NUMERICAL DISTANCE AND ANCHORING EFFECTS IN JUDGING CONFIDENCE

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

Abstract

On half of the trials in a sensory detection task, participants were required to decide whether they were more or less confident of the correctness of their decision than the confidence denoted by a number N ranging from 55 to 95. Subsequently, participants expressed confidence on a scale ranging from 50 to 100. On the other half of the trials, participants simply rendered confidence. Times to compare the internal confidence estimate to the numerical value N revealed numerical distance effects; as the distance between N and their subsequently rendered levels of confidence increased, response times decreased. Analyses of mean confidence ratings revealed anchoring effects, with mean absolute confidence ratings assimilating toward the value of N.

Research over the past decade has demonstrated how confidence processing in human judgement originates during the primary decision-making process and continues post-decisionally (Baranski & Petrusic, 1998, 2001; Petrusic & Baranski, 2000, 2003; Van Zandt & Maldonado-Molina, 2004). Although this literature establishes the locus of confidence processing, it has been relatively silent regarding the manner in which confidence information is represented.

Many models of human decision-making treat confidence as if its assessment were the result of an arithmetic operation. Models that predict a decisional locus of confidence, particularly those developed in the context of Signal Detection Theory (SDT), treat confidence as the difference between two points on an axis measuring magnitude of sensory observation (for example Egan, Schulman, & Greenberg, 1959; Balakrishnan & Ratcliff, 1996; for a review of SDT see Macmillan and Creelman, 2005). Models that predict a strictly post-decisional locus of confidence treat confidence as the difference between accumulated levels of evidence (Vickers, 1979; Petrusic & Baranski, 1989; Van Zandt, 2000; Van Zandt & Maldonado-Molina, 2004). Though each class of model suggests a different origin for confidence, each treat confidence as if its realization produces a discernable quanta of information which can be compared to a hypothetical maximum quantity of confidence: if the magnitude of the difference between the points/totals being compared is relatively large then a high confidence rating is expressed. If this assumption that confidence data is manipulated arithmetically holds true, one would expect confidence processing to be subject to the types of interference that affect other types of cognitive processes involving mathematics. This hypothesis is tested in the present study.

Moyer and Landauer (1967) established that the time it takes an observer to distinguish between two numerical magnitudes diminishes as the difference between the two values grows. This effect has come to be known as the numerical distance effect. If a participant's confidence rating is indeed being represented by a magnitude that can be compared to a number, N, one would expect to see numerical distance effects in a participant's time to decide confidence relative to N as N diverges from the participant's subsequently rendered absolute confidence level.

In one of the Tversky and Kahneman (1974) anchoring paradigms the anchor was made explicit through an initial comparative judgement. Subsequently, participants provided a numerical estimate of the quantity in question. In each case the estimate assimilated toward the value of the anchor provided in the comparison (see also Slovic & Lichtenstein, 1971). This effect has come to be known as the anchoring effect. The process of mapping something as subjective as magnitude of certainty onto an arbitrarily chosen set of confidence categories might constitute a difficult arithmetic calculation of the type described by Tversky and Kahneman. If so, the introduction of a relative confidence rendering task between the primary decision and the absolute confidence rendering phases of each trial might serve to anchor confidence judgments to the value of the number N. As well, since a comparative judgement of confidence with a number N parallels the classic Kahneman and Tversky explicit provision of an anchor, final absolute judgments of confidence may well assimilate toward the provided anchor. If so, such a finding would extend the anchoring effect nicely to the subjective, meta-cognitive domain; i.e. the expression of the certainty of the correctness of one's decisions.

Method

Participants. Twenty Carleton University undergraduate students participated in return for course credit. One participant failed to complete the task and was replaced.

Apparatus. The study was conducted using a desktop computer with a standard colour monitor. The computer was equipped with a Pentium-class processor 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'), two secondary response buttons (labelled '< / >'), and seven confidence response buttons (labelled 'X/50/60/70/80/90/100').

Stimuli. 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 ('noise' trials), 52%, 54%, 56%, or 58% active dots. Only 50% of the dots in the background squares were active on any given trial.

Procedure. Participants 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. Half of the trials in each block were noise trials and the other half signal trials. Each of the four signal strengths occurred equally often over four blocks of 200 trials. The first two blocks of trials represented one condition, while the second two blocks of trials represented a second condition. Condition order was varied between participants.

In one condition participants were asked, following each primary decision, to rate their confidence in the accuracy of the decision they had just made. The time taken to make this judgment will be referred to as confidence time. 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 were asked, following each decision, to decide whether their eventual confidence rating would be more or less than a number presented on the computer monitor following the disappearance of the primary stimulus. Participants were told to press the '<' button if they were less confident than this number, and to press the '>' button if they were more confident than this number. The numbers appearing on the screen were the midpoints of the confidence categories (55, 65, 75, 85, 95). Each of these numbers were presented an equal number of times at each level of signal strength within signal trials, and an equal number of times for both signal and noise trials. Following the greater-than/less-than decision participants rated their confidence.

Results

Response times (RTs) three standard deviations above each participant's mean and RTs less than 200 ms in duration were censored, as were trials where participants expressed a certainty of having made a mistake. This accounted for 1.04% of the 16000 trials. Dependent variables for each cell of the design are participant mean primary decision RT, mean time to render absolute confidence judgements, mean proportion of correct responses, and mean absolute confidence ratings. Proportion correct data were subjected to an arc-sine transformation prior to analysis. Within the cells of the second condition (where participants were asked to decide relative confidence prior to making absolute confidence judgements) mean time to compare confidence to N also served as a dependent variable.

In each of the following analyses of variance (ANOVA), Huynh-Feldt, epsilon adjustment degrees of freedom are used, but the degrees of freedom reported are defined by the design. Level of significance was set at 0.05 throughout and, for the most part, only significant results are reported.

Global analyses

Primary decision times. As expected, as is evident in Panel A of Figure 1, primary decision times decreased reliably as signal strength increased, nearly identically in both the relative and absolute blocks ($F(4, 72)=27.99$, $MS(Error)=848712.26$, partial $\eta^2=.609$).

Detective sensitivity. The plots in Panel B of Figure 1 are also clear in showing that detective sensitivity depended, as expected, on signal strength., $F(4, 72)=22.29$, $MS(Error)=0.43$, partial $\eta^2=.553$.

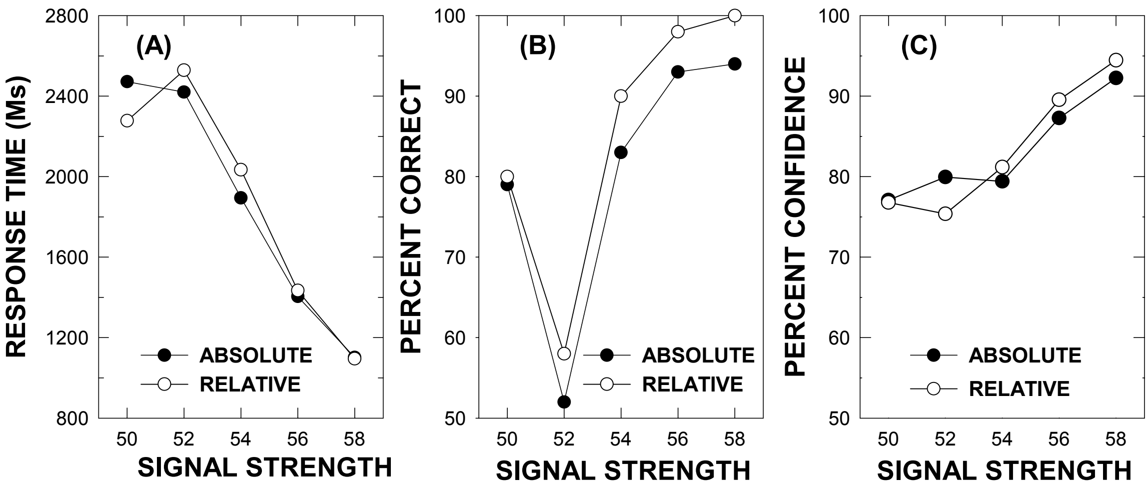

Fig 1. Mean Response Times (Panel A), mean Percent Correct (Panel B), and mean Percent confidence (Panel C) as a function of signal strength for each of the absolute and relative confidence conditions.

Absolute confidence. It is clear from the plots in Panel C of Figure 1, mean confidence also increased as signal strength increased, albeit not as clearly as either RTs or percent correct, in large part because of the massive overconfidence with the weaker signal and noise trials. As well, it should be noted that under-confidence is evident as the decisions become easier with the increases in signal strength. Thus, these data reveal the typical "hard-easy effect" in perceptual tasks (e.g., Baranski & Petrusic, 1994, 1999; Petrusic & Baranski, 1997). Nevertheless, the main effect of signal strength was reliable and uniform over the two types of blocks of trials, $F(4, 72)=47.17$, $MS(Error)=96.67$, partial $\eta^2=.724$.

Confidence times. Confidence times varied systematically with signal strength (i.e., decisional difficulty) in the absolute confidence block but remained approximately constant in the relative confidence block. Indeed, the overall linear trend component was reliable ($F(1, 18)=18.18$, $MS(Error)=119683.79$, partial $\eta^2=.502$. Moreover, the linear component of the strength by block interaction was also significant ($F(1, 18)=5.99$, $MS(Error)=170092.43$, partial $\eta^2=.250$).

Taken together, these analyses of confidence times indicate the presence of a clear post-decisional component in the processing of confidence, precisely as reported earlier in Baranski and Petrusic (2001) and Petrusic and Baranski (2000, 2003). The absence of an effect of signal strength on confidence times in the relative confidence block suggests an absence of a post-decisional component in the processing of confidence. Perhaps confidence processing was completed in parallel with relative confidence processing given that primary decisional RTs in the relative confidence block were not longer than those in the absolute confidence block.

Times to Render Relative Confidence

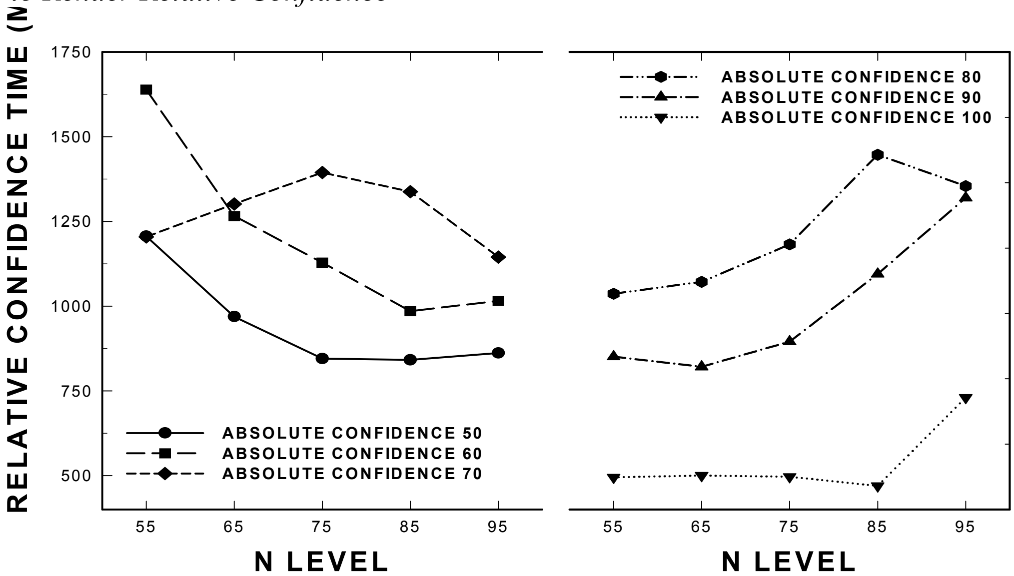

Fig 2. Mean times to decide relative confidence as a function of the comparison N with each level of rendered confidence.

A numerical distance effect would by definition require times to decide relative confidence be longest as N approaches the subsequently selected absolute confidence level. The plots in Figure 2, obtained after averaging over all signal strengths, show that this is precisely what has occurred. For example, when confidence "50" was rendered RTs were maximal at N=55 and then monotonically decreased as N increased. Similarly, for example, when confidence was "70" RTs were maximal at N=75 and then decreased systematically as levels of N decreased and increased. When participants were 100% confident, RTs were maximal at N=95 and remained flat and fast over the other levels of N. Thus, for each level of confidence rendered, RTs were maximal at the level of N nearest to the confidence category selected.

Anchoring effects

Importantly, within the relative confidence block, the level of N seemed to serve as an anchor for participants as they rendered absolute confidence, with participants expressing reliably greater absolute confidence as the level of N increased, $F(4, 72)=15.28$, $MS(Error)=18.69$, partial $\eta^2=.459$ (see Figure 3). Though the effect size of signal strength on mean absolute confidence was larger within this condition, $F(4, 72)=60.53$, $MS(Error)=185.91$, partial $\eta^2=.771$, the level of N does seem to have a reliable influence on the absolute confidence judgements of participants. The N level by signal strength interaction was also significant, $F(16, 288)=19.58$, $MS(Error)=7.60$, partial $\eta^2=.125$, with the 56% density level being the only strength level within which N did not appear to act as an anchor for confidence.

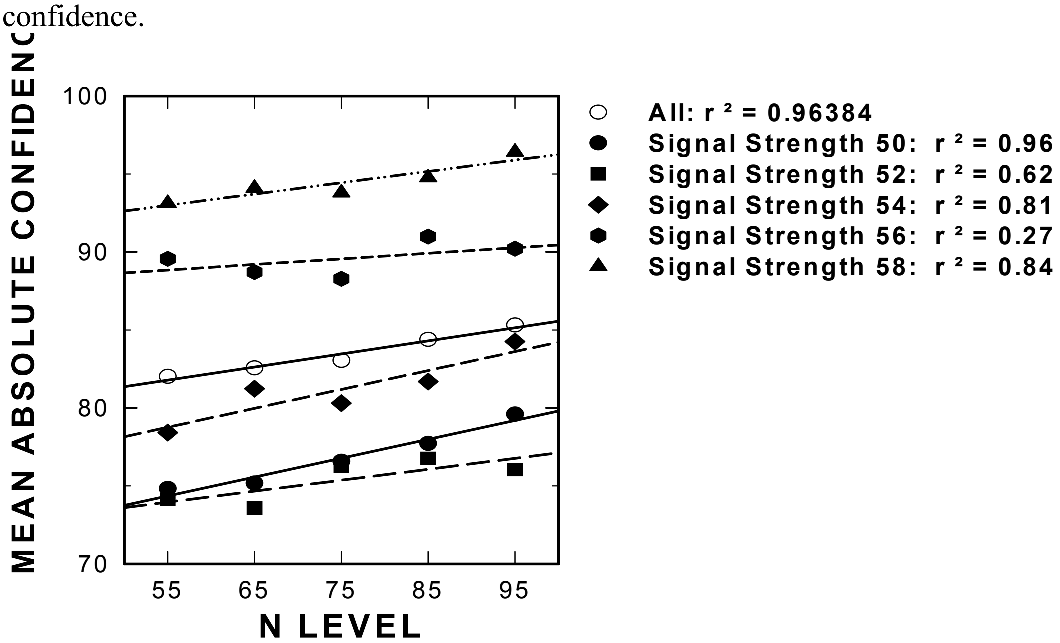

Fig 3. Mean confidence ratings as a function of the anchoring, N level, with each level of signal strength.

Discussion

The present study found that confidence judgements were subject to both numerical distance and anchoring effects. While these findings do not provide direct evidence for a numerically based internal representation of confidence, they do suggest that decision-makers are able to treat confidence as if it represents a magnitude and they seem able to perform arithmetic operations on it (specifically, they are able to calculate a difference between absolute confidence and N).

The present experiment provides clear evidence that, in the meta-cognitive domain, confidence judgements of the correctness of one's decisions can be subject to anchoring effects. As such, these findings converge nicely and extend the recent findings showing anchoring effects in the context of judgments of learning (Scheck, Meeter, & Nelson (2004).

References

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

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.

Macmillan, N.A. & Creelman, C.D. (2005). *Detection theory: A user's guide. (2nd ed.).* Mahwah, N.J. Lawrence Erlbaum Associates.

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.

Moyer, R. S., & Landauer, T. K. (1967), Time required for judgements of numericalal inequality. *Nature*, 215, 1519-1520.

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.

Scheck, P., Meeter, M., & Nelson, T.O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language*, 51, 71-79.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. O*rganizational behaviour and human performance*, 6, 649-744.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61(6), 401-409.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185 (4157), 1124-1131.

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

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1147-1166.

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

Acknowledgement

This research was supported by an Ontario Graduate Scholarship award to Carroll and a Natural Sciences and Engineering Research Council Discovery grant to Petrusic.