Judging confidence influences decision processing in comparative judgments

WILLIAM M. PETRUSIC
Carleton University, Ottawa, Ontario, Canada

and

JOSEPH V. BARANSKI
Defence Research and Development Canada, Toronto, Ontario, Canada

Current theories of confidence in human judgment assume that confidence and the decision it is based on are inextricably tied to the same process (decisional locus theories) or that confidence processing begins only once the primary decision has been completed (postdecisional locus theories). In the absence of auxiliary assumptions, however, neither class of theory permits the judgment of confidence to affect primary decision processing. In the present study, we examined the effect of rendering confidence judgments on the properties of the decision process in a sensory discrimination task. An examination of the properties of the time taken to determine confidence (i.e., the time taken to render the judgment of confidence) revealed clear evidence of postdecisional confidence processing. Concomitantly, the requirement of confidence judgments was found to substantially increase decisional response times, suggesting that some confidence processing occurs during the primary decision process. We discuss the implications of these findings for contemporary models of confidence in human judgment.

Theories of confidence in human judgment can be clearly distinguished on the basis of their presumed locus of confidence processing. One class of models, which we will refer to as the *decisional* locus models, assumes that confidence is inextricably tied to the decision process. Another class of models, which we will refer to as the *postdecisional* locus models, assumes that confidence processing begins only after the primary decision has been rendered. An example of the decisional locus view is evident in the earliest, most well-known, and most widely applied conceptualization of confidence developed in the context of signal detection theory (SDT). According to this view, decision and confidence emerge simultaneously, through a scaling of decision strength or distance from the decisional criterion or through the setting of multiple category boundaries (criteria) along the decision axis (see Egan, Schulman, & Greenberg, 1959; for alternative models, see Balakrishnan & Ratcliff, 1996, Ferrell & McGoey, 1980, Gu & Wallsten, 2001, and Wickelgren, 1968).[1] Importantly, in the absence of

auxiliary assumptions, this decisional locus view is clear in predicting that the judgment of confidence will have no effect on the properties of the primary decision process, precisely because "there is no basis within signal-detection theory for predicting that the elicitation of confidence judgments should fundamentally alter the detection process" (Balakrishnan, 1998, p. 605).

In the present paper, we first will provide a direct test of this hypothesis in a two-alternative forced-choice sensory discrimination task by requiring one group of participants to express confidence on each trial and another not to. According to SDT, discriminative accuracy should not be influenced in any way, either enhanced or diminished. As well, by examining response times (RTs), we can also determine whether rendering confidence alters the properties of the decision process. However, SDT was not developed with the intention of specifying the time course of confidence and decision processing. Thus, in the present experiment, we also sought to explore this time course. On the other hand, there are classes of theories that are explicit about the time course of confidence processing, and as such, the present study provides a first-order test of these theories. One such class can be referred to as postdecisional theories.

Examples of the postdecisional class of confidence models are rooted in a sequential-sampling decision process (e.g., Juslin & Olsson, 1997; Vickers, 1979). For example, according to Vickers's (1979; Vickers & Packer, 1982) accumulator model of comparative judgment (see also Van Zandt, 2000), stimulus difference information is sampled at discrete time intervals, with positive and neg-

ative information totals (t_p and t_n) accruing in separate accumulators. When one of these accumulators reaches its preset criterion, sufficient evidence has accrued for the selection of a particular response. According to the *balance of evidence* hypothesis, confidence is proportional to the difference in the accumulated totals when one of the accumulators reaches its criterion—that is, *confidence = $c|t_p - t_n|$*, where c is a proportionality constant.

According to Juslin and Olsson's (1997) sensory-sampling model (SESAM), the distribution of sensory differences is sampled sequentially, as in the Vickers accumulator. However, SESAM posits an evidence accumulation process that is subject to leaky integration. Notably, when an information buffer of size w is filled and a decision has not been reached, a sample observation (the first in the buffer) is deleted, and a new sample observation is added. According to SESAM, sensory observations are added to the buffer whenever the mean of the observations lies in an indifference interval—that is, whenever $c_1 \leq \overline{x} \leq c_2$, where c_1 and c_2 represent the boundaries of the indifference interval and $\overline{x}$ the mean of the sensory evidence. The overt decisional response occurs whenever $\overline{x} > c_2$ or $\overline{x} < c_1$ (but see Vickers & Pietsch, 2001, for a detailed critique of this model). Expression of confidence requires determining the proportion of observations in the sensory buffer that have the same sign as the mean, on which the decision was based.

Although these two models differ substantially and posit extremely different views of how confidence is computed, they are clear in postulating a postdecisional locus for confidence:

> The "balance of evidence" hypothesis effectively enables the observer to assess a likelihood ratio *after each decision* (italics added), and to assign to responses an appropriate degree of confidence, constituting an efficient description of the statistical information sampled from the stimulus. (Vickers & Packer, 1982, p. 196)

> When the participant makes the decision that $\mu > 0$, confidence is defined by the proportion of sensations X_i in the sample that are larger than 0; when the decision is that $\mu < 0$, confidence is defined by the proportion of sensations X_i smaller than 0. These assessments are consequent on a deliberate decision and are made *only after a noticeable difference has been detected*. (italics added; Juslin & Olsson, 1997, p. 349)

Importantly, as with the decisional locus models, there is no mechanism within the postdecisional class of confidence models to permit the judgment of confidence to influence primary decision processing.

Recently, Baranski and Petrusic (1998) initiated research aimed at determining the locus of confidence judgments through an examination of the properties of the time taken to determine confidence—that is, the time taken to render the judgment of confidence. Our findings provided preliminary support for the postdecisional models by revealing that some portion of confidence processing does indeed occur after the primary decision is rendered. However, we also found evidence that suggested a decisional

locus for the determination of confidence when participants were oriented to stress accuracy at the expense of speed. Unfortunately, our studies could not ascertain whether and to what extent the primary decision process was affected by the requirement of confidence judgments.

In the present study, we sought to examine this possibility directly. First, we examined the time taken to determine confidence, in order to determine the extent to which confidence processing occurs after the primary decision has been rendered. Following the logic developed in Baranski and Petrusic (1998), if the time to determine confidence were to be invariant across the levels of confidence selected by the participants, we could infer that there was no postdecisional confidence processing and, thus, that confidence was immediately available at the moment of decision. On the other hand, if the time taken to determine confidence were to vary systematically across the levels of confidence selected by the observer, we could infer that confidence was, to some degree, determined after the primary decision. Second, using a between-subjects design, in which one group of participants was instructed to render confidence whereas the other was not, we could gauge the effects, if any, on decisional processing through an examination of decision times and error rates in the two groups.

Thus, evidence that discriminative accuracy is altered when confidence is required would not be permitted by the SDT-based decisional views. However, RT analyses do not provide critical tests of the SDT view of decision and confidence because, as was pointed out earlier, these theories are silent about the time course of confidence processing. On the other hand, on a strict postdecisional view, the fundamental properties of decisional RT, especially its relationship with a priori difficulty of comparisons, cannot be altered when confidence is required; confidence cannot be computed until the decision process is complete and the evidence on which the computation of confidence is based is available. Of course, it is entirely possible on the postdecisional view that some increments in decision time are due to confidence processing. However, these increments cannot be larger than the postdecisional contributions, and they must not in any way alter the fundamental properties of decisional RTs.

METHOD

Participants
Twenty-eight Carleton University undergraduates served for a single session of approximately 1 h for course credit. All were naive concerning the nature and aims of the experiment.

Apparatus
Stimuli were presented on a Zenith ZCM-1492, flat screen CRT video monitor with 640 pixels horizontally and 480 pixels vertically. High-resolution graphics were permitted with VGA Wonder+ video card and MetaWindows graphics under Turbo Pascal (Version 7) software control. Timing, accurate to within ±1 msec, was possible with a Data Translation clock board and extensive software development. Graphics production, stimulus presentation, event sequencing, and the recording of responses and RTs were controlled by an IBM-PC/486DX 33-MHz clone computer.

The participants responded on a movable 13 × 17 cm panel containing three banks of response buttons. The first, at the bottom and center of the panel, contained a single home key, which was used to initiate a trial. The second, directly above, contained two primary response buttons that were used to indicate a left or a right decisional response. The third bank, located directly above and symmetrically around the second bank, contained six buttons arranged in a semicircle. The middle four of these keys were used in the confidence judgment, with the labels "Guess," "Little Certainty," "Moderate Certainty," and "Certain" placed directly above the respective buttons from left to right. A small desk lamp, positioned behind the video monitor, provided sufficient light for responding without interfering with the visual display.

Stimuli and Design

Stimuli were 10 pairs of squares that varied in a priori difficulty of discrimination. The sizes (x, y, r) are used to describe each of the 10 pairs, where x is the number of pixels defining the side of the larger square, y is the number of pixels defining the side of the smaller square, and r is the ratio of the areas of the squares. The 10 pairs of squares, ordered from easiest to most difficult to discriminate, were (5, 4, 1.5625), (12, 11, 1.1901), (53, 51, 1.0799), (166, 160, 1.0764), (58, 56, 1.0727), (186, 182, 1.01.0444), (53, 52, 1.0388), (58, 57, 1.0353), (184, 182, 1.0221), and (161, 160, 1.0125). To discourage the comparison of the areas of the squares on the basis of the length of a side, the two squares in each pair were separated horizontally by 10 pixels and were offset vertically so that the top of the lower square was aligned with the bottom of the upper square.

On half of the trials, the participants selected the "smaller" square and on the other half they selected the "larger" square. In addition, each square in the pair appeared equally often on the left and right and on the top and bottom of the display. Each of these two instructions by four positions was replicated four times (blocks) with each of the 10 pairs, for a total of 320 experimental trials. Thus, upon combining the data over the two instructions and the four different positional arrangements, 32 replications were obtained with each of the 10 pairs. The first block was preceded by 20 practice trials, which were randomly selected from the 80 cells of the design.

The participants were randomly assigned to the confidence condition and to a no-confidence condition, in which confidence judgments were not requested.

Procedure

The participants in each condition were read instructions, which explained the nature of the stimuli, the two instructions, and the four display arrangements. On each trial, an instruction appeared on the screen (to choose the "larger" or the "smaller" square), followed 1 sec later by the stimulus pair. In both the confidence and the no-confidence conditions, the participants' (left or right) decisional response removed the instruction and the stimulus pair. In the confidence judgment condition, the participants' decisional responses were followed by confidence ratings in which they selected one of the four labeled buttons on the top bank of the response panel. They then initiated the next trial by depressing the home key with the index fingers of their preferred hands, and the next trial began 1,500 msec later. In the no-confidence condition, there was no mention of confidence reporting, and there were no labels on the upper bank of response buttons; the participants simply depressed the home key to initiate the next trial. All the participants were instructed that accuracy and speed were both important but that accuracy was more important.

RESULTS

Trials on which primary decision time exceeded 5 sec were trimmed (approximately two standard deviations from the mean). This accounted for 393 of the 8,960 trials obtained in the study (i.e., 4.38%). The data points reported in the figures are based on the mean of all the participants. Significance levels for analyses of variance (ANOVAs) were set at $p < .05$ and were based on the Huyhn–Feldt adjusted degrees of freedom (although the degrees of freedom reported in the text were based on the design). The findings will be presented in two main sections. In the first, decisional RTs and confidence RTs are examined as a function of confidence category and difficulty of decision, with a view toward determining whether confidence processing is postdecisional; in the second, the effects of rendering confidence on primary decision times and discriminative accuracy are examined with a view toward determining whether and to what extent the requirement of confidence judgments affects primary decision processing.

Decision Time and Postdecisional Confidence Time

Decision time analyses. As has been known since Henmon (1911), primary decision times decline monotonically with increasing confidence. The top panel of Figure 1 provides a plot of primary decision time as a function of confidence category, separately for each of three levels of difficulty: *easy* [the three easiest to discriminate pairs with $p(\text{Error}) = .170$], *hard* [the three most difficult to discriminate pairs with $p(\text{Error}) = .345$], and *intermediate* [the four pairs of medium difficulty, with $p(\text{Error}) = .242$]. An ANOVA[2] was conducted with mean decision RTs as the dependent measure, with three levels of difficulty and four confidence categories as within-subjects factors. These analyses revealed significant linear trends [$F(1,8) = 16.16$, $MS_e = 666,152.39$] and a main effect of difficulty level [$F(2,16) = 5.48$, $MS_e = 28,142.53$]; at each confidence category, RTs increased as difficulty increased (cf. Baranski & Petrusic, 1994; Juslin & Olsson, 1997).

Confidence RT analyses. Trials on which confidence RTs exceeded 2.5 sec (within the subset of RT constraints on the primary RTs) were trimmed. This accounted for 33 of 4,217 trials on which confidence was rendered (i.e., 0.78%). In replication of the findings reported by Baranski and Petrusic (1998), the plots in the bottom panel of Figure 1 exhibit four important properties. First, confidence RTs were substantially faster than the primary decision times. Second and most important, confidence RTs varied systematically as a function of confidence category—that is, the time taken to determine confidence was not a constant and, thus, some confidence processing was occurring postdecisionally. Third, in contrast to the strictly linear dependence of primary decision times on confidence category, confidence RTs displayed both significant linear [$F(1,8) = 5.34$, $MS_e = 19,822.96$] and quadratic [$F(1,8) = 5.60, MS_e = 9,626.41$] trends. Indeed, as was shown in Baranski and Petrusic (1998), the nonmonotonicity arose from faster confidence responding with the "Guess" category, referred to as "Fast-Guessing." Importantly, the specific form of the confidence-RT–confidence-category relation was not a consequence of the relative frequencies of category
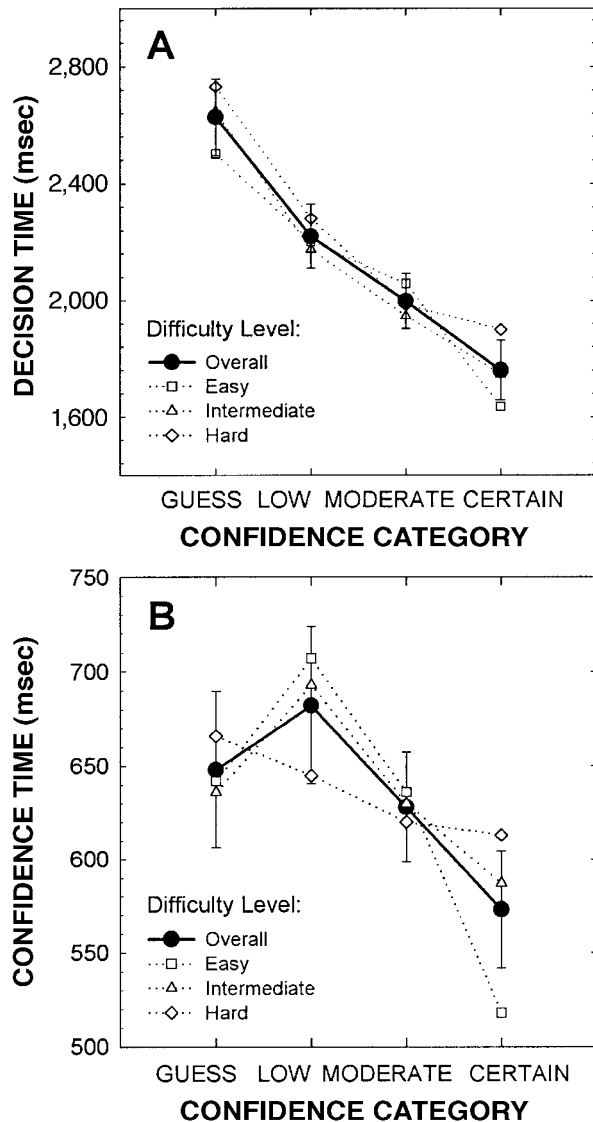
Figure 1. (A) Decision time as a function of confidence category for the easy, intermediate, and hard difficulty levels and overall, upon combining over difficulty levels. (B) Confidence response time as a function of confidence category for the three difficulty levels and overall difficulty levels. Error bars are provided for the overall plots and are based on the standard error of the mean across participants.

usage; for example, confidence RTs were fastest for the "Certain" category, but that category was not the most frequently used. Finally, in contrast to primary decision times, confidence RTs did not reliably depend on decision difficulty (i.e., the interaction between difficulty level and confidence category was not significant).[3]

Effects of Rendering Confidence on Decision Processing

Decisional response time analyses. Figure 2A, the upper panel, presents mean overall decisional RTs for

each pair in each of the two conditions; the lower panel provides mean proportion of errors. For each of the 10 comparison pairs, primary decision RTs were substantially increased when confidence judgments were required, and the difference between the conditions was particularly large for the easier comparisons. Importantly, the increase in decision time when confidence was required cannot be attributed to the participants' simply withholding their confidence judgment until after the decision was rendered. If this were the case, the pattern of confidence RTs in Figure 1B would show no evidence of postdecisional confidence processing. Rather, the findings in Figures 1B and 2A, taken together, suggest that confidence processing occurred during the decision process and postdecisionally.

An ANOVA was conducted with confidence rendered or not as the between-subjects condition and the 10 pairs and four blocks as the within-subjects conditions. The main effect of group was significant [$F(1,26) = 5.03$, $MS_e = 7,189,199.62$]; RTs were 359 msec longer when confidence was rendered than when it was not. The pairs differed reliably as a function of a priori difficulty [$F(9,234) = 30.70, MS_e = 125,336.59$], and the interaction between pair and group was also reliable [$F(9,234) = 2.80$]. As is evident in Figure 2A, the difference between the two groups was diminished as the pairs became more difficult to discriminate—that is, the easiest to discriminate pair exhibited the largest RT difference (452 msec), and the most difficult pair exhibited that smallest RT difference (175 msec). As well, RTs decreased with practice; the effect of block was reliable [$F(3,24) = 21.42$, $MS_e = 374,810.96$], as was the interaction between block and group [$F(3,78) = 4.97$]. The latter reflects the fact that the increase in RT due to rendering confidence, although evident in every block, was decreased as RTs become faster with each block. However, the largest disruptions of decisional processing due to the expression of confidence with the easiest comparisons and the smallest with the most difficult comparisons occurred uniformly across the four blocks.

Error analyses. The plots in the lower panel of Figure 2B indicate that the requirement of confidence judgments had no clear effect on discriminative accuracy. Overall, errors occurred on 18.25% of the trials when confidence was not required and on 21.29% of the trials when it was. An overall ANOVA with the percentage of error measure for each respondent as the dependent variable revealed main effects of stimulus pair [$F(8,18) = 112.23, MS_e = 0.0188$], but not of group [$F(1,28) = 2.28$, $MS_e = 0.1131$]. The issue of whether discriminative accuracy is affected by the requirement of confidence judgments is critically important and, thus, should be reexamined in future research, given that the difference between groups was not significant by conventional ANOVA but that it was the case that error rates were higher in 8 out of the 10 stimulus pairs when confidence was required.

Correct versus error RTs. An important assumption of the preceding analyses is that decision processing was
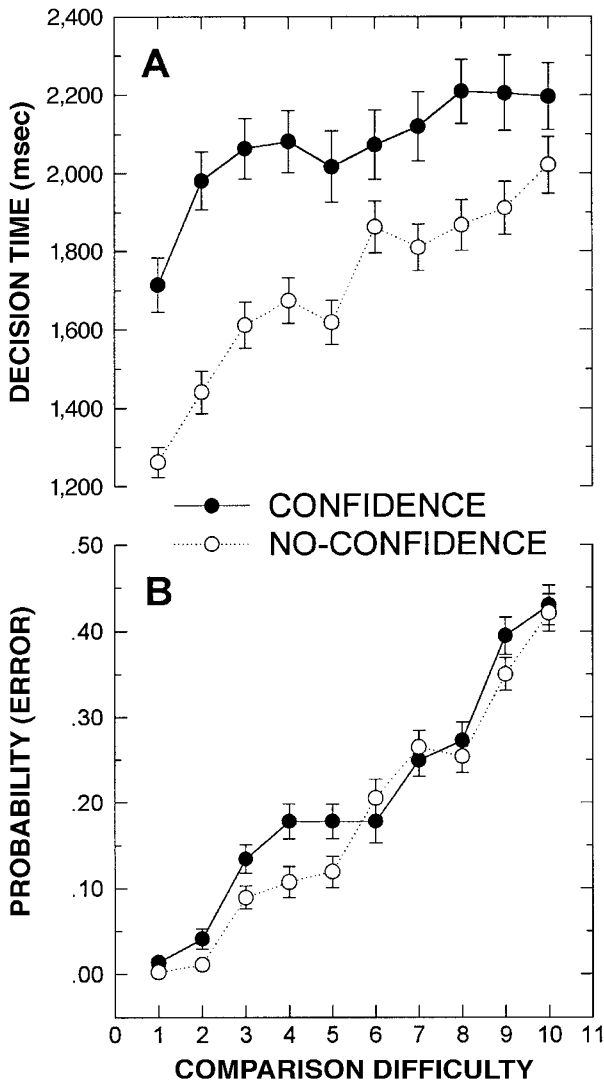
Figure 2. Decision time (A) and probability of error (B) as a function of the ordinal difficulty of the comparison for the confidence (filled points) and no-confidence (unfilled points) conditions. Error bars denote the standard error of the mean across participants.

fidence) as the between-subjects factors and correct responses and errors as a within-subjects factor showed that, indeed, mean error decisional RTs were reliably longer than correct RTs [$F(1,26) = 45.55$, $MS_e = 17,073.04$]. Moreover, the interaction between group and whether the response was correct or in error was significant [$F(1,26) = 6.17$]; error times were 322 msec longer than correct times in the no-confidence condition and 149 msec longer when confidence was required. Alternatively, this interaction shows that the disruptive effect of rendering confidence was significantly greater for correct responses (385 msec) than for error responses (212 msec). Thus, we can conclude that the instructions emphasizing accuracy at the expense of speed were effective for both groups and that rendering confidence alters the precise relationship between correct and error RTs.

GENERAL DISCUSSION

The finding that the properties of decision time are altered when confidence judgments are required confirms Baranski and Petrusic's (1998) conjecture that some portion of confidence processing occurs during the primary decision. On the other hand, it appears that decision accuracy (i.e., discriminative sensitivity) may not be affected by the requirement of confidence ratings (see also Baranski & Petrusic, 2001). Although these results in no way challenge the fundamental decision-processing assumptions of the decisional (i.e., SDT) and postdecisional (e.g., Juslin & Olsson, 1997; Vickers, 1979) locus theories, they do highlight the need for a more complete specification of the time course of confidence processing and its relation to decision processing. Specifically, SDT accounts of confidence would have to provide a clearer distinction between decision (relative to a criterion) and confidence (based on distance from the criterion), as well as a specification of how the judgment of confidence affects decision-processing time. For example, SDT can account for basic RT properties under *strength theory* assumptions (see, e.g., Norman & Wickelgren, 1969). On this view, RTs decrease as the distance between the sampled observation and the criterion increases. This is because more *fundamental operations* are required to resolve on which side of the criterion the sensory observation lies the closer the observation is to the criterion and that each operation takes some small amount of time. However, once the observation is located relative to the criterion, the distance from the criterion can be scaled, and confidence can be determined fairly automatically, presumably without any effect on decision-processing time.

Although the postdecisional models are supported by evidence of postdecisional confidence processing (Figure 1B), they too cannot predict the increase in decision time when confidence is required, without making additional assumptions. For example, one intuitively compelling view of the requirement to render a confidence judgment is that two tasks must be performed concurrently—that is, the primary decision task is carried out under di-

carried out under an emphasis on accurate, as opposed to speeded, responding. Accordingly, it is important to show in the present experiment that the instructions to emphasize accuracy at the expense of speed were indeed effective. It has long been known that the relationship between RTs for correct and error responses is indicative of whether participants set primary decisional criteria emphasizing speed or accuracy. Notably, when accuracy is stressed, decision RTs for errors are longer than those for correct responses, whereas when speed is stressed, RTs for errors are as fast as or even faster than those for correct responses (see, e.g., Luce, 1986; Petrusic, 1992; Pike, 1968; Ratcliff, Van Zandt, & McKoon, 1999; Vickers, 1979). An ANOVA with group (confidence vs. no con-

vided attention, concurrently with the determination of confidence. This view, taken in the context of the currently popular theories of evidence accrual in *discrete time* (see Link, 1992, Luce, 1986, Petrusic, 1992, and Vickers, 1979, for alternative decision processes), predicts that the largest and most disruptive effects of confidence processing on the primary decision process will be with the most difficult decisions. This follows because, according to these theories, the number of evidence accrual events increases with the difficulty of the decision. If it is assumed that, under conditions of divided attention, the accrual process is slowed, the largest effects of rendering confidence on primary decisional processing must be with the most difficult decisions. However, in the present study, the largest effects of rendering confidence were observed with the easiest judgments (Figure 2A).

Alternatively, the more general diffusion models of the accumulation of evidence in *continuous* time (Ratcliff, 1978) may well accommodate a multitasking view of the effects of rendering confidence, although the basis for confidence according to these views is in some doubt (see Baranski & Petrusic, 1994, p. 418; Vickers, 1979, 2001). Specifically, the diffusion rate for the easier judgments could well be slowed more than that for the more difficult judgments because multitasking is more likely to take place when the primary task demands are light, precisely as was found.[4]

The present findings also suggest that the collection of confidence data may greatly complicate the mathematical modeling of primary decisional RTs and using RTs for scaling and measurement analyses. Specifically, current theories of the decision process assume that $RT = DT + C$, where DT denotes decisional components and C denotes residual nondecisional components (e.g., sensory input and motor execution). However, the fact that decisional RTs are increased by the requirement of confidence and that this increase is dependent on the difficulty of the comparison cannot be modeled simply by changes in the additive constant. Similarly, the value of the constant C would have to depend on whether the response was correct or not when confidence is rendered, because rendering confidence increased RTs more for correct responses than for errors. Thus, these effects must be modeled within the context of theories that make explicit provision for the time taken during decisional processing for the determination of confidence within the multitasking framework outlined earlier.

The recent neural-based leaky-integrator model of RT developed by Usher and McClelland (2001) provides precisely such a direction. According to this winner-take-all idea, competing streams of activation grow with time and are also subject to mutual inhibition. A decision is reached when one of the streams of activation reaches a stable threshold level and all other streams are inhibited. Although Usher and McClelland do not provide a basis for confidence in their model, a more general form of Vickers's (1979) balance of evidence can be considered

for the computation of confidence. However, this computation will require both ongoing monitoring of the dynamic course of activation growth and inhibition and a postdecisional interrogation of the evidence for and against each of the competing responses accrued over the course of decision processing. On the assumption that active monitoring of the competing streams of activation and inhibition slows the growth of activation, rendering of confidence will slow decisional processing, without having much effect, if any, on discriminative sensitivity. Such a view may thus permit both the very large decisional and the smaller postdecisional components of confidence processing observed in the present experiment.

REFERENCES

Balakrishnan, J. D. (1998). Measures and interpretations of vigilance performance: Evidence against the decision criterion. *Human Factors*, **40**, 601-623.

Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 615-633.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, **55**, 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 929-945.

Baranski, J. V., & Petrusic, W. M. (2001). Testing architectures of the decision-confidence relation. *Canadian Journal of Experimental Psychology*, **55**, 195-206.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America*, **31**, 768-773.

Ferrell, W. R., & McGoey, P. J. (1980). A model of calibration for subjective probabilities. *Organizational Behavior & Human Performance*, **26**, 32-53.

Gu, H., & Wallsten, T. S. (2001). On setting response criteria for calibrated subjective probability estimates. *Journal of Mathematical Psychology*, **45**, 551-563.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, **18**, 186-201.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, **104**, 344-366.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: Erlbaum.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

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

Norman, D. A., & Wickelgren, W. (1969). Strength theory of decision rules and latency in short-term memory. *Journal of Mathematical Psychology*, **6**, 192-208.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 962-986.

Pike, A. R. (1968). Latency and relative frequency of response in psychophysical discrimination. *British Journal of Statistical & Mathematical Psychology*, **21**, 161-182.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Comparing connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.

Usher, M., & McClelland, J. L. (2001). The time course of percep-

tual choice: The leaky, competing accumulator model. *Psychological Review*, **108**, 550-592.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **26**, 582-600.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Vickers, D. (2001). Where does the balance of evidence lie with respect to confidence? In E. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001: Proceedings of the seventeenth annual meeting of the International Society of Psychophysics* (pp. 148-153). Lengerich: Pabst Science.

Vickers, D., & Packer, J. S. (1982). Effects of alternating set for speed versus accuracy on response times, accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica*, **50**, 179-197.

Vickers, D., & Pietsch, A. (2001). Decision-making and memory: A critique of Juslin and Olsson's (1997) sampling model of sensory discrimination. *Psychological Review*, **108**, 789-804.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, **5**, 102-122.

NOTES

1. In forced-choice sensory discrimination, as in the experiment to be reported, on any trial with two stimuli—say S_1 and S_2—the task for the observer is to respond R_1 when stimulus S_1 is presented on the left (first) and S_2 on the right (second), with this order being denoted as (S_1, S_2). When the stimulus pair is presented in the order (S_2, S_1), the observer is to respond R_2 (see Macmillan & Creelman, 1991, pp. 127–130). When ratings of confidence are required, the observer selects a response ranging from high confidence that the stimulus pair (S_1, S_2) was presented to high confidence that it was the other pair, (S_2, S_1). Confidence-rating–based plots of the probability of the response R_1 given that the pair (S_1, S_2) was presented against the probability of the response R_1 given that the pair (S_2, S_1) was presented permit construction of a receiver operating characteristic (ROC) curve from which an estimate of discriminative sensitivity can be obtained. As Macmillan and Creelman (1991, pp. 50–60) indicated, confidence-based ROCs can be obtained in any one of three ways. For example, a numerical scale ranging from 0 to 100 (full range) could be employed with 0 denoting a subjective probability of zero that the pair (S_1, S_2) had been presented and 100 a subjective probability of 1.00 that the pair (S_1, S_2) had been presented. Of course other number categories, not necessarily mapping so clearly onto subjective probabilities, might also be used (e.g., ranging from 1 to 5). Second, the confidence categories might be defined verbally, ranging from *completely certain it was (S_1, S_2)* to *completely certain it was (S_2, S_1)*, with intermediate categories reflecting varying degrees of confidence in which stimulus pair was presented. Finally, precisely as in the present experiment, two responses are required on each trial. The first requires the binary response R_1 or R_2, and the second requires an expression of confidence. For example, a half-range numerical scale, ranging from 50, indicating a .50 subjective probability (a guess) to 100 (indicating complete certainty) could be used, or verbally defined categories, such as *guess*, and *moderate certainty* to *complete certainty*, might be also be used. *Formally*, according to SDT, each of these three variants for expressing confidence is identical. Each results in the activation of an ordered series of cut-points, c_1 to c_n, on the decision axis defined by the sensory difference, $d = X - Y$, where X and Y represent the sensory effects of the stimuli presented on the left (first) and right (second), respectively, with the confidence response C_i occurring whenever $c_i < d < c_{i+1}$.

2. It is important to note that not all the participants used all the confidence categories at each difficulty level. Consequently, statistical analyses involving the confidence category factor were complicated by the inclusion of this factor, and required combining the data over pairs to define aggregate difficulty levels. Thus, these analyses and the subsequent analyses on decision time and confidence RTs were based on but 9 of the 14 participants who provided confidence judgments in each of the four confidence categories at each of the three aggregate difficulty levels.

3. In contrast, exclusion of the confidence category factor permits examination of the full range of difficulty levels at each block for each participant. Indeed, an ANOVA with mean overall confidence RTs, with block and pair as within-subjects factors, revealed a reliable effect of block [$F(3,39) = 9.41$, $MS_e = 76,192.35$], confirming an effect of practice. The effect of pair was also reliable [$F(9,117) = 2.81$, $MS_e = 17,238.04$]. The fastest expression of confidence occurred with the very easiest pair (553 msec), and the slowest time to make a confidence judgment occurred with the most difficult comparison (659 msec). Tukey paired-comparisons of mean confidence RTs showed only the easiest pair to differ from the two most difficult pairs (HSD = 95.71 msec). This dependence of confidence RT on decision difficulty may reflect an influence of decisional processing on confidence RTs.

4. We are grateful to John Wixted for suggesting this aspect of the multitasking view.