accurately reflect the local slope in the region just above threshold.

The final advantage of the same–different task is that biases in discrimination responses will no longer create apparent differences among stimuli. Subjects cannot be biased to respond ''blue,'' for example, because ''blue,'' is no longer a possible response, so this experiment will provide a better test of whether subliminal perception varies across stimulus words. There is the possibility, of course, that subjects could be biased to respond ''same'' more often than ''different'' or vice versa. To avoid any problems that such biases may introduce, the same–different responses, like the confidence judgments, can be analyzed with techniques from signal detection theory. In particular, hits and misses can be computed from trials on which the stimulus and the comparison are the same, whereas false alarms and correct rejections can be computed from trials on which the stimulus and the comparison are different (see Table 5). In any case, such biases will not be problematic because they will not create differences between stimulus words.

In this experiment, the four stimulus words used in Experiment 2 were split into two pairs. The words *Blue* and *Green* were presented to one group of subjects, and the words *Yellow* and *Orange* were presented to another group of subjects. This was done because it would have been impractical to obtain 24 separate estimates for each subject in a single experimental session (i.e., estimates of three points on the psychometric function for perception and three points on the psychometric function for awareness for each of the four stimulus words). With only two stimulus words with each subject, only 12 estimates per subject were needed. Thus, for each stimulus word, perception and awareness thresholds were estimated to assess whether the words could be perceived without awareness, and SOAs corresponding to 10 and 20% performance levels (corrected for guessing) were also estimated for perception and awareness, for each word, to assess slope differences.

In summary, incorporating the same–different task and the slope measurements allows us to eliminate the potential artifactual explanations of subliminal perception raised in connection with Experiment 2. The basic logic of this experiment is quite

TABLE 5
Categorization of Hits, Misses, False Alarms, and Correct Rejections for the Same–Different Task

Comparison stimulus	Same–different response: ''Same''	Same–different response: ''Different''
Stimulus 1	Hit	Miss
Stimulus 2	False alarm	Correct rejection

Note. These categorizations apply to trials in which Stimulus 1 is presented as the target stimulus. When Stimulus 2 is the target, the categorizations are the same except that the roles of comparison stimuli 1 and 2 are exchanged.

similar to that of the previous one, however, despite these improvements in design. If there really is subliminal perception, then the estimated perception thresholds should again be lower than the estimated awareness thresholds. If there is no subliminal perception (i.e., if the evidence of subliminal perception obtained in Experiment 2 resulted from one of the artifacts), however, then there should be no difference between perceptual and awareness thresholds.

Method

Subjects. Half of the 64 undergraduate subjects participated to partially fulfill a course requirement, 31 participated for $12, and 1 participated for a combination of course credit and $6. As in Experiment 2, the length of each session varied across subjects because Multiple PEST was used to estimate the desired points on the psychometric functions. Subjects were scheduled to participate for 2 h; actual sessions ranged from 45 min to 2 h, with an average session time of 82 min. Session times were reduced to 2 h because subjects often reported fatigue when scheduled for 3 h, even when encouraged to take breaks. This reduction in session time resulted in a much higher mortality rate than in Experiment 2. The data from 22 additional subjects were excluded from the analyses because they did not complete the experiment in the scheduled 2-h session. Of these 22 subjects, 1 subject had thresholds lower than the minimum presentation limit of the equipment. The data from 2 additional subjects were excluded because the experimenter later discovered that they had participated in an earlier subliminal perception experiment, and the data from 1 other subject were excluded because the subject had forgotten to bring her corrective lenses and could not see clearly during the experiment.

Apparatus and stimuli. The equipment and stimuli from the previous two experiments were used in Experiment 3. The stimuli were constructed as in Experiment 2. For both groups of subjects, the keys *Z* and *?* corresponded to ''same'' and ''different,'' respectively, and the keys *S* and *:* corresponded to low and high confidence, respectively.

Design and procedure. Two groups of subjects were tested: Group 1 was presented with the words *Blue* and *Green,* and Group 2 was presented with the words *Yellow* and *Orange.* For each subject in each group, three points on the psychometric function were estimated (0, 10, and 20%, corrected for guessing) for each function (perception and awareness) and each of two stimulus words. The points corresponding to 0% were used to test for subliminal perception, whereas the points corresponding to 10 and 20% were used to test for slope differences.[6]

Thus, the design was a 2 Group (Group 1 versus Group 2) × 2 Word (*Blue* and *Green* in Group 1 and *Yellow* and *Orange* in Group 2) × 2 Psychometric Function (Perception versus Awareness) nested design with Word nested in Group. Two separate analyses were planned: one for the Thresholds (corresponding to 0% on the psychometric functions) and one for the slopes of the psychometric functions (corre-

[6] Although slopes can be estimated by using the points corresponding to 0 and 10% performance, this was not done because of the PEST biases for estimates at the 0% threshold (Kunimoto, 1994).

sponding to the difference between the 20 and 10% points on the psychometric functions).

Each block consisted of 48 trials, of which 24 were used to estimate points on the psychometric function for perception and 24 were used to estimate points on the psychometric function for awareness (4 trials per stimulus word for each target PC). These trials were presented in random order without replacement. The first block was considered practice and was not used in the estimation of points on the psychometric functions. As in Experiment 2, subjects were forced to take at least 2 minutes' break after the first three blocks and after every four subsequent blocks in order to reduce fatigue.

A same–different task was used in this experiment instead of the 4-AFC task of Experiment 2. Thus, the trials were presented as follows. First, a fixation box was displayed on the Tektronix 602 display unit for 1 s. Then the stimulus was presented for a variable duration (determined by the Multiple PEST procedure) followed immediately afterward by the mask for 200 ms. A comparison word, selected randomly with replacement, was then presented on the CRT. After the subject indicated whether the comparison was the same or different from the stimulus word, the CRT prompted the subject for a confidence response. The confidence response was confirmed for 250 ms, followed by a 250-ms blank screen. The display unit then presented the fixation box for the next trial.

Unlike Experiments 1 and 2, subjects did not ''bet'' hypothetical chips to indicate their confidence. Instead, subjects directly made ''High Confidence'' and ''Low Confidence'' responses. Because these responses were equated with the chip bets in the previous experiments, and because it took significant time to instruct the subjects on betting chips, it was felt that these direct statements would be easier to understand and thus save time and avoid possible confusion during the instruction phase of the experiment.

Despite the absence of chips, subjects were still paid based on their responses as an incentive for good performance. As in the previous experiment, they were shown a payoff matrix, but instead of having rewards in cents, points were awarded or decremented based on their responses: hits were +3 points, misses +2 points, false alarms −2 points, and correct rejections −1 point. The experimenter emphasized that it was always best to make a correct same–different response because it was associated with positive points and that it was always best to have the confidence level ''match'' the correctness of the same–different response: ''High'' confidence when correct and ''Low'' confidence when incorrect. At the end of each block of trials, the number of points accumulated for that block, each previous block, and the total number of points accumulated were given as feedback to the subjects. Subjects were told that they would receive up to $1 as an incentive, dependent on their performance. Unfortunately, because the PEST procedure automatically adjusts stimulus durations depending on performance, any simple performance evaluation was impractical to implement. Thus, at the end of each experimental session, the computer randomly generated a dollar amount between $0.76 and $1.00, and the experimenter rounded that value up to the nearest quarter: thus, all subjects were given a dollar at the end of their experimental session, in addition to the class credit or payment they were expecting to receive.

The number of times subjects could respond ‘‘High Confidence’’ and ‘‘Low Confidence’’ was changed in order to account for the nonthreshold values that were also being estimated. Subjects could respond ‘‘High Confidence’’ up to 28 times and ‘‘Low Confidence’’ up to 24 times in each block of trials. Subjects did not have to remember how many times they had selected each response during the course of a block because the CRT displayed the number of remaining high and low confidence responses the subject was allowed to make whenever the computer prompted subjects for a confidence response. The number of ‘‘High Confidence’’ and ‘‘Low Confidence’’ responses subjects could make was reset to 28 and 24, respectively, at the beginning of each block of trials.

The experimenter emphasized that the high and low confidence responses should be made relative to each other. Subjects were told that at some point in the experiment they may wish to respond ‘‘Low Confidence’’ for almost all of the trials. But because the computer will not let them do this, they should respond ‘‘High Confidence’’ on those trials in which they had ‘‘More Confidence’’ such that they were responding ‘‘High Confidence’’ and ‘‘Low Confidence’’ about equally often. This idea of relative confidence responses was made more concrete by giving subjects a continuum analogy: They could think of their confidence as a continuum from absolutely no confidence to complete confidence, and they should set a criterion such that they responded ‘‘High Confidence’’ and ‘‘Low Confidence’’ approximately equally often.

The experimenter emphasized the accuracy of responses and not response time. Subjects were told to do their best on each trial, including those trials in which they thought they did not see anything, because words were flashed on every trial. The experimenter suggested that they should go by their ‘‘feelings’’ whenever they thought they didn’t see anything.

The first block of trials was practice, and all stimulus words were flashed for 100 ms so that subjects could clearly see the stimuli. After the first block of trials, a ‘‘quick descent’’ procedure was implemented in order to quickly find an SOA near each subject’s thresholds: Multiple PEST was used to converge on 60% correct same–different performance, pooling across stimuli. Once this target was reached, Multiple PEST was implemented for each stimulus and each target point on the psychometric functions for perception and awareness. The preliminary quick descent procedure was implemented in order to improve the starting points for Multiple PEST and thereby reduce the total duration of the experiment.

Both perceptual and awareness responses were analyzed with signal detection techniques before being input into the Multiple PEST procedure. This followed the same procedure used in Experiment 2 for the awareness responses. For the points on the function for perception, the correctness of the subject’s same–different response was recorded; a d' for perception was computed and this d' was converted to PC by using the conversion table given by Elliot (1964); it was this PC that Multiple PEST used to adjust the SOA for the next trial. Similarly, for the points on the function for awareness, both the correctness of the subject’s same–different response and the subject’s confidence in making that response were recorded; a d' for confidence was computed (as described in the introduction); and then this d' was converted to PC and input into Multiple PEST. An experimental session was completed when Multiple PEST converged on all the targeted points on the psychometric functions. Thus, for each subject and each of the two stimulus words, 3 points on the psychometric function for perception and awareness were estimated corresponding to target PCs of 0, 10, and 20% (corrected for guessing).

Results and Discussion

Table 6 shows the mean SOAs for perception and awareness as a function of Target PC for each stimulus word (*Blue* and *Green* for Group 1 and *Yellow* and *Orange* for Group 2). Interestingly, virtually identical average SOAs were obtained for Target PCs of 0, 10, and 20% correct, both for perceptual judgments and for awareness judgments. This indicates that both psychometric functions are extremely steep in the near-threshold region and thus allays any concerns over the possibility of judgment difficulty effects (Appendix B). Further evidence against an explanation in terms of judgment difficulty effects is given below.

Close inspection of the SOAs within the table indicates that SOA does not increase

TABLE 6
Experiment 3: Mean SOAs (in Milliseconds) and Standard Errors in Parentheses for Perception and Awareness as a Function of Target PC (0, 10, and 20%) for Group 1 (*Blue* and *Green*) and Group 2 (*Yellow* and *Orange*)

	Target PC		
Stimulus	0%	10%	20%
Group 1			
Blue			
Perception	18.9 (1.4)	18.5 (1.6)	19.0 (1.6)
Awareness	22.1 (1.8)	24.2 (2.0)	22.8 (1.9)
Green			
Perception	17.5 (1.1)	17.5 (1.1)	16.4 (1.1)
Awareness	19.0 (1.7)	18.6 (1.4)	19.7 (1.9)
Group 2			
Yellow			
Perception	21.1 (1.4)	20.2 (1.3)	21.5 (1.5)
Awareness	23.3 (1.6)	21.8 (1.3)	21.8 (1.5)
Orange			
Perception	16.0 (0.7)	17.1 (1.3)	17.2 (0.8)
Awareness	20.5 (1.5)	19.2 (1.1)	22.2 (1.6)
Average			
Perception	18.4	18.3	18.5
Awareness	21.2	21.0	21.6

systematically with Target PC for every word and threshold type; presumably, this reflects random error in the process of estimating the desired SOA. Because there is apparently a very small true difference between the SOAs at Target PCs of 10% versus 20% (i.e., because the psychometric functions are very steep), it is not surprising that random error could produce nonmonotonicities in the relationship between SOA and Target PC for specific words.

The fact that the psychometric functions are very steep for both perception and awareness also indicates that the slope difference artifact is not a concern. To conduct a formal statistical test for slope differences, the SOA differences between the 10 and 20% points were computed for each psychometric function (perception and awareness) and each stimulus in the two groups. A 2 Group × 2 Word (within Group) × 2 Psychometric Function (Perception versus Awareness) ANOVA on these differences yielded no reliable main effects or interactions (all p's > 0.10), consistent with the hypothesis that there are no differences in slope between the psychometric functions for perception and awareness.

Because neither the slope difference artifact nor judgment difficulty effects seems to be a problem, the thresholds for perception and awareness were compared directly. The 2 Group × 2 Word (within Group) × 2 Threshold ANOVA yielded a reliable main effect for Threshold [$F(1, 62) = 11.21$, $p < 0.01$]. The mean perception threshold was 18.4 ms, whereas the mean awareness threshold was 21.2 ms; moreover, this difference is also clear at Target PCs of 10 and 20%. This result provides further support for the hypothesis of perception without awareness.

In addition to the difference in thresholds, there was also a significant Word (within Group) effect [$F(2, 62) = 9.29$, $p < 0.01$]. In Group 1, the mean threshold for *Blue* was 20.5 ms, whereas the mean threshold for *Green* was 18.3 ms. In Group 2, the mean threshold for *Yellow* was 22.2 ms, whereas the mean threshold for *Orange* was 18.3 ms. This pattern is consistent with the results found in Experiment 2.

Finally, the Group × Threshold interaction was not significant [$F(1, 62) < 1.00$, $p > .10$]. Thus, there was no evidence that the difference between perception and awareness thresholds varies across groups. This result, in combination with the significant effect of Threshold, is consistent with the hypothesis that subliminal perception does *not* vary across the stimulus words. On the basis of this result, although we cannot rule out the hypothesis that there are some small word-to-word differences in the magnitude of subliminal perception, it seems safest to conclude that the differences found in Experiment 2 were probably due to discrimination response biases (Experiment 2).

With average perceptual thresholds approximately 3 ms lower than average awareness thresholds, there should be a small range of SOAs within which perceptual discriminations are above chance but awareness judgments are at chance. Although the present experiment was not explicitly designed to monitor performance in these intervals, it is possible to examine such performance in a post-hoc fashion. For each subject and word, we first determined the SOA corresponding to the awareness threshold, SOA_{AT}. We then identified all the trials for that subject and word in which the SOA had been in the range of SOA_{AT} to $SOA_{AT} - 3$ ms. Each of these trials was categorized with respect to the accuracy of its discrimination response, and a d' for perceptual discriminations was computed for this set of trials. The average d' value thus computed was 0.59, which is significantly different from zero [$t(63) = 5.06$, $p < .01$]. Thus, these data support the view that subjects do have some perceptual sensitivity for stimuli below the awareness threshold. The awareness d' value computed for this same set of trials was −0.06, as is to be expected because these trials were selected for having SOAs just below the awareness threshold.

As shown in Table 7, the d' values for perceptual discriminations are also greater in sets of trials in which SOAs were 4–7 ms below the awareness threshold, with no evidence of awareness for this set of trial either. With small-enough exposure durations (8–11 ms below the awareness threshold), of course, perceptual sensitivity does clearly decline. It is perhaps surprising that perceptual sensitivity is so stable over the range of SOAs 0–7 ms below the awareness threshold. Based on an average difference in thresholds of 3 ms, one might predict reasonably high perceptual d's for SOAs within 3 ms of the awareness threshold and virtually zero d's for SOAs lower than that. And it is certainly possible that the true psychometric functions would show that pattern, with the different pattern in Table 7 arising by chance. Another possibility, however, is that the average subliminal effect—a difference in threshold SOAs of 3 ms—is produced by a mixture of cases with and without a larger subliminal effect. The present results, for example, are all consistent with the hypothesis that the subliminal effect is 10 ms for one-third of the subjects and 0 ms for the other two-thirds.

The fact that perceptual discrimination is considerably above chance with SOAs at or just below the awareness threshold (cf. Table 7) provides further evidence that the results are not due either to differences in the slopes of the psychometric functions or to judgment difficulty effects. These two potential artifacts both involve misestimation of the true threshold by the PEST procedure. The results in Table 7, however, represent a direct comparison of perceptual discrimination and awareness responses in a fixed set of trials with a given range of SOAs, so they are independent of the PEST procedure.

In summary, Experiment 3 provides strong support for the hypothesis of perception without awareness. Even controlling for sources of artifact such as differential stimulus discriminability, differences in psychometric functions slopes, and judgment difficulty effects, the evidence indicates that subjects can discriminate among stimuli

TABLE 7

Experiment 3: Perception Performance and Awareness Level as a Function of SOA at and below Subjects' Awareness Thresholds

	Performance type	
Exposure duration	Perception	Awareness
AT–0 ms to *AT*–3 ms	0.59	−0.06
AT–4 ms to *AT*–7 ms	0.44	0.02
AT–8 ms to *AT*–11 ms	0.11	0.04

Note. Measures of perception performance and awareness levels were pooled across subjects and stimulus words. AT = Awareness Threshold.

even when they have no idea which of their discrimination responses are correct and which are errors. Furthermore, using a same–different task so that discrimination response biases would not be a factor, there was no evidence that subliminal effects varied across stimulus words.

EXPERIMENT 4

Experiment 4 was designed to serve two functions: (1) to replicate the finding that perceptual thresholds are lower than awareness thresholds even when possible artifacts are controlled and (2) to extend this finding to a task with more than two stimulus alternatives. This experiment was therefore a replication of Experiment 2, except that the same–different perceptual task was substituted for the 4-AFC task to control for the differential discriminability artifact. Given the steep and approximately equal slopes found in Experiment 3, it seemed reasonable to assume that artifacts due to differential slopes or judgment difficulty would be unlikely in a further experiment with the same stimuli. Thus, Multiple PEST was used to estimate eight thresholds (four stimulus words $\times$ perception vs awareness) for each subject, and slopes were not estimated.

Method

Subjects. Nine of the 36 undergraduate subjects participated to partially fulfill a course requirement and 27 subjects participated for $12. As in Experiments 2 and 3, the length of each session varied across subjects because Multiple PEST was used to determine thresholds. Subjects were scheduled to participate for 2 h, and sessions ranged in duration from 40 to 120 min. The data from 5 additional subjects were excluded from the analyses because they did not complete the experiment in the scheduled 2 h. The average session time for the 36 subjects who completed the experiment was approximately 1 h and 10 min.

Apparatus and stimuli. The equipment and stimuli from the previous experiments were again used in Experiment 4. The stimuli were constructed as in Experiments 2 and 3, and the mapping of responses to keys followed that of Experiment 3.

Procedure. The Multiple PEST procedure was again used to estimate perception and awareness thresholds for each of the four stimulus words. The sequence of events in a single trial was the same as that in Experiment 3. There were four possible stimulus words rather than two, however. The only other change was the number of each type of confidence response the subject could make in each block of trials. Because only thresholds were estimated in this experiment, subjects were allowed to respond ‘‘High Confidence’’ up to 26 times and ‘‘Low Confidence’’ up to 26 times in each 48 trial block.

Results and Discussion

Average thresholds for perception and awareness are shown in Table 8, and it is clear that the mean perceptual threshold is again lower than the mean awareness threshold [20.9 vs 23.4 ms, $F(1, 35) = 9.54$, $p < .005$]. This result provides further support for the hypothesis of perception without awareness, consistent with the evi-

TABLE 8

Experiment 4: Mean Threshold (SOA in Milliseconds) as a Function of Threshold Measured (Perception and Awareness) and Word (*Blue, Green, Yellow,* and *Orange*)

	Threshold	
Word	Perception	Awareness
Blue	23.2	23.5
Green	18.9	24.3
Yellow	21.7	24.4
Orange	19.7	21.4
Average	20.9	23.4

dence found in Experiments 2 and 3. In addition, this result confirms that subliminal perception can be obtained using the present technique in experiments with stimulus sets containing more than two items.

The mean thresholds of the four stimuli were marginally significantly different [$F(3, 105) = 2.23$, $p < .10$], possibly reflecting different perceptual sensitivities to the different words. As in Experiment 2, the perceptual threshold was longest for the word ''blue,'' possibly because this word is made up of the fewest letters. In addition, the size of the subliminal effect appears to vary somewhat with the stimulus word presented [$F(3, 105) = 2.16$, $p < .10$], and the means suggest that the stimulus *Blue* displays little or no subliminal effect, unlike the other stimuli. Given that *Blue* has consistently displayed a subliminal effect in the previous experiments, however, we are inclined to attribute this pattern to random variation.

GENERAL DISCUSSION

The question of the existence of subliminal perception has had a very controversial history, especially because of disagreements about what constitutes an acceptable operational definition of awareness (for recent discussions see, e.g., Henley, 1984, and Merikle, 1984). Researchers favoring subjective definitions have generally left the definition to the subjects, with the result that awareness assessments may have been contaminated by response biases. Researchers favoring objective definitions, on the other hand, have used measures that not only eliminated response bias but also generally eliminated the phenomenal experience of the subject.

This article has developed an operational definition of awareness that combines the virtues of both objective and subjective definitions. Using the techniques of signal detection theory to monitor the relation between confidence judgments and response accuracy, we suggest that the subject's ability to discriminate correct responses from errors can be used as an indication of the subject's awareness of the stimuli. The reliance on the subjects' confidence judgments makes the new operational definition sensitive to subjects' phenomenological states. At the same time, the signal detection analysis filters out the response biases that have previously plagued phenomenologically valid subjective reports.

In Experiment 1 the proposed SDT-based definition was compared with the subjective awareness criterion used by Cheesman and Merikle (1984, 1986). Awareness of performance, as assessed by each of the two definitions, was determined relative to actual performance in a word discrimination task. When subjects' discrimination performance was between 30 and 60%, subjects had some awareness of the stimuli according to the SDT definition, but had no awareness according to Cheesman and Merikle's definition. Because the main difference between the two definitions is the removal of response biases by the SDT definition, this experiment strongly suggests that response biases can significantly contaminate subjects' estimates of their own overall performance. As proponents of objective definitions have long argued (e.g., Eriksen, 1960), then, an acceptable operational definition of awareness must take such response biases into account.

In Experiment 2, perceptual thresholds were compared directly against awareness thresholds measured with the new operational definition. The former were found to be reliably lower, supporting the hypothesis of perception without awareness. With that design, however, there were three potential alternative explanations of the evidence for perception without awareness: (a) differences in pairwise stimulus discriminabilities, (b) difference in the slopes of the psychometric functions for perception and awareness, and (c) judgment difficulty effects.

Experiment 3 incorporated two major modifications designed to address these alternative explanations. First, the perceptual task was changed from word identification to same–different discrimination because this would eliminate the effects of differences in pairwise discriminability. Second, the slopes of the psychometric functions for perception and awareness were estimated. Both functions were quite steep, and approximately equally so, suggesting that neither the differential slope artifact nor judgment difficulty effects constitute a problem with these stimulus displays. As in Experiment 2, perceptual thresholds were found to be reliably lower than awareness thresholds for all stimulus words, providing additional support for the hypothesis of perception without awareness under circumstances where the three alternative explanations can be ruled out. Experiment 4 replicated the effect obtained in Experiment 3, again using the same–different task to eliminate the differential discriminability artifact, and extended the effect to the task with four rather than two alternative stimuli.

At a theoretical level, the present results provide strong new support for the hypothesis of subliminal perception. In some sense this support is not surprising because there is a growing consensus that at least some phenomena indicative of unconscious perceptual processing are real (e.g., Greenwald, 1992). Nonetheless, the question of how best to document and study such phenomena is still hotly debated (see, e.g., Dosher, 1998; Greenwald & Draine, 1998; Greenwald et al., 1995; Merikle & Reingold, 1998; Miller, 2000; Reingold & Merikle, 1988). A major contribution of the present report, then, may be methodological in showing how to dissociate perception from awareness in the very specific sense that people can discriminate among stimuli at better than chance levels even with displays so brief that their confidence is unrelated to their accuracy. This dissociation of confidence and accuracy suggests that subjects are simply guessing, as far as they know, and that they are therefore unaware of any discriminative information that they might be extracting.

At a more practical level, it seems reasonable to consider the magnitude of the subliminal effect documented here. With discrimination thresholds only 3 ms or 15% less than awareness thresholds (Experiments 2–4), one might be tempted to conclude that the effect is rather small. But this would be a simplistic conclusion because it ignores the question of how much 3 ms represents to the sensory and perceptual systems. Maybe quite a lot of information can be obtained in an extra 3 ms.

One way to judge the importance of the extra 3 ms is to look at performance in the discrimination task with displays presented at the awareness threshold. In Experiment 3, for example, we selected trials presented at or just below the awareness threshold and computed d' for the discrimination judgments made in those trials (Table 7). When subjects could no longer discriminate correct responses from errors, they still obtained discrimination task d's of approximately 0.6, corresponding to approximately 61% correct in a two-choice task. Inasmuch as 61% is not a lot better than chance, it seems fair to conclude that perceptual judgments do not actually reflect a lot of information in these trials. Clearly, this information disappears rapidly as SOA is decreased, but that may say more about the rapid dropoff in sensory sensitivity than about the size of the subliminal effect. Surely, there could well be other display conditions—yielding flatter psychometric functions—that would produce differences between awareness and perceptual thresholds much larger than the 3 ms effect obtained here.

Finally, the method employed in the present research may have important applications in the study of consciousness throughout various areas of cognitive psychology. Past researchers have often relied on their own experience when speculating about the nature and functions of consciousness. Indeed, according to Allport (1988) some researchers ''refer to *the* phenomenon of consciousness . . . accompanying their words with some sort of vigorous gesture, as though pointing inwards towards their own mental life, their own 'phenomenon of consciousness.' '' But the problem with this approach, as Allport points out, is that ''[there is] no clear conception what people are talking about when they talk about 'consciousness' or 'phenomenal awareness' . . . .''

The experiments presented here illustrate that there is at least one measurable criterion of awareness—the relation between confidence and accuracy—that can have both phenomenological appeal and methodological rigor. Fortunately, the definition of awareness proposed here is relatively general: Its use requires only the collection of both performance data and confidence responses. The definition was applied here in very simple visual tasks. It would also be possible, however, to apply the definition to auditory perception, memory, or other cognitive tasks in order to determine the extent to which various types of performance are carried out with or without awareness.

For example, Treisman and her colleagues have argued that a serial, limited capacity attentional device, presumably conscious, is required to perceive conjunctions of features, such as color and form, but is not required to perceive single features (e.g., Treisman & Gelade, 1980). On this view, one might imagine that with sufficiently brief displays subjects might be able to discriminate featural stimuli even though their confidence is unrelated to their accuracy, but that this should not be possible for discrimination of conjunction stimuli. As a second example, subjects in studies of implicit learning sometimes seem to acquire the ability to control complex systems without acquiring any explicit knowledge of how the systems work (e.g., Berry & Broadbent, 1988). The lack of explicit knowledge is often inferred from subjects' inability to describe the rules underlying the system. Given that subjects often have difficulty verbalizing their knowledge (e.g., Berry, 1990), a stronger test might be to see whether subjects' confidence in their specific actions predicts the success of those actions. If subjects are truly learning to control the system without awareness, then it is hard to see how they could know in advance which of their actions would be most likely to succeed.

One particularly interesting application was proposed by Weiskrantz (1992). Citing an earlier version of this manuscript (Kunimoto, Miller, & Pashler, 1990), Weiskrantz noted that the definition of awareness proposed here is not limited to human cognition but could even be used for the measurement of awareness in animals. When subjects are asked to make a confidence judgment about their perceptual response they are essentially making an independent, two-choice ''commentary'' about their response. Weiskrantz speculates that this independent commentary can be trained into an animal's repertoire to assess whether the animal is aware of what it is perceiving. In fact, just such an independence of responses (forced-choice discriminations versus commentary responses on the other) was proposed by Weiskrantz (1986) for use with monkeys and other nonlinguistic creatures. Thus, it seems that studies of the relation between confidence and accuracy can be quite revealing in general about conscious access to the information on which various types of judgments are based.

APPENDIX A

Implementation of the PEST Procedure

Creelman and Kaplan's (1973) psychophysical procedure, Multiple PEST (Parameter Estimation by Sequential Testing), was used to estimate thresholds and other points on the psychometric functions in Experiments 2–4. Multiple PEST is a multi-threshold version of Taylor and Creelman's (1967) PEST procedure, and it was designed to estimate psychometric function values efficiently by concentrating data collection near the desired performance level. In the present experiments, PEST was used to determine the stimulus-mask onset asynchrony (SOA) for each trial. PEST adjusted SOA over a full experimental session to find the SOA corresponding to a particular target performance value on a psychometric function, as follows.

A starting SOA was selected by the experimenter, and a set of trials was run using that SOA. The initial SOA used by PEST can be chosen arbitrarily, with no effect on the estimated target value. Naturally, however, choosing an initial testing level close to the target performance level reduces the time needed to estimate the target SOA.

Based on performance at the initial SOA, PEST selected a new SOA, at which another sequence of trials was conducted. Based on performance at this second SOA, a third SOA was selected and tested, and so on, until termination criteria had been satisfied.

Although this general procedure is conceptually straightforward, there are several methodological questions one may wish answered. In particular, (a) How does PEST decide when the SOA should be changed? (b) How does PEST decide what SOA should be tested next? (c) How does PEST decide when no more trials are needed to estimate the target SOA? and (d) How does PEST estimate the target SOA?

First, target level is changed according to the sequential likelihood ratio test of Wald (1947). This test determines whether the current SOA yields performance better or worse than the desired target performance level and produces a decision of any given power in as few trials as possible. The power of the test depends on an adjustable parameter, *W*, the deviation limit of the sequential test. Small values of *W* yield quick but not very powerful decisions, whereas large values of *W* yield slow, but more powerful decisions.

If the Wald test finds that the current SOA is too large or too small, then SOA is changed in the appropriate direction. The size of the change is either half, the same, or double (up to a maximum step size) the size of the previous change, according to a set of rules given by Taylor and Creelman (1967).

PEST stops when the rules call for an SOA change of some predetermined minimum step size, which must be chosen to be small enough to have adequate precision for the final estimate of target SOA. In fact, the estimate of the target SOA is that testing level called for by the last minimum step, although no trials are run at this level.

Table 9 shows the specific parameter values used in the current implementation of PEST for Experiments 2–4. These parameters were chosen based on the recommendations of Creelman and Kaplan (1973) and Taylor and Creelman (1967) as well as on experience gained during testing of pilot subjects.

Two related complications arose in the current implementation of PEST: First, the sequential likelihood ratio test is based on percentage correct rather than d'. We preferred to use d' in order to eliminate response biases in both discrimination and awareness responses, in all cases except estimation of perceptual thresholds in the four-choice task of Experiment 2. Our solution was to compute d' and then convert it to PC using the conversion table given by Elliott (1964). Thus, Multiple PEST used a value of PC estimated from an actual performance-based d' to decide whether SOA should be adjusted. Second, the d' for trials at the current SOA could not be computed until there had been at least one hit, miss, false alarm, and correct rejection. Under most circumstances, we simply did not start conducting the likelihood ratio test within any set of trials until one of each type of trial had occurred. The exception to this rule was that we used a d' of 4.6 if the subject performed perfectly for 10 or

TABLE 9
Parameters of PEST Used in Experiments 2–4

Parameter	Experiment 2	Experiments 3 and 4
W	0.75	1.1
Starting level	100 ms	100 ms
Starting step size	10 ms	4 ms
Minimum step size	2 ms	1 ms
Maximum step size	10 ms	4 ms

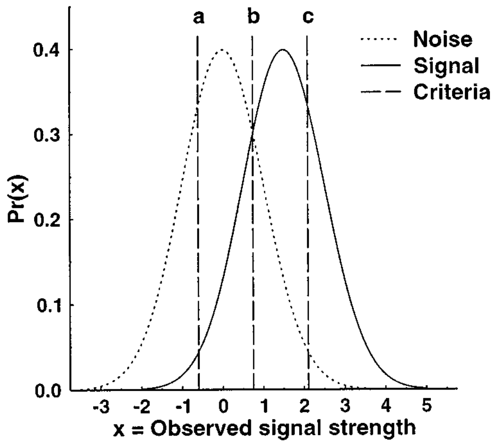

FIG. 4. Signal detection theory representation of the yes/no task followed by a confidence judgment. The distributions of sensory strength, *x*, on signal and noise trials are indicated by lines with small and large dashes, respectively. The three response criteria *a, b,* and *c* are used to select responses: The subject responds ''signal'' if $x > b$ and ''noise'' if $x < b$. The subject responds ''high'' confidence if $x < a$ or $x > c$ and ''low'' confidence if $a < x < c$.

more trials at a given SOA, which occasionally happened at the beginning of an experimental session when SOA was long.

APPENDIX B

Judgment Difficulty Effects

How would an ideal observer respond, given the requirement to discriminate correct responses from errors as in the present experiments? This appendix examines this question within the context of a well-known formalism for the behavior of an ideal observer, that of signal detection theory (Green & Swets, 1966). Predictions of this theory are developed within the context of the familiar yes/no detection task.[7]

Suppose that in each trial, an ideal observer must decide whether a signal or noise has been presented using only a unidimensional measure of stimulus strength, *x*. As shown in Fig. 4, this strength value is commonly assumed to be a random variable from one normal distribution when noise is presented and from another normal distribution, having a different mean but the same variance, when the signal is presented.

In order to decide whether to respond ''signal'' or ''noise,'' the ideal observer is assumed to establish a response criterion, *b*. Then, the subject responds ''noise'' if $x \leq b$ and responds ''signal'' if $x > b$. To meet the additional task requirement of responding ''High'' or ''Low'' confidence, the ideal observer would presumably establish two additional criteria, *a* and *c*, to divide each response category (i.e.,

[7] There are many possible ways to extend signal detection theory to identification and same–different tasks with multidimensional stimuli (Macmillan & Creelman, 1991), as we used in Experiments 2–4. Because we wish to make only qualitative rather than quantitative predictions from this theory, however, we chose to present the problem in terms of the simplest case.

''noise'' versus ''signal'') into two subcategories corresponding to high and low confidence. The subject would then respond that ''High Confidence'' if $x \leq a$ or if $x > c$, and the subject would respond ''Low Confidence'' if $a < x \leq c$.

Using this formalism, it is possible to compute sensitivity measures for both perceptual and awareness judgments, analogous to those computed for subjects in the present experiments. With regard to perceptual judgments, the hit and false alarm probabilities are $\Pr(\text{hit})_p = \Pr(x > b|\text{Signal})$ and $\Pr(\text{false alarm})_p = \Pr(x > b|\text{Noise})$ and the sensitivity measure is $d'_p = \phi^{-1}[\Pr(\text{hit})_p] - \phi^{-1}[\Pr(\text{false alarm})_p]$, where ϕ^{-1} is the inverse of the cumulative normal probability function.

For awareness judgments, the hit and false alarm probabilities are $\Pr(\text{hit})_a = 0.5 \times \Pr(x \leq a|\text{Noise}) + 0.5 \times \Pr(x > c|\text{Signal})$ and $\Pr(\text{false alarm})_a = 0.5 \times \Pr(x \leq a|\text{Signal}) + 0.5 \times \Pr(x > c|\text{Noise})$ (cf. Table 1), assuming that signal and noise trials are presented equally often. These probabilities allow us to compute the sensitivity measure for awareness, $d'_a = \phi^{-1}[\Pr(\text{hit})_a] - \phi^{-1}[\Pr(\text{false alarm})_a]$.

Figure 5 characterizes the relationship between d'_a and d'_p. As is well known, d'_p depends only on the separation between the means of the distributions on signal and noise trials, not on the placement of the decision criterion, b. For a given value of d'_p, however, the value of d'_a does depend on the placement of the criteria a and c. Thus, the five lines in the figure correspond to cases in which these criteria are placed so that the probability of a high confidence response is .1, .3, .5, .7, or .9.

For the present purposes, two crucial facts are illustrated by this figure. First, d'_a is consistently less than d'_p. This indicates that, even for an ideal observer, the discrimination of correct responses from errors (i.e., the judgment used to test for awareness) is more difficult than the discrimination of signal trials from noise trials (i.e., the judgment used to test for perception). In other words, judging the stimuli is inherently easier than judging one's own accuracy. Clearly, then, there is some danger that awareness thresholds are larger than perceptual thresholds just because the former require a more difficult judgment rather than because of subliminal perception.

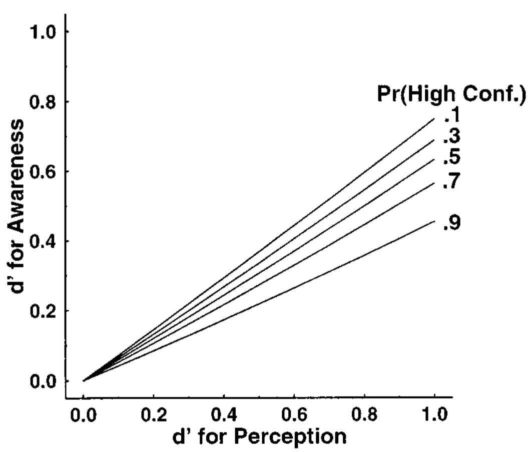

FIG. 5. Predicted relationship between d' for perception and d' for awareness as a function of the criterion used to decide ''high'' vs ''low'' confidence.

The second fact is that d'_a is greater than zero everywhere that d'_p is greater than zero. In principle, this means that the ideal observer's threshold for $d'_a = 0$ is the same as this observer's threshold for $d'_p = 0$. Thus, this fact partly mediates the danger that awareness thresholds are larger just because they are measured with a more difficult judgment; the difference in judgment difficulty can explain a reduction in sensitivity, but not its complete elimination.

There is nonetheless some danger of an artifact analogous to that involving different slopes of the psychometric functions for perception and awareness. Suppose that d'_p increases slowly as a function of SOA; for example, assume that $d'_p = 0$ at an SOA of 5 ms (i.e., threshold) and that $d'_p = 0.01$ at an SOA of 100 ms. As discussed in connection with the differential slope artifact, with such a flat psychometric function the threshold estimation procedure would very likely stop too soon on many occasions, concluding that the perceptual threshold was much larger than its true value of 5 ms. Critically, the fact that $d'_p \leq d'_a$ means that the psychometric function for awareness is even flatter than the one for perception. Thus, stopping too soon would be even more likely to occur for the awareness threshold than for the perceptual threshold, leading to even greater overestimation of the awareness threshold than of the perceptual threshold. Thus, if there is a relatively flat psychometric function for perception, the reduced sensitivity associated with the more difficult awareness judgment could cause an artifactual difference in threshold measures.

Note that the danger of an artifact is quite small if d'_p increases rapidly with SOA. Suppose, for example, that d'_p went from 0 to 2 as SOA went from 5 to 6 ms. Suppose further than d'_a were half of d'_p, which is a conservative estimate judging from Fig. 5. Then $d'_a = 1$ at an SOA of 6 ms, and the threshold estimation procedure would consistently produce threshold estimates very close to the true value, 5 ms. With such a steep psychometric function relating d'_p to SOA, differences in judgment difficulty would clearly produce miniscule differences in estimates of perceptual versus awareness thresholds—far less than 1 ms and much smaller than those observed in the present experiments.

In summary, differences in judgment difficulty could cause even an ideal observer to produce perceptual thresholds less than awareness thresholds. This effect will be quite small, though, as long as d'_p and d'_a increase rapidly with SOA. To evaluate this potential artifact, then, the Experiment 3 included measures of the rate at which these d' values increase with SOA.

REFERENCES

Allport, D. A. (1977). On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.), *Attention and performance* VI (pp. 505–533). Hillsdale, NJ: Erlbaum.

Allport, D. A. (1988). What concept of consciousness? In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 159–182). Oxford, UK: Clarendon Press.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior*, **22,** 88–104.

Berry, D. C. (1990). Talking about cognitive processes. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking* (Vol. 2, pp. 87–98). New York: Wiley.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, **79,** 251–272.

Bisiach, E. (1986). Through the looking-glass and what cognitive psychology found there. *Behavioral and Brain Sciences,* **9,** 24–25.

Björkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics,* **54,** 75–81.

Bricker, P. D., & Chapanis, A. (1953). Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychological Review,* **60,** 181–188.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics,* **36,** 387–395.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology,* **40,** 343–367.

Creelman, C. D., & Kaplan, H. L. (1973). Simultaneous independent threshold estimates: Multiple PEST. *Behavior Research Methods and Instrumentation,* **5,** 89–92.

Dosher, B. A. (1998). The response-window regression method: Some problematic assumptions: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General,* **127,** 311–317.

Duncan, J. (1985). Two techniques for investigating perception without awareness. *Perception & Psychophysics,* **38,** 296–298.

Elliott, P. B. (1964). Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers* (pp. 651–684). New York: Wiley.

Ellis, A. W., & Marshall, J. C. (1978). Semantic errors or statistical flukes? A note on Allport's ''On knowing the meaning of words we are unable to report.'' *Quarterly Journal of Experimental Psychology,* **30,** 569–575.

Emerson, P. L. (1986). Observations on maximum-likelihood and Bayesian methods of forced-choice sequential threshold estimations. *Perception & Psychophysics,* **39,** 151–153.

Erdelyi, M. H. (1974). A new look at the new look: Perceptual defense and vigilance. *Psychological Review,* **81,** 1–25.

Erdley, C. A., & D'Agostino, P. R. (1988). Cognitive and affective components of automatic priming effects. *Journal of Personality & Social Psychology,* **54,** 741–747.

Eriksen, C. W. (1956). Subception: Fact or artifact? *Psychological Review,* **63,** 74–80.

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review,* **67,** 279–300.

Fowler, C. A. (1986). An operational definition of conscious awareness must be responsible to subjective experience. *Behavioral and Brain Sciences,* **9,** 33–35.

Fowler, C. A., Wolford, G. L., Slade, R., & Tassinary, L. (1981). Lexical access with and without awareness. *Journal of Experimental Psychology: General,* **110,** 341–362.

Galvin, S. J. (1988). *The theory of Type II ROC analysis.* Unpublished master's thesis, Department of Psychology, Victoria University of Wellington, Wellington, New Zealand.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. K. (2000). Extending the theory of signal detectability to discrimination between correct and incorrect decisions. Unpublished manuscript.

Gazzaniga, M. S. (1970). *The bisected brain.* New York: Appleton–Century–Crofts.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance,* **8,** 875–894.

Goldiamond, I. (1958). Indicators of perception. 1. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychological Bulletin,* **55,** 373–411.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. *American Psychologist,* **47,** 766–779.

Greenwald, A. G., & Draine, S. C. (1998). Distinguishing unconscious from conscious cognition—

Reasonable assumptions and replicable findings: Reply to Merikle and Reingold (1998) and Dosher (1998). *Journal of Experimental Psychology: General,* **127,** 320–324.

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition,* **17,** 35–47.

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible (''Subliminal'') stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General,* **124,** 22–42.

Harvey, L. O. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments, and Computers,* **18,** 623–632.

Henley, S. H. A. (1984). Unconscious perception re-revisited: A comment on Merikle's (1982) paper. *Bulletin of the Psychonomic Society,* **22,** 121–124.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences,* **9,** 1–23.

Howes, D. (1954). A statistical theory of the phenomenon of subception. *Psychological Review,* **61,** 98–110.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General,* **118,** 126–135.

James, W. (1890). *Principles of psychology.* New York: Holt.

Joordens, S., & Merikle, P. M. (1992). False recognition and perception without awareness. *Memory & Cognition,* **20,** 151–159.

Kellogg, W. N. (1931). The time of judgment in psychometric measures. *American Journal of Psychology,* **43,** 65–86.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature,* **377,** 336–338.

Kunimoto, C. (1994). *Perception without awareness: A new definition and methodology.* Unpublished Ph.D. thesis, University of California.

Kunimoto, C., Miller, J. O., & Pashler, H. E. (1990). Perception without awareness confirmed: A bias-free procedure for determining awareness thresholds. Unpublished manuscript.

Latto, R., & Campion, J. (1986). Approaches to consciousness: Psychophysics or philosophy? *Behavioral and Brain Sciences,* **9,** 36–37.

Lazarus, R., & McCleary, R. (1951). Autonomic discrimination without awareness: A study of subception. *Psychological Review,* **58,** 113–122.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics, *Journal of the Acoustical Society of America,* **49,** 467–477.

Lupker, S. J. (1986). Conscious identification: Where do you draw the line? *Behavioral and Brain Sciences,* **9,** 37–38.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, UK: Cambridge Univ. Press.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology,* **15,** 197–237.

McConnell, J. V., Cutler, R. L., & McNeil, E. B. (1958). Subliminal stimulation: An overview. *American Psychologist,* **13,** 229–242.

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychological Review,* **57,** 235–240.

Merikle, P. M. (1982). Unconscious perception revisited. *Perception & Psychophysics,* **31,** 298–301.

Merikle, P. M. (1984). Toward a definition of awareness. *Bulletin of the Psychonomic Society,* **22,** 449–450.

Merikle, P. M. (1992). Perception without awareness: Critical issues. *American Psychologist,* **47,** 792–795.

Merikle, P. M., & Cheesman, J. (1986). Consciousness as a ''subjective'' state. *Behavioral and Brain Sciences,* **9,** 42–43.

Merikle, P. M., & Reingold, E. M. (1998). On demonstrating unconscious perception: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General,* **127,** 304–310.

Miller, J. O. (2000). Measurement error in subliminal perception experiments: Simulation analyses of two regression methods. *Journal of Experimental Psychology: Human Perception and Performance,* **26,** 1461–1477.

Morton, J. (1986). What do you mean by conscious? *Behavioral and Brain Sciences,* **9,** 43.

Navon, D. (1986). On determining what is unconscious and what is perception. *Behavioral and Brain Sciences,* **9,** 44–45.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General,* **106,** 226–254.

Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology,* **26,** 505–527.

Nolan, K. A., & Caramazza, A. (1982). Unconscious perception of meaning: A failure to replicate. *Bulletin of the Psychonomic Society,* **20,** 23–26.

O'Regan, J. K., & Humbert, R. (1989). Estimating psychometric functions in forced-choice situations: Significant biases found in threshold and slope estimations when small samples are used. *Perception & Psychophysics,* **46,** 434–442.

Paap, K. R. (1986). The pilfering of awareness and guilt by association. *Behavioral and Brain Sciences,* **9,** 45–46.

Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. *Memoirs of the National Academy of Sciences,* **3,** 73–83.

Pollack, I. (1968). Methodological examination of the PEST (parametric estimation by sequential testing) procedure. *Perception & Psychophysics,* **3,** 285–289.

Posner, M. I., & Snyder, C. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Postman, L. (1953). On the problem of perceptual defense. *Psychological Review,* **60,** 298–306.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics,* **44,** 563–575.

Sidis, B. (1898). *The psychology of suggestion.* New York: Appleton.

Snodgrass, M., Shevrin, H., & Kopka, M. (1993). The mediation of intentional judgments by unconscious perceptions: The influences of task strategy, task preference, word meaning, and motivation. *Consciousness & Cognition,* **2,** 169–193.

Stroh, M., Shaw, A. M., & Washburn, M. A. (1908). A study in guessing. *American Psychologist,* **19,** 243–245.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* **18,** 643–662.

Swensson, R. G. (1972). The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception & Psychophysics,* **12,** 16–32.

Taylor, D. A. (1977). Time course of context effects. *Journal of Experimental Psychology: General,* **106,** 404–426.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America,* **41,** 782–787.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology,* **12,** 97–136.

Van Selst, M., & Merikle, P. M. (1993). Perception below the objective threshold? *Consciousness & Cognition,* **2,** 194–203.

Vickers, D., Burt, J., Smith, P., & Brown, M. (1985). Experimental paradigms emphasising state or process limitations. I. Effects on speed-accuracy tradeoffs. *Acta Psychologica,* **59,** 129–161.

Vokey, J. R., & Read, J. D. (1985). Subliminal messages: Between the devil and the media. *American Psychologist,* **40,** 1231–1239.

Wald, A. (1947). *Sequential analysis.* New York: Wiley.

Weiskrantz, L. (1986). *Blindsight: A case study and implications.* Oxford, UK: Oxford Univ. Press.

Weiskrantz, L. (1992). Introduction: Dissociated issues. In A. D. Milner & M. D. Rugg (Eds.), *The neuropsychology of consciousness. Foundations of neuropsychology: A series of textbooks, monographs, and treatises* (pp. 1–10). New York: Academic Press.

Wiener, M., & Schiller, P. H. (1960). Subliminal perception or perception of partial cues. *Journal of Abnormal and Social Psychology,* **61,** 124–137.

Williams, P. C., & Parkin, A. J. (1980). On knowing the meaning of words we are unable to report—A confirmation of the guessing explanation. *Quarterly Journal of Experimental Psychology,* **32,** 101–107.

Wolford, G. L. (1986). A review of the literature with and without awareness. *Behavioral and Brain Sciences,* **9,** 49–50.

Received April 6, 2000