ON THE LOCUS AND TIME COURSE OF CONFIDENCE PROCESSING

Steven R. Carroll and William M. Petrusic
Department of Psychology, Carleton University, Ottawa, Canada
srcarroll@canada.com; bill_petrusic@carleton.ca

Abstract

Participants were assigned to one of two types of psychophysical tasks: signal detection or line-length discrimination. In one of three blocks of trials participants rated their confidence in the accuracy of each decision by selecting one of six confidence categories. In a second block participants never rendered confidence. In a third block participants always expected to have to render confidence, but were instructed not to do so if a tone sounded. The tone sounded on 50% of the trials at times varying from 100 to 500 ms following stimulus presentation. Primary decision reaction times increased whenever confidence was rendered. Importantly, in the stop-confidence processing signalled task, primary decision response times increased linearly as the time between the stimulus presentation and the stop-confidence signal increased. These findings are clear in implicating both a decisional and post-decisional locus for the basis of confidence. Taken together, the findings support the idea that confidence processing evolves at a constant rate over the course of primary decision processing.

Baranski and Petrusic (1998) initiated the study of the locus and the time course of confidence processing in comparative judgements through investigation of the properties of the time to determine confidence. The time to determine confidence is defined by the time between the expression of the primary decision and the selection of a confidence category corresponding to the subjective probability that the decision rendered was correct. Baranski and Petrusic concluded that when the primary decision is made under speed stress confidence is computed post-decisionally. On the other hand, under accuracy stress, there was a strong suggestion that confidence is initiated and perhaps even completed during the primary decision process.

Indeed, Petrusic and Baranski (2000, 2003) provided direct evidence that confidence is processed during the primary decision process under a stress for accuracy at the expense of speed. Using a between participants design in a line length discrimination task, one group of participants rendered confidence following the primary decision and the other simply made the comparative judgement. The requirement of confidence judgements substantially increased primary decisional response times (RTs). Moreover, confidence times varied systematically with confidence category, thereby implicating post-decisional confidence processing.

Using a within participants design, Baranski and Petrusic (2001) replicate and extend the Petrusic and Baranski findings, showing increased primary decisional RTs when confidence was rendered, as well as clear evidence of post-decisional confidence processing, in one experiment requiring binary discrimination of visual extents and of the populations of Canadian cities in another.

These findings are problematic for two classes of theories of confidence in human judgement distinguishable on the basis of the presumed locus of confidence processing. The class of *decisional* locus models assumes that confidence is inextricably tied to the decision process. Perhaps the most well-known and widely applied view of confidence was developed in the context of signal detection theory (SDT). On this view, confidence and the reported decision arise simultaneously such as through the setting of multiple category boundaries (criteria) in obtaining confidence based ROCs (e.g., Egan, Schulman, & Greenberg, 1959), for example. This decisional locus view is clear in predicting that rendering confidence will have no effect on the properties of the primary decision process.

According to the class of *post-decisional* locus models, confidence processing begins only after the primary decision process has been completed. Vickers' (1979) *balance of evidence* theory of confidence, developed in the context of his accumulator model of comparative judgement, is indeed the most successful and most fully developed (see, e.g., Van Zandt & Maldonado-Molina, 2004). On Vickers' view, confidence arises as a scaling of the difference in the amounts of evidence accumulated favouring each of the two alternative decisions. Thus, on a strict interpretation of the balance of evidence hypothesis, there is no basis on which confidence processing could result in substantial increases in primary RTs.

The Baranski and Petrusic (2003) and Petrusic and Baranski (2000, 2003) work shows that rendering confidence increases primary decisional RTs but it fails to specify the time course of the computation of confidence. For example, the balance of evidence may be computed during the final stages of decisional processing. On the other hand, the balance of evidence may evolve, continuously, over the course of decision processing. The present experiments were designed to further examine both the loci and the time course of confidence processing. To accomplish this we employed three conditions (within participants), with both a sensory detection and a line-length discrimination task. In one condition, participants always rendered confidence and in a second, confidence was not required. The third condition used a variant of Logan's (1983) "stop-signal" paradigm. In particular, in the Stop-Confidence condition, participants always expected to have to render confidence, but were instructed not to do so if a tone sounded. The tone sounded on 50% of the trials at times varying from 100 to 500 ms following stimulus presentation.

If confidence evolves over the course of decision processing then the increases in primary decisional RTs will increase monotonically with increases in the stimulus-tone onset asynchrony. On the other hand, if confidence is computed only as the decision process approaches completion and prior to motor execution, then RTs to only the longest probe will be increased, provided the probe is sufficiently close to the output stage.

Method

Participants. Forty-eight Carleton University undergraduate students participated in return for course credit.

Apparatus. The study was conducted using a desktop computer with a standard colour monitor. The computer was equipped with a Pentium-class processor, a Soundblaster soundcard, and a Windows 98 operating system. Stimulus presentation and response data collection was controlled via Superlab Pro v. 2.0. Participant responses were made via a control panel with two primary response buttons (labelled 'yes/no' or 'left/right' depending on the task) and seven confidence response buttons (labelled 'X/50/60/70/80/90/100').

Stimuli. Signal detection stimuli were 16, 100 x 100 pixel squares arranged to form a larger 4 x 4 square. Each of the 16 squares contained 2500 2 pixel x 2 pixel 'dots'. Each dot was either inactive (coloured white) or active (coloured black). The density level of the dots in the four centre squares varied from trial to trial, but on any given trial all of the centre squares were homogenously dense, with either 50% active dots (the 'noise' condition), 52%, 54%, 56%, or 58% active dots. Only 50% of the dots in the background squares were active on any given trial.

Line-length stimuli consisted of single, horizontal lines bisected by a vertical line 10 pixels in length. The vertical line was positioned so one of the two horizontal line-segments was always 100 pixels long, while the remaining segment was always longer being either 102, 104, 106, or 108 pixels in length. The lines were offset horizontally by 24 pixels to either the left or right of the centre of the screen, and were positioned 178 pixels above an instruction (‘shorter’ or ‘longer’) which was centred on the screen and written in a 12 point Arial bold font.

Procedure. Half of the participants were assigned to the signal detection task, while the other half were assigned to the line-length discrimination task. Participants in the signal detection task were asked, on each trial, to decide whether there was a greater density of active dots in the centre of the display than there were around the edges of the display. Participants in the line-length discrimination group were asked, on each trial, to decide which of the two line-lengths was either shorter or longer, depending on the instruction. Half of the trials in each block were noise trials and the other half signal trials, with each of the four signal strengths occurring equally often for a total 336 trials in each of three blocks for the signal detection group. Each of the 4 line length pairs appeared with each of the two instructions in each of the two left-right orders and this factorial combination was replicated 24 times for a total of 384 trials in each of four blocks for the line-length discrimination group.

Each participant received each of the three conditions. Participants were randomly assigned to one of the six different orders in which the treatments were administered. In one condition, participants were asked, following each primary decision, to rate their confidence in the accuracy of the decision they had just made. Participants were told that a rating of ‘50’ was indicative of a guess, a rating of ‘100’ indicated certainty, and the ratings 60-90 were to be used accordingly. Participants were further instructed to select a confidence rating of ‘X’ if they were certain they had made a mistake.

In a second condition, participants never had to render confidence. In the third condition, participants were told to expect to have to rate their confidence, but were told they would not have to do so if they heard a short tone. The tone was presented on one-half of the trials, and its onset following stimulus presentation varied, in 100 ms increments, from 100 ms to 500 ms. As well the tone was presented simultaneously with the onset of the stimulus pair.

Results and Discussion

At the outset, RTs three standard deviations above the mean were censored. This accounted for 2.29 % of the 24192 trials in the detection task and 1.98 % of the 27648 trials in the line length task. Throughout, the primary dependent variables for each cell in the design are the mean of each participant’s overall primary decisional RTs, mean confidence time, mean proportion correct responses, and mean confidence rating. The findings are presented in five sections. The first provides mean primary decision times in each of the three conditions for each task. The second examines mean decision times in the Stop-Confidence condition, the third provides analyses of confidence times, the fourth a summary of discriminative accuracy, and the final section provides analyses of the confidence judgements. In each analysis of variance (ANOVA), Huynh-Feldt, epsilon adjustment of degrees of freedom was used. However, the degrees of freedom associated with each value of F are defined by the design and the Mean Square Errors provided in the text are those given by the conventional degrees of freedom. Level of significance was set at 0.05 throughout.

Primary Decisional RT Analyses

Table 1. Mean decisional RTs (Ms) for the each condition for the detection and the discrimination tasks.

	Condition			
		Stop-Confidence		
Task	No Confidence	Tone	No-Tone	Confidence
Signal Detection	1541.23	1548.21	1810.31	2316.72
Line-length	2118.18	1653.36	1845.83	2521.92

The mean RTs for each of the three conditions with two sub-conditions (Tone and No-Tone) for the Stop-Confidence condition for each task are presented in Table 1. As is evident in Table 1, the requirement to render confidence resulted in the longest RTs with both tasks, thus, replicating the findings of Petrusic and Baranski (2000, 2003) and Baranski and Petrusic (2001). The main effect of condition (No confidence, Tone, No-Tone, Confidence) on primary decisional RTs was reliable for both the signal detection task, F(3, 54)=9.09, partial $\eta^2 = .335$, and the line-length discrimination task, F(3, 54)=15.57, partial $\eta^2 = .464$.

Decisional difficulty. RTs varied systematically with the density of the dots in the display and the length of the comparison line. As the density of the signal increased and the length of the comparison line increased RTs decreased. The main effect of dot density on RT was reliable in the detection task, F(4, 72)=25.04, partial $\eta^2 = .582$, as was the effect of comparison line length, F(3, 54)=27.03, partial $\eta^2 = .600$.

The Effect of Tone Onset Delay on Primary Decisional RTs

The plots in Figure 1 provide mean primary decisional RTs for each of the stop-confidence signal onset delays with the stop-confidence condition for each task. As is strikingly evident for each task, decisional RTs increase *linearly* with the delay in the stop signal; i.e., with the amount of confidence processing. An ANOVA showed the linear component of the overall sum of squares for the effect of tone onset delay to be marginally significant for the signal detection task, F(1, 23)=3.21, p>0.08, partial $\eta^2=.123$, and reliable the line-length discrimination task, F(1, 23)=7.74, partial $\eta^2 = .252$.

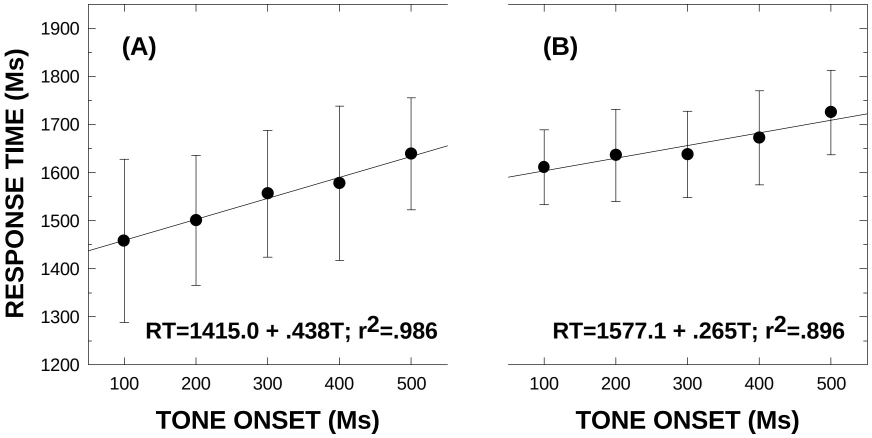

Figure 1. Primary decisional RTs with 95% confidence intervals during the Stop-Confidence condition for the signal detection (Panel A) and the line-length discrimination tasks (Panel B).

Time to Render Confidence

Confidence times varied non-monotonically, albeit systematically, with confidence category in the detection task, implicating a post-decisional component. On the other hand, for the line-length task, confidence RTs do not vary in any systematic manner with confidence category, indicating minimal, if any, post-decisional confidence processing.

It took considerably more time to render confidence in the stop-confidence condition than in the confidence condition, with the main effects of condition evident for the both the detection, F(1,18)=70.808, partial η^2 = .797, and for the line-length discrimination tasks, F(1,18)=48.233, partial η^2 = .728. The random nature of the stop-confidence condition, in which 50% of the trials required confidence rendering and 50% did not, may have resulted in a mixing-cost where a great deal more confidence processing necessarily occurred post-decisionally.

Replicating the findings of Petrusic and Baranski (2003), a main effect of decisional difficulty on time to render confidence was significant for the detection task, F(1,18)=6.475, partial η^2 = .265; confidence times decreased as the differences between the compared densities increased. In contrast, no effect of discriminative difficulty was found for the line-length discrimination task, F(1,18)=1.42, p>.25, partial η^2 = .073, consistent with the view that little, if any, confidence processing occurred post-decisionally.

Detective Sensitivity and Discriminative Accuracy

As is evident from the plots in Figure 2, proportion correct varied systematically with the density of the dots in the display and the length of the comparison line. As the density of the signal increased detective sensitivity increased, F(1,18)=28.004, partial η^2 = .609. Similarly, accuracy improved as the difference between the standard and the comparison increased, F(3,18)=26.881, partial η^2 = .599. As is also clear from Figure 2, both detective sensitivity and discriminative accuracy were uninfluenced by condition.

Confidence Judgements

The plots in Figure 2 are clear in showing that participants were generally under-confident although considerable overconfidence is evident at the 52 % signal. As well, it is clear confidence ratings showed little variation with the difficulty of the judgement in contrast to

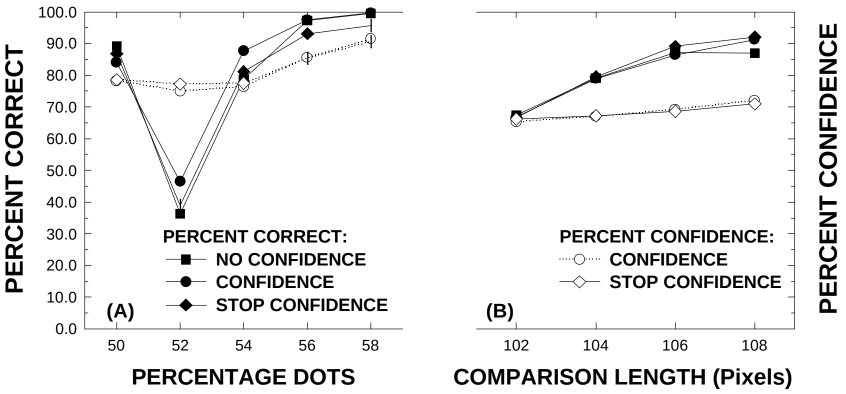

Figure 2. Percent correct and percent confidence as a function of percentage dots (Panel A) and the length of the comparison stimulus (Panel B) in each condition.

the accuracy measures. Nevertheless, an ANOVA showed confidence varied reliably as a function of dot density, F(4, 92)= 32.07, partial $\eta^2 = .582$ and as a function of line-length, F(3, 69)=48.08, partial $\eta^2 = .676$. As the plots in Figure 2 also show, mean confidence did not differ in the two confidence conditions in either task.

Summary and Conclusions

The present findings provide a clear replication and important extension of the earlier work reported in Petrusic and Baranski (2000, 2003) and in Baranski and Petrusic (2001). First, primary decisional RTs are substantially increased when confidence is rendered, clearly implicating a decisional locus for the computation of confidence. Second, it is evident that confidence processing does not have ballistic properties; once initiated it can indeed be stopped. Third, as is also evident from the Stop-Confidence condition, the judgement of confidence is initiated at the outset of decisional processing and the monotonic increase in primary decisional RTs with delays in tone onset indicates that confidence processing continues steadily, evolving throughout decisional processing. In fact, the linearity of the increases in primary decisional RTs with stop-tone delay indicates confidence processing is occurring at a *constant rate* during the course of decision processing, at least for the range of probes used.

References

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 and 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.

Logan, G. D. (1983). On the ability to inhibit simple thoughts and actions: 1. Stop-signal studies of decision and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 585-606.

Petrusic, W.M. & Baranski, J.V. (2000). Effects of expressing confidence on decision processing: Implications for theories of RT and confidence. In C. Bonnet (Ed.). *Fechner Day 2000. Proceedings of the Sixteenth Annual Meeting of the International Society for Psychophysics*. (Pp. 103-108). Strasbourg, France: The International Society for Psychophysics.

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgements. *Psychonomic Bulletin & Review*, **10**, 177-183.

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

Acknowledgement

This work was supported by an Ontario Graduate Studies Scholarship to Carroll and a Natural Sciences and Engineering Research Council of Canada Discovery grant to Petrusic.