Supplementary Materials

Method

Full methodological details can be found in Rae et al. (2014), and Osth et al., (2017) original manuscripts. Here we report an abridged method for the reader's ease.

Participants were 48 in Rae et al., and 46 in in Osth et al., from the University of Newcastle, Australia, with self-reported normal or corrected to normal vision. Stimuli were nouns with frequency and concreteness ratings from the MRC Psycholinguistics database.

Participants completed two 1-hour sessions on different days. Session one consisted of practise and experimental blocks. In the practise blocks participants were presented with practise study-test lists in one speed and one accuracy condition. Instructions for each condition emphasised speed or accuracy of responses, respectively. In the experimental block, participants were presented with four speed and four accuracy study-test lists. Session two consisted of experimental blocks only with the same structure as session one. Condition order was alternated within and between sessions and counterbalanced between participants, with a 30 second rest break between conditions.

In the study phase words from the study list were presented one at a time on the centre of the monitor and the participant required to read them aloud. On each trial, a fixation period of 150 milliseconds (ms) was proceeded by study words that remained on screen for 1000 ms. A 250 ms blank period followed each trial. A 7000 ms warning to prepare for the test phase as either a speed or accuracy condition appeared on the screen between the study and test phases.

In the test phase words for the test list were presented one at a time on the centre of the monitor. On all trials, words appeared until a response of "old" or "new" was made using the "z" or "/" keys, or 6000 ms elapsed, at which point the warning

"TIME LIMIT EXCEEDED! NO RESPONSE RECORDED" was displayed on the screen. Response keys were counterbalanced between participants. In the speed condition a "TOO SLOW" warning appeared on the screen for any response slower than 650 ms, and a "TOO FAST" warning appeared on the screen for any response faster than 250 ms. In the accuracy condition "CORRECT" and "INCORRECT" was displayed after correct and incorrect responses, respectively. Accuracy feedback, and all other warnings, appeared for 500 ms. A 750 ms blank period followed each trial. In Osth et al. (2017), a the completion of each trial of the test phase, participants were also asked to rate their confidence in their previous response.

Results

Confirmation of Manipulation

In Rae et al., (2014) participants were more accurate and responded more slowly when accuracy (M = 771ms, M = 81.5%) was emphasised over speed (M = 549ms, M = 73.6%). Paired samples t-tests showed these differences to be significant for both accuracy, $t(46) = 10.34$, $p<.001$, and mean RT, $t(46) = 12.08$, $p<.001$. The results for Osth et al., (2017) replicated this pattern when accuracy (M = 928ms, M = 78%) was emphasised over speed (M = 684ms, M = 68.4%) for both accuracy, $t(45) = 8.75$, $p<.001$, and mean RT, $t(45) = 12.81$, $p<.001$. The results for Osth et al., (2017) were slower and less accurate overall, likely due to the increased load caused by the addition of confidence responses.

Consistent with differing dominance of error types, in Rae et al., (2014) paired t-tests showed that under accuracy emphasis, the mean RT for errors (M = 848ms) was considerably slower than correct responses (M = 757ms), $t(46) = 8.31$, p <.001, but under speed emphasis, there was no difference between the mean RT for errors (M = 550.8ms) and correct responses (M = 551.5ms), $t(46) = .13$, p = .90. This pattern was again replicated in Osth et al., (2017) where under accuracy emphasis the mean RT for errors (*M* = 998ms) was slower than correct responses (*M* = 917ms), $t(45) = 5.28$, $p < .001$, but under speed emphasis the mean RT for errors (*M* = 679ms) and correct responses (*M* = 692ms) was similar, $t(45) = 1.42$, $p = .16$.

Figure 1 provides quantile conditional accuracy functions (Q-CAFs; Ridderinkhof, 2002) from both experiments to explore these results further. They allow a detailed examination of the relative distribution of errors and correct responses under speed and accuracy emphasis by showing the probability of an error in RT bins bounded by quartiles (i.e., the fastest 25% of responses, the 2nd fastest 25%, and so on).

The top row of Figure 1 highlights that under speed emphasis, participants in both experiments made more errors in the fastest RT bin than in other bins. Under accuracy emphasis on the other hand, participants made more errors in the slowest RT bin than in other bins. The overall pattern demonstrates a successful manipulation of the distribution of errors, and thus the proportion of errors types, between the two conditions. The bottom row of Figure 1 represent the same data, but using the median RT, rather than percentiles, as the x-axis measure. This highlights that the RT bins common to both conditions represent the fastest responses in the accuracy condition and the slowest in the speed condition.

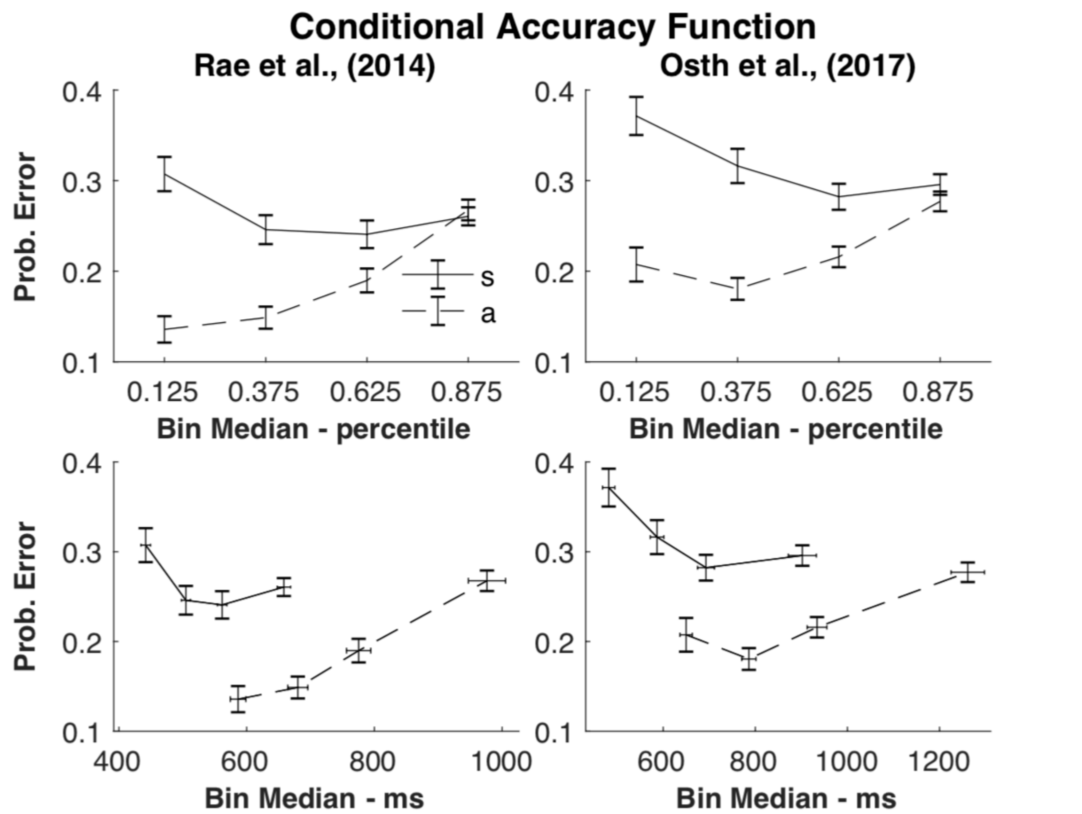

Figure 1. Quantile conditional accuracy functions for and for speed and accuracy emphasis. ‘s’ = speed emphasis condition. ‘a’ = accuracy emphasis condition. The probability of an error in each bin is calculated for each participant in each bin, and then averaged across participants. In the left plot bins are data for Rae et al., (2014). In the right plot bins are data for Osth et al., (2017). In the top row RT bins are represented by their median percentile, while on the bottom row RT bins are represented by their median response time. X-axis error bars represent the across participant standard error of the bin median, while Y-axis error bars represent the within-subjects standard error of the mean (Morey, 2008).

Measuring Post-Error Changes

Post-error-adjustment measures aim to quantify the short-term sequential (*local*) effect of previous trial outcomes on RT and accuracy for the subsequent trial. The *standard* method of quantifying post-error slowing involves subtracting the mean RT of each participant's post-error trials from the mean RT of their post-correct trials. This type of global averaging has potential drawbacks. For example, local effects like post-error slowing can be confounded with long-term (global) effects like tiredness, distraction, or boredom (Dutilh et. al., 2012). Dutilh et. al., (2012) proposed a solution where post-error trials are paired with their immediately preceding pre-error trials (that are also post-correct trials), and pairwise differences calculated (i.e., post-error RT – pre-error and post-correct RT). The mean of the differences is described as a *robust* measure of post-error RT changes. Dutilh et. al., (2012) showed the robust method is able to differentiate true post-error RT changes from global sequential effects in both empirical and simulated data. Dutilh et. al. (2013) also extended the robust method to post-error accuracy changes by pairing post-error and pre-error accuracy.

A second drawback is that when errors differ in speed from correct responses, post-error effects may be confused with effects arising from the speed of the previous response (Hajcak & Simons, 2002). Hajcak and Simons (2002) suggested comparing post-error and post-correct performance for pairs of correct and error response closely matched on RT. They developed this *matched* approach to protect against artifact confounds in ERP's but the reasoning is equally valid for behavioral changes - it is plausible that participants are sensitive to the speed of their previous response and therefore their subsequent response could systematically differ based on this perception (Williams et al., 2016).

Statistical Results for Within-Condition Analyses

Table 1. Two-Way Within Subject ANOVA results for Experiment 1.

	df			*F*			*p*		
	R	**M**	**S**	**R**	**M**	**S**	**R**	**M**	**S**
Emphasis	1	1	1	8.77	17.92	17.24	.005	<.001	<.001
Error Position	1	1	1	4.94	4.74	1.7	.032	.035	.199
Emphasis * Error Position	1	1	1	2.02	.16	.11	.162	.695	.739
Error	43	44	44						

*note: Figures represent results for the robust, matched, and standard calculation methods. Participants who had less than 5 suitable errors in each cell were removed entirely from analyses. Raising this cut-off to 10 suitable errors strengthens the statistical reliability of these results, however, 15 participants would be excluded for the robust method. ‘R’, ‘M’, and ‘S’ indicate robust, matched, and standard, respectively.

Table 2. Two-Way Within Subject ANOVA results for Experiment 2.

	df			*F*			*p*		
	R	**M**	**S**	**R**	**M**	**S**	**R**	**M**	**S**
Emphasis	1	1	1	3.33	16.73	17.57	.075	<.001	<.001
Error Position	1	1	1	8.01	5.12	2.9	.007	.029	.096
Emphasis * Error Position	1	1	1	.55	.27	.55	.46	.604	.464
Error	44	44	44						

*note: Figures represent results for the robust, matched, and standard calculation methods. Participants who had less than 5 suitable errors in each cell were removed entirely from analyses. 'R', 'M', and 'S' indicate robust, matched, and standard, respectively.

Post-Error Accuracy Analyses

Figure 2 shows post-error accuracy results for both experiments. For Rae et al., (2014) the data suggest errors had no interpretable effect on accuracy, and for Osth et al., (2017) the data indicate a clear decrease in accuracy following an error, which may be larger for the accuracy condition. Paired-sample t-tests bear out these visual interpretations. For Rae et al., (2014), the difference between conditions was not significant for any method. For Osth et al, (2017) the decrease in accuracy following errors is larger under accuracy emphasis for both the matched (Ms = -.019, Ma = -.052, t[45] = 3.10, p = .003) and standard (Ms = -.025, Ma = -.059, t[45] = 3.50, p = .001) calculation methods, but not the robust method. The addition of confidence responses in the Osth et al (2017) experiment errors appeared to result in errors having a more deleterious impact on accuracy. This effect was largest in the accuracy emphasis condition, where we also found considerable post-error speeding. We elaborate further on the confidence manipulation in the following section.

The data from both experiments conclusively indicate that accuracy did not improve following errors, even in the speed emphasis conditions where we found post-error slowing. We return to the theoretical implications of these results in discussion.

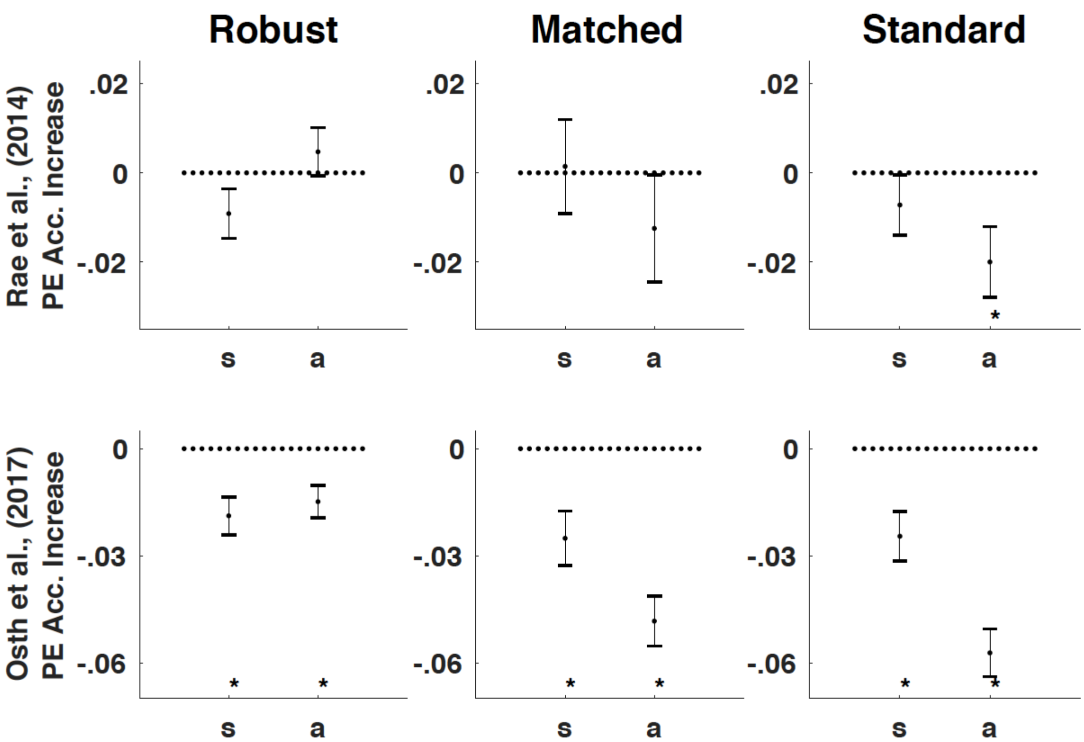

Figure 2. The across-participant post-error accuracy changes for speed and accuracy emphasis for the robust, matched, and standard calculation methods. The top row of plots represents data for Rae et al., (2014). The bottom row of plots represents data from Osth et al., (2017). Above the dotted line indicates an increase in accuracy following errors. 's' and 'a' indicate speed and accuracy, respectively, while '*' indicates a two-tailed t-test using a critical p-value of .05 showed a significant difference from zero for the corresponding across-participant distribution. Error bars indicate the standard error of the mean.

Analyses of Confidence Manipulation in Osth et al., (2017)

Rabbitt and Vyas (1970) demonstrated error awareness by showing participants could make a response signalling or correcting an error about 80% of the time. In recognition memory, Van Zandt and Maldonado-Molina (2004) showed that participants were typically aware of response-speed errors and, if permitted to do so, reversed their erroneous responses.

Based on these findings, in Osth et al.'s (2017) experiment we expected that under speed emphasis and following an error, 'low confidence' (in being correct) responses would likely reflect error awareness, which has been linked to an increase in post-error slowing (Nieuwenhuis et al., 2001; see Danielmeier and Ullsperger, 2011, for a review). It is possible then that under speed emphasis we will see increased post-error slowing for low confidence responses compared to high confidence responses.

Given the paucity of research regarding evidence-quality errors, it is difficult to empirically predict how confidence may influence post-error changes under accuracy emphasis. Theoretically, it follows from our thesis that errors under accuracy emphasis are more likely to occur when the evidence is equivocal or systematically favours the incorrect decision. It is possible then that a low-confidence response might indicate awareness of poor evidence-quality, rather than awareness of an error per se. If so, it may accentuate speeding tendencies. Conversely, a belief that evidence quality is good (i.e., high-confidence responses) may accentuate slowing tendencies.

In sum, we suspect an interaction between confidence and speed vs. accuracy emphasis. Under speed emphasis we might expect more post-error slowing following low confidence errors compared to high confidence errors. Under accuracy emphasis we might expect more post-error speeding following low confidence errors compared to high confidence errors.

We performed a two-way within-subjects ANOVA for each calculation method, this time using emphasis (speed/accuracy) and confidence response (high/low) as factors. Figure 2 suggests that under speed emphasis, the amount of post-error slowing was similar for both high and low confidence responses. However, under accuracy emphasis, low confidence responses tended toward more post-error speeding.

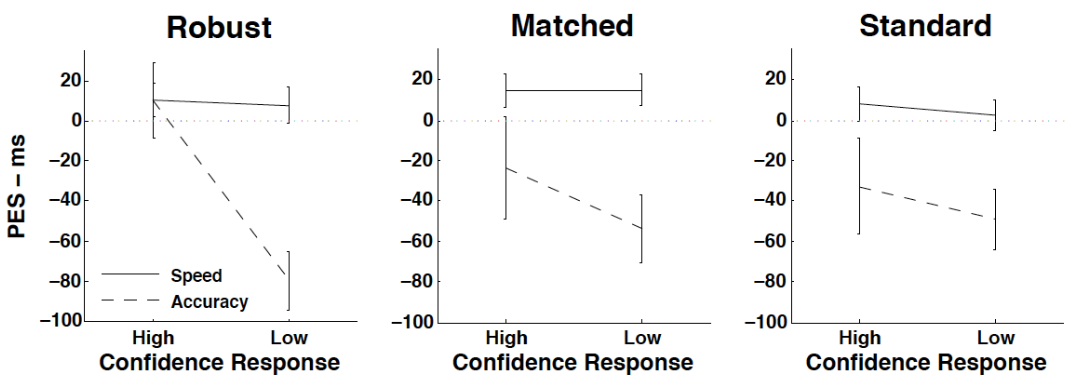

Figure 2. Post-error RT changes using the robust, matched, and standard calculation methods for each confidence response (high/low) and each level of condition (speed/accuracy). Participants who had less than 5 suitable errors in a cell were removed entirely from analyses. Error bars represent the within-subjects standard error of the mean (Morey, 2008). Statistical comparisons appear in Table 3.

Table 3 confirms that the robust method found strong evidence for an interaction between emphasis and confidence response, while the matched and standard methods found no reliable effects of confidence response. The strong finding by the robust method of more post-error speeding following low confidence responses suggests longer-term effects may have masked the effect under the matched and standard measurements methods. In particular, under accuracy emphasis high-confidence error responses could have occurred in a generally slower region, increasing apparent post-error speeding, whereas low-confidence errors could have occurred in a generally faster region, decreasing post-error speeding.

Table 3. Two-Way Within Subject ANOVA results for Experiment 2 confidence analysis.

	df			*F*			*p*		
	R	**M**	**S**	**R**	**M**	**S**	**R**	**M**	**S**
Emphasis	1	1	1	12	12.43	9.57	.001	.001	.003
Confidence Response	1	1	1	17.43	1.68	1.16	<.001	.202	.288
Emphasis * Confidence Response	1	1	1	22.04	1.56	1.43	<.001	.219	.548
Error	44	44	44						

*note: Figures represent results for the robust, matched, and standard calculation methods. Participants who had less than 5 suitable errors in each cell were removed entirely from analyses. ‘R’, ‘M’, and ‘S’ indicate robust, matched, and standard, respectively.

Additional Supplementary References

Danielmeier, C., & Ullsperger, M. (2011). Post-error changes. *Frontiers in Psychology, 2*, 1-9. doi: 10.3389/fpsyg.2011.00233

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology, 38*, 752-760.

Ridderinkhof, K. R. (2002). Micro- and macro-changes of task set: activation and suppression in conflict tasks. *Psychological Research, 66*, 312-323. doi: 10.1007/s00426-002-0104-7

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1147. doi.org/10.1037/0278-7393.30.6.1147