SELECTIVE ATTENTION AND CONFIDENCE CALIBRATION

Jordan Schoenherr, Craig Leth-Steensen, and William M. Petrusic
psychophysics.lab@gmail.com, craig_leth_steensen@carleton.ca,
bpetrusi@carleton.ca
Carleton University, Ottawa

Abstract

This experiment examined the effect of attentionally irrelevant visual stimuli on subjective confidence calibration. Participants identified target letters that were flanked by letters that differed with respect to their response congruency to the target response. The key result was the finding of higher overconfidence in incongruent flanker conditions, even when presentation time manipulations were successful in equating accuracies across the three flanker conditions.

This work extends a program of research undertaken by the authors examining the relationship between attentional and confidence processing in perceptual judgement. In our initial work (Schoenherr, Leth-Steensen, & Petrusic, ISP Proceedings, October, 2005), the effects of dividing attention over visual and auditory input modalities on the confidence judgments provided to subsequent unidimensional perceptual discrimination responses were examined. In that work, both confidence calibration and resolution became degraded under divided attentional conditions.

In this work, we attempt to answer the question: How is confidence affected by failures in selective visual attention? We regard this as a very pertinent question in a practical sense given that failures of selective visual attention can occur in everyday life and also that it is typically our metacognitive impressions of what we have just perceived that serve to drive further action. It is also an important question theoretically because the effect of visual context (i.e., within a stimulus display itself) on confidence processing in perceptual judgement has not yet been examined to any great extent and also because such work would represent a novel set of findings that could provide some important new theoretical constraints for existing theories of the confidence process (e.g., Vickers, 1979; Petrusic & Baranski, 1997).

The selective attention paradigm that we used is the Eriksen and Eriksen (1974) flankers task. In this paradigm, an array of letters is presented that consists of a middle target letter and a set of surrounding flanking letters. The task is to focus visual attention on and identify the target letter while selectively ignoring the flankers. The identity of the flankers can either be the same as the target (congruent condition) or different in the sense that they represent stimuli that are either associated with an alternative response to the present target (incongruent condition) or are truly irrelevant with respect to the assigned response mappings (neutral condition). The now classic findings are that responses to incongruent flanker stimuli are slowed and more error prone in comparison to responses to neutral and congruent flanker stimuli.

For our present purposes, it suffices to note that these findings are typically attributed to a failure to selectively attend solely to the middle target letter - although there has been quite a bit of controversy as to whether this failure arises due to late-selection attentional processing or simply to early selection with attentional leakage (Paquet & Lortie, 1990; Yantis & Johnston, 1990) and also with respect to whether it is due to actual perceptual confusion, response competition, or both (and, relatedly, whether or not this effect can be regarded as evidence for a continuous flow, as opposed to a discrete stage, model of information processing; Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Sanders & Lamers, 2002).

The only previous research that is specifically related to our current work was performed by Keren (1988). Keren's main focus was a direct comparison of confidence calibration in perceptual and general knowledge judgement tasks, as there was some controversy at the time as to whether the phenomenon of overconfidence in intellectually based judgements extends to perceptually based ones (for subsequent work supporting this notion see Baranski & Petrusic, 1994). In Keren's experiment of interest here (Experiment 2), participants first viewed a two-letter display for about 50-85 ms and then reported the identity of a target letter that was indicated by a post-stimulus position cue. Similar to the flanker task, the uncued letter could either be the same, neutral, or response incongruent. As is typical of the findings obtained using this paradigm, Keren found that accuracy performance was substantially degraded when the target letter was flanked by itself than by another letter (i.e., a repeated letter inferiority effect of 61 vs. 81% correct, respectively). With respect to confidence measures, though, participants were quite overconfident in the repeated letter condition and quite underconfident otherwise (i.e., the mean confidence ratings were 70 and 74%, respectively - but note that Keren did not regard this as evidence for overconfidence in perceptual judgement because he deemed this task to be only "perceptual-like"). However, two important aspects of Keren's findings were that (a) given the rather counterintuitive nature of these accuracy findings, the cognitive nature of the processes that gave rise to them is not clear, and (b) given the overall observed differences in accuracy, the phenomenon of overconfidence in the repeated letter condition could simply represent a manifestation of the well-known "hard-easy" confidence calibration effect (i.e., that, in general, overconfidence occurs for any kind of difficult judgment and underconfidence for easy ones).

Method

Twenty undergraduates participated in a single 45 min. session. Stimuli consisted of three capital letters, a target in the centre of the array with two identical flankers on either side. Two sets of targets and flankers were used: F, P, K, T and H, N, L, M. The former two letters in each set were used as both targets or flankers and latter two were used only as neutral (non-target) flankers. These stimuli were presented side-by-side in the centre of a computer monitor in standard DOS font size. Immediately following the presentation of the stimulus, a pattern mask (XXX) was presented in the same location as the stimulus.

A 486 computer using Mel 2.0 experimental software randomly presented the stimuli within each block to the participants. In an attempt to minimize practice effects, a short initial training phase was of 20 trials familiarized the participants with the procedure. Two experimental blocks of 144 trials (24 trials per flanker congruency condition for each stimulus set) were performed by each participant, one block where confidence was also required and one block where the primary decision was the sole requirement (their order was counterbalanced across participants).

Each trial began with the cue "Ready" presented in the centre of the top portion of the screen for 2500 ms, after which a stimulus was presented below it. Because we were interested in confidence processing under conditions within which a substantial amount of errors are being made, our flanker stimuli were presented for only very short durations.

Moreover, in order to try to minimize differences in accuracy so as to selectively isolate any effects on confidence processing that result solely from the difference in congruency conditions, the congruent, neutral, and incongruent flanker stimuli were presented for 50 ms, 66 ms, and 116 ms, respectively.

Using the index fingers of each hand, participants selected the appropriate response key (either “z” or “/”) when they had determined the identity of the target. Two targets, one from each different stimulus set (F, H and P, N, respectively) were assigned to each response key. If the block of trials required confidence judgements, once the primary decision had been rendered, a prompt to indicate the level of confidence that they had in their response would follow. Confidence was indicated by pressing number keys (on the top row of the computer keyboard) labelled from 50 to 100% confidence. For the no-confidence block, a corresponding prompt appeared that simply instructed participants to press the space bar.

Results

Two separate 2 (confidence block) × 3 (flanker congruency) × 2 (stimulus set) × 2 (confidence block order) ANOVAs, with the latter factor between subjects, were performed on the mean participant decisional response times (RTs) and proportions correct (PCs). For these analyses, PCs were arcsine transformed and RTs beyond 3 *SD*s of the overall mean RT of each participant as well as those below 200 ms were excluded (2.83%). All significant results are reported (with their Greenhouse-Geisser adjusted levels of significance).

In the PC analyses, the main effect of confidence block was significant $F(1, 18) = 7.03$, $p \leq .016$, such that when confidence judgements were required, participants were more accurate (.744 vs. .708 , respectively). Stimulus set was also significant $F(1, 18) = 39.26$, $p \leq .001$. Flanker congruency, $F(2, 36) = 1.50$, $p \leq .240$, was not significant, although it did interact significantly with stimulus set, $F(2, 36) = 3.62$, $p \leq .046$. Importantly, mean PCs were very similar across flanker congruency conditions for the more accurately responded to F, P stimulus Set 1 (see Table 1) indicating that, at least for this stimulus set, stimulus presentation time manipulations were successful in eliminating accuracy differences. No PC effects associated with confidence block order were significant.

In the RT analyses, the main effect of confidence block was significant $F(1, 18) = 13.69$, $p \leq .002$, such that when confidence judgements were required, participants showed dramatic increases in RT (1405 vs. 1088 ms, respectively). Flanker congruency, $F(2, 36) = 15.47$, $p \leq .001$, was also significant (1233, 1338, and 1168 for the congruent, neutral, and incongruent flanker conditions, respectively - although this is not the typical flanker RT effect note that stimulus presentation times differed in each congruency condition in this experiment). Stimulus set was also significant $F(1, 18) = 14.38$, $p \leq .001$, as was the two-way interaction of Stimulus Set x Confidence Block, $F(1, 18) = 6.85$, $p \leq .017$. RTs were longer for the H, N stimulus set but less so when confidence was not required (see Table 2).

Confidence block order interacted significantly with confidence block, $F(1, 18) = 10.60$, $p \leq .004$, such that the RT difference when confidence was and was not required was much smaller when the confidence judgement block was performed second (38 ms) than when it was performed first (597 ms), a finding which almost certainly reflects the influence of practice with the letter identification task. Confidence block order also interacted significantly with flanker congruency, $F(2, 36) = 4.73$, $p \leq .022$, and entered into a significant three-way interaction with both confidence block and flanker congruency together, $F(2, 36) =3.79$, $p \leq .041$, due to the presence of very slow RTs to neutral flanker stimuli in the confidence judgement block for the participants who performed this block first.

Table 1: Average PC by Confidence Block, Flanker Congruency, and Stimulus Set

		Congruent	Neutral	Incongruent
CON	Stimulus Set 1	0.804	0.821	0.813
	Stimulus Set 2	0.694	0.706	0.625
	M	**0.749**	**0.764**	**0.719**
NC	Stimulus Set 1	0.773	0.808	0.752
	Stimulus Set 2	0.694	0.654	0.565
	M	**0.733**	**0.731**	**0.658**

Table 2: Average RT by Confidence Block, Flanker Congruency, and Stimulus Set

		Congruent	Neutral	Incongruent
CON	Stimulus Set 1	1307	1384	1234
	Stimulus Set 2	1472	1630	1403
	M	**1389**	**1507**	**1318**
NC	Stimulus Set 1	1029	1143	945
	Stimulus Set 2	1123	1195	1089
	M	**1076**	**1169**	**1017**

Four separate 3 (flanker congruency) × 2 (stimulus set) ANOVAs were performed on the individual mean confidence measures and the calibration indices of over/underconfidence, calibration, and resolution (Baranski & Petrusic, 1994; group confidence calibration curves are given in Figure 1). Mean confidence level was significantly affected by both flanker congruency, $F(2, 38) = 9.45$, $p \leq .001$, and stimulus set, $F(1, 19) = 43.20$, $p \leq .001$ (see Table 3). Importantly, over/underconfidence also exhibited significant effects of both flanker congruency, $F(2, 38) = 6.08$, $p \leq .005$, and stimulus set, $F(1, 19) = 6.84$, $p \leq .017$. Overconfidence was essentially twice as high for incongruent flanker stimuli.

Calibration also differed significantly with flanker congruency, $F(2, 38) = 3.35$, $p \leq .046$, and stimulus set, $F(1, 19) = 10.72$, $p \leq 0.004$. Moreover, for calibration the two-way interaction of Flanker Congruency x Stimulus Set was also significant, $F(2, 38) = 3.73$, $p \leq .033$. Calibration was worse for incongruent flanker stimuli, and especially so for the H, N stimulus set. Perhaps surprisingly, however, resolution was only significantly affected by stimulus set, $F(1, 19) = 5.59$, $p \leq 0.29$.

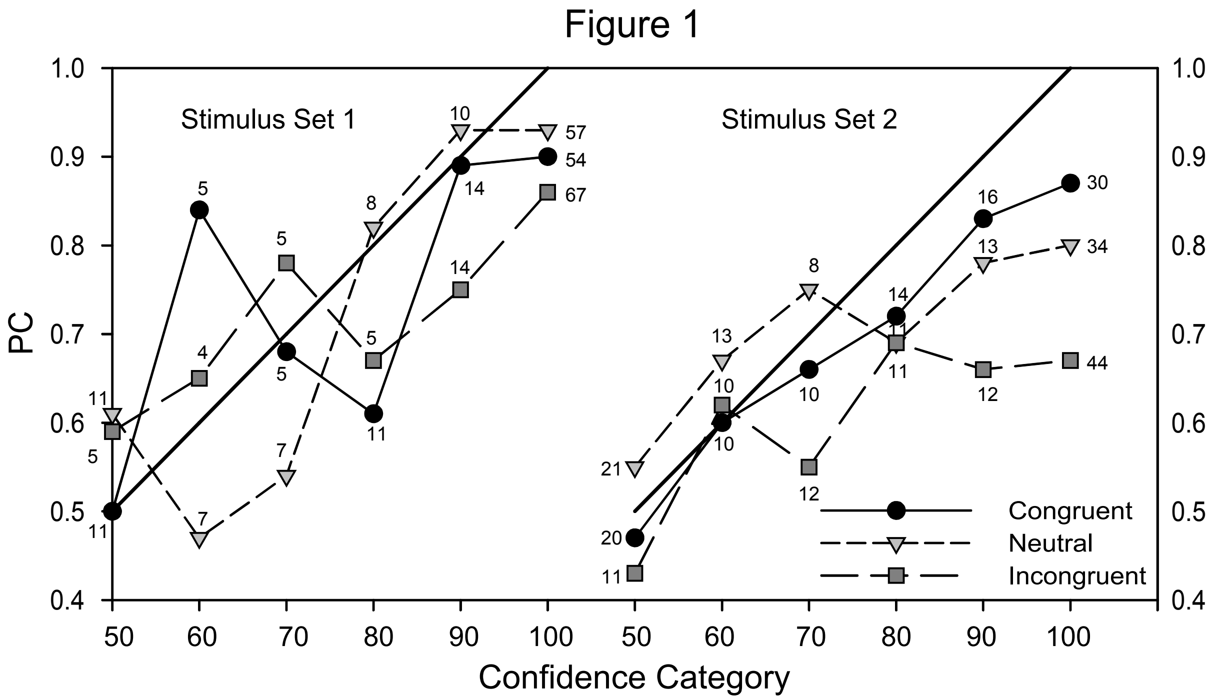
Table 3: Average Calibration Indices by Stimulus Set and Flanker Congruency

Set	Congruency	P(correct)	M_{Conf}	O/U	Cal	Res
1	Congruent	0.804	87.41	0.070	0.06	0.034
	Neutral	0.821	87.09	0.050	0.047	0.038
	Incongruent	0.813	92.5	0.112	0.072	0.013
2	Congruent	0.694	78.77	0.094	0.069	0.056
	Neutral	0.706	79.15	0.086	0.074	0.03
	Incongruent	0.625	83.71	0.212	0.138	0.041
M		**0.744**	**84.77**	**0.104**	**0.077**	**0.035**

Discussion

The key result from this study (among several others that will not be discussed here) was the finding of sizeably enhanced overconfidence in the incongruent flanker condition. The fact that it occurred even for the case where accuracy was closely equated across flanker congruency conditions (i.e., stimulus Set 1), allows us to conclude that it does not simply represent a manifestation of the hard-easy effect but, instead, indicates that failures of selective attention can have real and quite dramatic effects on confidence. More specifically, it suggests that the nature of the errors being made in the incongruent flanker condition was qualitatively quite different from that in the other two flanker conditions, in that for the former case, such errors were associated with much higher overall confidence (indeed, it seems that in the incongruent flanker condition, participants were often completely unaware that their selective attention had failed).

Such findings are also consistent with those obtained in previous investigations of the effect of incongruent flanker stimuli on measures of partial response activation at the psychophysiological level. In particular, Coles et al. (1985) showed that about half of the time when an error occurred to their incongruent flanker stimuli (but much less so proportionally for errors to their congruent flanker stimuli), no response-related activation (in terms of subthreshold dynamometer squeezing) was present on the side of the correct response. That finding indicates that much of the time when an incorrect response to incongruent flanker stimuli “wins out”, it “wins big”. Correspondingly, in terms of theories of the confidence process, most evidence-accrual-based theories specifically assume that confidence is derived by comparing the amounts of covertly accumulated evidence for each response, with higher confidence reports being associated with more discrepancy between those amounts. Hence, we conclude that it is these kinds of theories that are the most promising with respect to providing an account for our findings.



References

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

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception & Performance*, *11*, 529-553.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics*, *16*, 143-149.

Keren, G. (1988). On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration studies. *Acta Psychologica*, *67*, 95-119.

Paquet, L. & Lortie, C. (1990). Evidence for early selection: Precuing target location reduces intereference from same-category distractors. *Perception & Psychophysics*, *48*, 382-388.

Petrusic, W. M., & Baranski, J. V. (1997). Context, feedback, and the calibration and resolution of confidence in perceptual judgements. *American Journal of Psychology*, *110*, 543-572.

Sanders, A. F., & Lamers, J. M. (2002). The Eriksen flanker effect revisited. *Acta Psychologia*, *109*, 41-56.

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

Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focussed attention tasks. *Journal of Experimental Psychology: Human Perception & Performance*, *16*, 135-149.