ASSESSING POSNER'S THEORY OF ALERTING: A META-ANALYSIS OF SPEED-ACCURACY EFFECTS

Colin R. McCormick & John Christie

Abstract

Posner and his colleagues proposed a seminal theory of how alerting influenced information processing over 50 years ago (Posner, Klein, Summers, and Buggie, 1973). In this study, participants were presented with warning signals at varying intervals before a target, and participants were asked to produce a spatial discrimination response. Trials in which participants were played a warning signal were compared to trials without a warning signal to understand the effect of phasic alerting using reaction time (RT) and error rate (ER). Posner and colleagues observed a general speed-accuracy trade-off (SAT) across conditions, in which faster RTs led to higher ER, and concluded that phasic alertness shifts response criteria without improving the efficiency of information processing. More recent research has questioned whether this theory of alerting applies generally across all time-courses and conditions. The current meta-analysis aimed to test Posner's theory of alerting (1975) using all available data in the field that closely matches the methodology used in Posner et al.'s (1973) influential study. After including data from sixteen published experiments across three different signal-target foreperiod durations, our conclusions support that while a speed-accuracy trade-off is likely present at shorter foreperiods (50 msec), the longer foreperiods (200 and 400 msec) show evidence of an increase in the rate of information processing when the participant was alerted.

Keywords: attention, temporal processing

Open Practices Statement

The R code used to conduct this meta-analysis, as well as the data that was pulled from the relevant papers, is available at www.osf.io/y4wb7. This was not a pre-registered project.

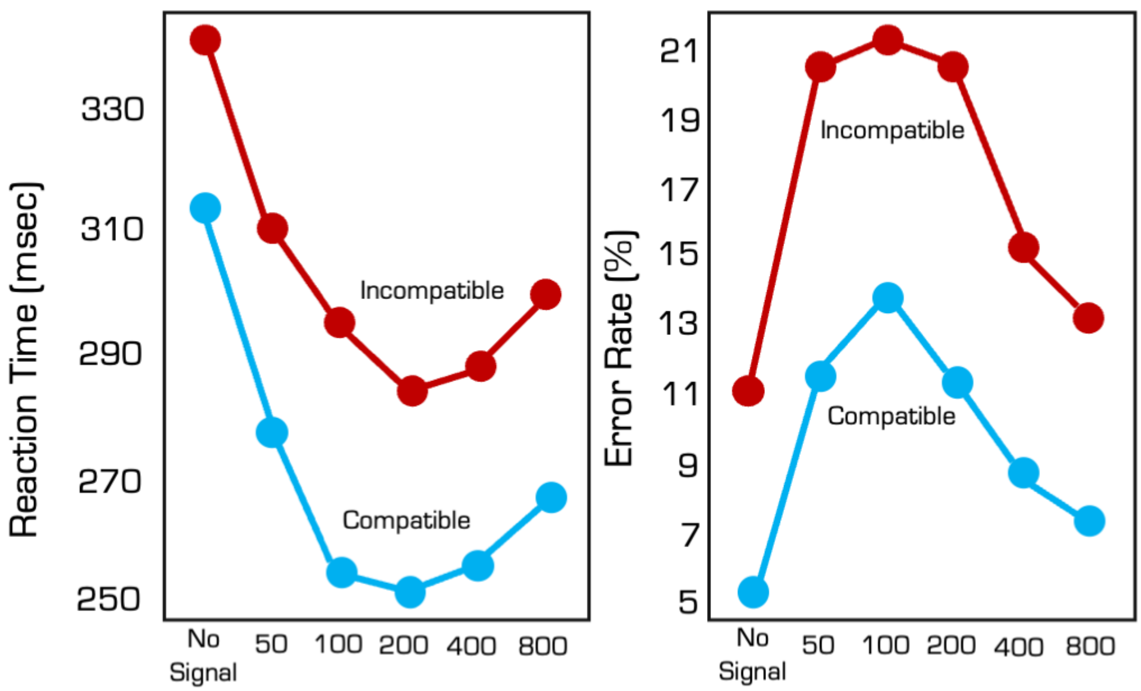
Introduction

A popularized approach to understanding attention is categorizing it as three isolable networks that influences the processing of information (Petersen & Posner, 2012; Xuan et al., 2016). Alerting is one of these networks, and its activation helps generate a heightened sensitivity to external stimuli (Posner & Petersen, 1990). This allows an individual to be in a state of increased response readiness, providing a behavioural advantage when response speed is required (Posner, 2008). Typically, experiments on alerting will initiate the state through either an auditory or visual stimulus, often called a 'warning signal'. The warning signal can provide a varying degree of temporal certainty as to when this target will be presented, depending on the experimental design. Research studying how warning signals impact performance extends over a century, with questions related to finding the optimal foreperiod (interval of time) between the signal and the target, how the consistency of a foreperiod impacts performance, and how it impacts the speed at which individuals accumulate information about the target (Niemi & Näätänen, 1981).

Almost 50 years ago, Posner published a seminal theory on how alertness impacts information processing (Posner, 1975). This theory was informed through the analysis of speed and accuracy performance within alerting paradigms that presented warning signals before targets. In a task in which participants were asked to provide a speeded response discriminating whether two letter stimuli were the same or different, Posner and Boies manipulated the encoding of information using form cues that provided one of the letter stimuli in advance, as well as alerting via a warning signal which provided timing information (1971). These pre-target stimuli could be presented by themselves, or together. The encoding and alerting effects were additive with one another for response speed when presented in tandem, and the alerting signal did not increase the speed at which the encoding cue information could be utilized to inform the response. Posner and Boies reported that the rate of information buildup was unaffected by alerting (1971). In a follow-up study, Posner et al. (1973) identified that error rates (ER) were quite low within Posner and Boies' (1971) data, and the task only covered a limited number of foreperiod durations. Error rates are an important performance metric to determine whether the improvements to response speed associated with alerting are because participants are trading off accuracy for improved speed. Posner et al. (1973) sought to test this outcome with a follow-up experiment that asked for participants to provide a speeded response as to what side of the screen a target was presented and manipulated response compatibility to increase error rate through upping task difficulty (compatible= right target requires a right button response; incompatible= right target requires a left button response). Posner et al. included a range of foreperiods conditions between the warning signal and the target (50 milliseconds (msec), 100msec, 200msec, 400msec, 800msec) along with a no warning signal condition to map out the temporal nature of alerting effects. The results of this experiment showed a clear U-shaped pattern for RT (reaction time), in which the fastest responses were found between the 100 to 400msec foreperiods, along with an inverted U-shape for ER that peaked around 100msec (see Figure 1).

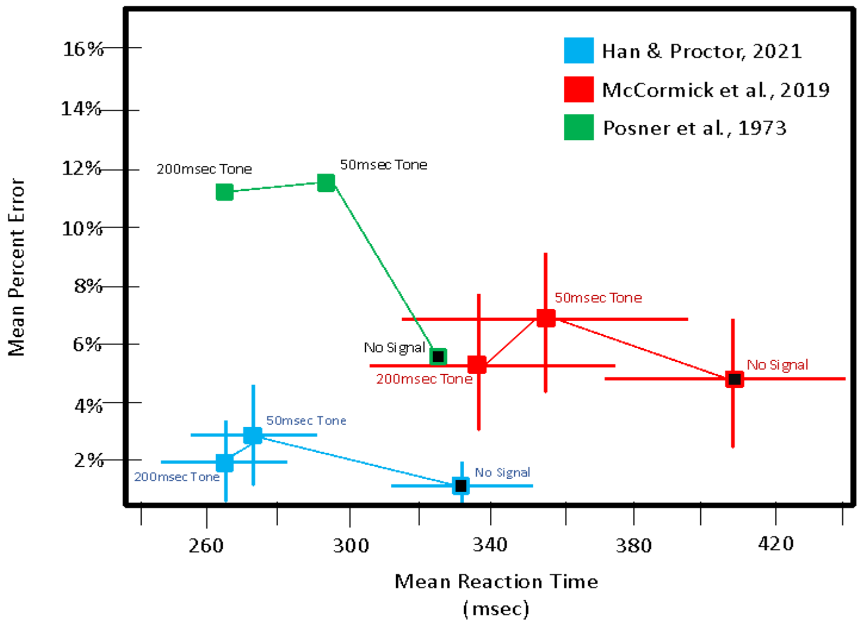
In general, lower RTs were associated with higher ERs. Posner et al. used the significant main effect of foreperiod within their Analysis of Variance (ANOVA) for both RT and ER, along with the visualized U- shaped pattern of results, to conclude that alerting does not increase information accumulation speed in participants. Instead, alerting was theorized to shift response criterion, so that responses are generated at a point in time in which less information about the target had been accumulated. This shift in response criterion is referred to as a speed-accuracy trade- off (SAT), wherein one forfeits accuracy performance to improve response speed, or vice versa.

Figure 1. Redrawn from the results of Posner, Klein, Summers, and Buggie's 1973 experiment. The '0' foreperiod condition represents the 'no warning signal' condition, and the various other foreperiod conditions represent the interval between the signal and target onsets.



The theory generated from this research paper long stood as a cornerstone of alerting research. There has recently been a renewed interest in this topic. McCormick, Redden, Hurst, and Klein (2019) replicated Posner et al.'s (1973) experiment with a larger sample size. McCormick et al. declared that the results from the original study were reproduced, as they obtained the same significant effects of foreperiod from their ANOVAs as Posner et al.'s study, along with similar U-Shaped patterns for RT and ER. Han and Proctor (2022) discovered a pattern that had been overlooked by McCormick et al. (2019) when also closely replicating key features of Posner et al.'s methods (1973). When providing RT feedback on each trial, Han and Proctor had comparable RT effects to Posner et al (1973) and McCormick et al's (2019) experiments in both the 50msec and 200msec condition but noted that ER effects were much smaller than the original study. This difference in effect size is also true for McCormick et al. (2019) but went unnoticed because these authors were fixated on the outcomes of the ANOVA to test the hypothesis. In addition, Han and Proctor's (2022) replication without RT feedback shows RT effects are still present in a similar magnitude to Posner's study, but there is a numeric improvement in the signaled trials compared to the non-signal trials for ER. Based on these outcomes, Han and Proctor make the argument that while a SAT may be present when contrasting the 50 msec signaled foreperiod condition from the no signal condition, such a shift in criterion cannot fully explain what is going on at the 200 msec contrast, as RT is faster with a numerically smaller ER in comparison to the 50 msec foreperiod condition. Therefore, at least some of the speed improvement associated with alerting may be a result of an increase in the rate at which information accumulates. In another set of experiments that conceptually replicated Posner et al., Los and Schut (2008) also showed that faster response times were not as associated with the same increases in error that Posner's theory of alerting would have predicted. However, recently Klein (2023) reanalyzed this data from Los and Schut (2008) in an effort to increase the power of the analysis. To achieve this increased analytic power, Klein collapsed the data across the different experiments and foreperiod conditions and found a similar pattern of results as Posner et al. (1973), in which the fastest-half of conditions, based on mean RTs, generated more errors than the slower half of conditions. Considering that there have been recent studies that both support (McCormick et al., 2019; Klein, 2023) and challenge (Los and Schut, 2008; Han and Proctor, 2022) Posner's theory of alerting (1975), a meta-analysis is warranted to establish consensus within the field and reassess Posner’s theory using the aggregate speed and accuracy data.

Figure 2 Posner et al. (1973), McCormick et al. (2019), and Han and Proctor (with feedback; 2022). The x-axis is mean reaction time, while the y-axis is mean percent error. The ‘no signal’ conditions are filled in with black, and this represents the reference point to compare the two other tone conditions within each study. This figure helps with understanding SATs, as movements left and up represent a trade-off between speed and accuracy, and movements left and flat or left and down represent pure improvements to performance. Error bars are standard deviation of mean (data unavailable for Posner et al., 1973).



The current meta-analysis looks to address Posner’s theory of alerting by analyzing the relationship between alerting and speed-accuracy performance across a collection of experiments with very similar methodologies. The analysis will involve comparing RT and ER effects between trials in which the participants received an alerting signal, and trials where participants do not receive an alerting signal, across three different foreperiod conditions. This will allow for us to apply the logic used in Posner et al. to see if the warning-signal improvements to RT expected across all three foreperiod conditions are most likely a result of a shift in response criterion (meaning that RT improvements are associated with increased ER effects), or whether there is evidence of improvements to information processing (improvements to RT are not associated with increasing ER effects). The three foreperiods chosen (50, 200, and 400 msec) represent a miniature U-Shaped RT pattern like the one presented in Posner et al. (1973). Based on the three previously presented studies which motivated the current analysis (Posner et al, 1973; McCormick et al., 2019; Han and Proctor, 2022), there is the strongest evidence of a SAT when the foreperiod between signal and target is 50 msec, but once the foreperiod reaches 200 msec, alerting appears to improve information processing and reduce signaled trial RT without additional cost to ER.

Within this analysis, we will be contrasting the size of RT effects, between signaled and no-signal trials, with their associated ER effects. This analysis will also permit an assessment of the typical RT and ER effect size across the different foreperiods. To correct for differences in the overall accuracy across experiments, additional analyses will be performed using post-hoc log-odds transformations of error rate. This compensates for accuracy changes closer to perfect performance, as these represent larger shifts in the probability of outcomes and should be treated as larger effect sizes (Jaeger, 2008, Dixon, 2008). This will not generate the same values as a true logistic regression but provides a useful approximation to handle the differences in error rate across experiments. It can be seen in Figure 2 that the overall ER values can vary quite a bit, and this impacts the comparisons which can be made. Han and Proctor had a shift in ER performance from .7% in the no signal condition to 3.1% in the 50 msec foreperiod condition, which in log odds translates to a difference of 1.51, whereas Poser saw a shift from 5.2% to 11.6%, which translates to a difference of .87.

The aggregate RT and ER effects across the different foreperiods in this meta-analysis will be contrasted with the effects observed in Posner et al (1973). This is because Posner et al.

was the main influence on Posner's theory of alerting (1975) and inspired many of the follow-up studies looking to assess alerting and its impact on response criterion and information processing (Los and Schut, 2008; McCormick et al., 2019; Han and Proctor, 2022; Klein, 2023).

Methods

Eligibility Criteria

For a study to be included in this meta-analysis, the following conditions had to be met:

- The task had to involve a condition that provided a warning signal, along with a condition in which no warning signal was provided. The 'warning signal' could not provide spatial information. The no-signal condition involves the same trial procedure as a signaled trial, but without the presence of a warning signal. The warning signal could be presented auditorily or visually.
- The design should be a 2-Alternative Forced Choice (2-AFC) task. This means the target can be one of two possible stimuli, or appear at one of two locations in space, and participants must make a speeded response by pressing one of two buttons. This allows for the measurement of Reaction Time (RT) and Error Rate (ER). Reaction time is the amount of time (in msec) required to react after target onset, and accuracy is the binary recording of whether they responded correctly (1) or incorrectly (0). Enough information must be available to extract necessary information for the meta-analysis, such as effect variance.
- The foreperiod between the warning signal and the target had to be 50 msec, 200msec, and/or 400 msec. If a foreperiod condition was within 100msec of these values, it was also included with the closest foreperiod. Deviations in the foreperiod were reported within the data table (Appendix A Table A2).

- The participants within these experiments should be typically developing, and without any reported or induced[1] impairments to cognitive ability.

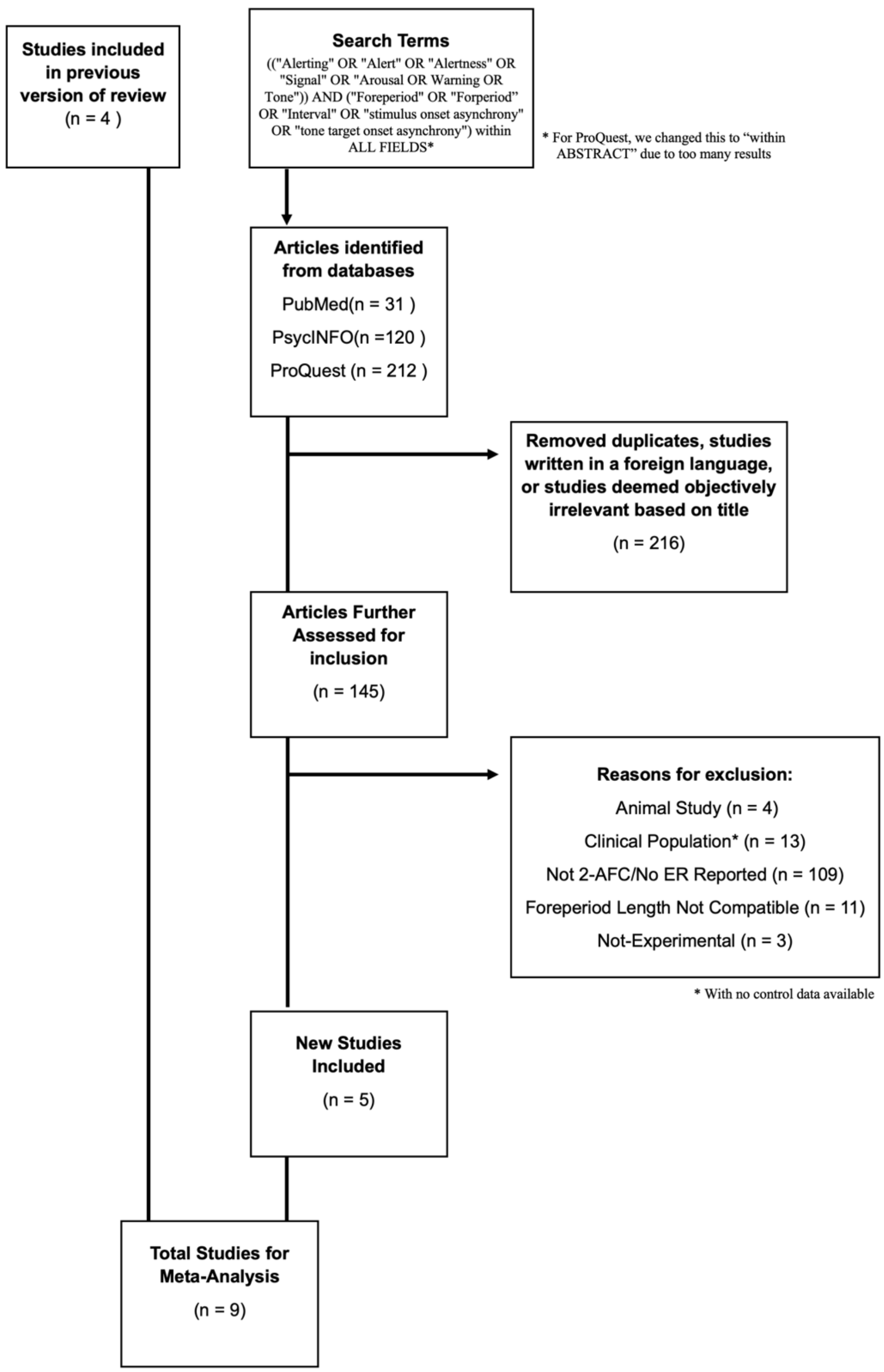
Information Sources & Search Strategy

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines have been followed in the reporting of how we conducted our review of the literature (see Figure 3). One of the authors of this paper (CM) was the sole member conducting the review and did not rely on automation tools. Four experiments were identified prior to the literature review (reported below). The review of the literature took place in October 2023.

The first step of the literature review was setting search terms for the three databases used. This resulted in a combined search-sum of 363. On each of the result pages provided by the database, CM chose to download papers that appeared to plausibly be related to the topic of interest. After removing duplicate results, this left 145 papers to be further inspected for methodological relevance. Of those 145 papers, five were determined to be applicable, in addition to the four previously identified, generating a total of nine experiments. The most common reason for exclusion was that the alerting study only reported RT or ER, but not both. A more detailed reporting of the review of the literature can be found in our flowchart (Figure 3). Considering the methodological restraints surrounding inclusion in this experiment, this is a reasonable total and will provide valuable insight into the likely range of effect sizes for reaction time and accuracy, allowing for a re-inspection of Posner's theory of alerting.

[1] experiments involving the administration of drugs

Figure 3. A flowchart showing the review of the literature that led us to including these nine papers.



Study Selection

The four papers identified prior to the review process were Posner, Klein, Summers, & Buggie (1973), Han & Proctor (2022), McCormick, Hurst, Redden, & Klein (2019), and Dietze & Poth (2022). Posner et al. was the motivating paper for this review, McCormick et al. and Han & Proctor were replications of this experiment, and Dietze & Poth was separately found via use in another research paper. The other five papers included are Han & Proctor (2023), Dietze, Recker, & Poth (2023), Dietze & Poth (2023), Kazen-Saad (an unpublished thesis work; 1983), and He et al. (2020). In total, these nine papers contribute sixteen different experiments.

The temporal distribution of studies is worth noting: two of the papers included are from the 1970s and 1980s, and then seven are from a four-year span from 2019 to 2023. There are a number of studies between this large temporal gap which were interested in the dynamics of alerting, however they often used either just RT or ER, or an alternative methodology (i.e. Simon tasks, Go No-Go, etc.). One of the reasons for this recent spike in research that matches our criteria is a renewed interest in the study of attention within the temporal domain, as it has been identified as a rich and understudied area of research (Nobre & van Ede, 2018; also see Nobre & van Ede, 2023). Additionally, the replication crisis and OpenScience movement inspired the direct replication of Posner's seminal study after 45 years (McCormick et al., 2019).

Data Collection Process

The meta-analysis was performed in R (R Core Team, 2023) using functions from the 'metafor' package (Viechtbauer, 2010). Relevant metrics were extracted from the raw data when possible[2], which included 11 of the 16 total experiments. Data were extracted from figures/supplementary tables otherwise. These metrics include the mean values for RT and ER,

[2] Either through online repositories or by contacting researchers directly requesting the data: McCormick et al., 2019; Han & Proctor 2022; 2023; Dietze, Recker, & Poth, 2023; Dietze & Poth, 2023; Dietze & Poth, 2022.

standard deviations, the sample size, trials per condition, along with t and F values. If information was not reported within the paper, either within the text or in supplementary materials, relevant equations were used (in the case of calculating effect standard deviation (SD), which was never reported), or estimates were generated based on other experiments (in the case of Kazen-Saad, 1983). If compatibility manipulations were included, as was the case in Posner, Klein, Summers, & Buggie (and the subsequent replications), only the compatible conditions were included.

Additionally, we identified and included four moderating variables: signal modality (was the signal auditory or visual), block structure (was the foreperiod fixed within a block vs intermixed), feedback (was RT feedback provided), and trials per condition (numeric). These moderators were included in the models once it was identified there was very high heterogeneity between the included experiments, despite their generally similar methodology. Signal modality was included as a moderator because the visual and auditory processing systems are distinct from one another and have been shown to differently impact performance (Posner, Nissen, & Klein, 1976; Dietze and Poth, 2023). Block structure was included as it determines whether participants could use the signal to predict when a target was presented or not, which impacts the influence of volitional preparation (endogenous mode) on performance. RT feedback was included as it was one of the features of Han and Proctors' replication of Posner et al. (2022) and was found to impact ER. If participants are provided feedback on performance in cognitive tasks, it will impact the response criterion they put forth and can lend to faster performance and less accurate performance (Hines, 1979). Trials per condition was used to account for the variance in mean estimates across the included experiments. Other demographic and design notes were recorded and were reported in the Appendix (Appendix A Table A1 and A2), but these four moderators were the only moderators tested on the model.

Data Items

Reaction time

Reaction time was the amount of time it took participants to respond to target stimuli within the task. Only the RTs for correct responses were included in the analysis. For experiments in which the data files were not available, RTs and their SDs were provided within tables or supplementary materials. For Posner et al. (1973) and Kazen-Saad (1983), RTs were extracted from tables, and condition SD was unavailable.

Error Rate

Error rate is the binary measure of accuracy (1=correct, 0=incorrect) averaged across available trials. For experiments in which the data files were not available, ER and its SDs were provided within tables or supplementary materials. For Posner et al. (1973) and Kazen-Saad (1983), ERs were extracted from tables, and condition SD was unavailable.

Log Odds

In addition, a transformation of ER into log odds values was conducted using the qlogis function in R (R Core Team, 2023). The effect SD for the log-odd values were based on the theoretical variance for values and can be found within the included R script. The theoretical variance was generated by running a high number of simulations with the logistic distribution, until the variance of the output stayed consistent to the second decimal place. This was required for the transformation into log odd values as we did not have enough information on individual participant performance in each condition from the included experiments, which is required for generating 95% confidence intervals (CI).

Summary Measures

Mean RT and ER effects are included for RT, ER, and Log Odds, which are difference scores between our mentioned comparisons of interest (no signal and signal conditions). Difference scores were compared for each signal-foreperiod condition a study included. Log-odds were generated by transforming ER values. These are presented in forest plots.

Effect Variance

Effect variance was not provided by any of the included papers, so it was calculated via the F-values from the ANOVAs for non-raw data extraction (using *effect variance* = $effect\ size/\sqrt{F} * \sqrt{n}$). This is except for Kazen-Saad (1983), in which no statistical test was reported for ER. In this case, we applied the same effect SD as Posner et al., 1973, as it was a conservative estimate and methodologically near identical. In the case of the raw-data analysis, we applied this equation to the data for each foreperiod difference score:

$$\sqrt{(variance(A)\ +\ variance(B)\ -\ 2 * (Covariance(AB))}$$

Risk of Bias in Individual Experiments

All the experiments included in this meta-analysis involve within-group comparisons, so comparisons between our different conditions are not impacted by sampling differences. Abnormalities within any of the experiments, including information on methodological abnormalities, statistical reporting abnormalities, data trimming, and any other possible concerns which could indicate a bias, were reported.

Risk of Bias Across Experiments

Risk of bias was assessed by presenting the included experiments in a funnel plot, which allows for a comparison of precision and effect size that can signal biases in the field. If the experiments are unbiased, they should be distributed symmetrically around the mean effect. Two of the experiments from the same paper use the same participants (Dietze & Poth, 2023; split into visual & auditory signals), but have no repeated data between them. Interpretations of the funnel plot and the possible sources of the biases were presented. Moderating variables were included to observe whether this controls for some of this bias.

Additionally, although there were 16 experiments total, six of these experiments included the authors 'Dietze and Poth', three were conducted by 'Han and Proctor', and Klein was an author on three of the papers. It is worth noting that these experiments likely share subtle design or sampling characteristics which could influence results (which could be speed/accuracy emphasis, geographical location. etc.), so we note any clustering of effects that may occur. We were able to find an unpublished dissertation involving two relevant experiments (Kazen-Saad, 1983) which have very similar methodological designs to Posner et al's experiment (1973). This adds some richness to our experiment sample, as we know academic journals have a bias towards publishing significant research outcomes.

Synthesis of Results

Three meta-analyses were run on RT data, three on ER data, and three on log-odd data, for a total of nine. This involved looking at the RT, ER, and log-odds effect of a warning signal with a 50 msec foreperiod before the target, 200msec before the target, and 400 msec before the target compared to when no signal was used. The effects for each condition involve difference scores of mean RT, ER when subtracting the 'no signal' condition mean from the 'signaled' condition. Each foreperiod condition had its own analyses, as the amount of time between the warning signal and target influences the size of behavioural effects (Posner et al., 1973; Lawrence and Klein, 2013) and involve different amount of influence from endogenous (volitional) and exogenous (reflexive) modes of alerting (Lawrence and Klein, 2013; Klein, 2022). The time-course of endogenous and exogenous influence is discussed at-length later when interpreting analysis outcomes. Moderating variables were included for signal modality (was the warning signal auditory or visual), block structure (were the foreperiods fixed or intermixed), and feedback (was RT feedback provided or not). The data from these analyses are presented in forest plots, which include calculated 95% CIs for each effect. A mixed-effects model was run through the rma.uni function (within the metafor package in R; Viechtbauer, 2010) with a restricted maximum likelihood estimator (REML) selected as the method. This is because the effects across experiments are not fixed and running the model as if they are fixed biases the model towards the experiments with very low overall ERs. This model also produces measures of heterogeneity (I^2 and T), which are reported and discussed.

The models reported represent the aggregate effect at the 'mean' of each of our dummy coded moderator variables. This accounts for differences generated by our three moderator variances by weighing their influence in proportion to the number of experiments which included the manipulation. This makes it so the reported effect best represents the overall effect for 2-AFC alerting experiments, and controls for more of the heterogeneity across experiments. This 'Meta-Regression' or 'adjusted' model, will be presented alongside the RE model, which does not include the moderating variables, within the forest plots.

Results

Study Selection

Sixteen experiments were included in this meta-analysis, pulled from nine different research papers. These experiments provided a methodologically homogeneous sample to allow us to explore how alerting impacts speed and accuracy performance across three foreperiod conditions.

Characteristics of the Experiments

All 16 experiments matched the inclusion criteria outlined in the methods section. They all used a 2-AFC task which compared signaled and no signaled trials across various foreperiods (50msec, 200msec, and 400msec).

Not all the included experiments contributed data to each foreperiod analysis. One-hundred and sixty-five participants were included in the 50 msec analysis (from six experiments), 298 for the 200 msec comparison (from 13 experiments), while four-hundred and four participants were included for the 400 mesc foreperiod comparison (from 11 experiments). Deviations in the exact length of foreperiod conditions is reported in the Appendix table (Appendix A Table A2). The participants were all typically developing, with an average/median age around the mid to low 20s, with one study having an experimental group with an average age of 75. The median number of trials per-condition cell was 60 (min: 15, max: 202, mean = 81).

Ten of the included experiments used auditory warning signals and six used a visual signal. For McCormick et al.'s (2019) experiment, there was an isointense manipulation that involved a shift from mono to stereo white noise in the task without a change in intensity (analogous to isoluminance in vision). This did not have any substantial non-additive effects within their study, so we have included it among the other experiments. Twelve of the experiments had consistent foreperiod durations within a block, while four intermixed the presentation of foreperiods throughout the experiment. Five experiments provided RT feedback, while 11 did not. These distinctions can be found in the notes of the Appendix table (Table A1), along with other relevant methodological distinctions.

Risk of Bias Across and Within Experiments

Moderators were included in the models to attempt to control for the unexplained heterogeneity across experiments. This included signal modality (was the signal auditory or visual), block structure (was the foreperiod fixed within a block vs intermixed), trial feedback (was RT feedback provided), and trials per condition (numeric). For the 50 msec foreperiod models, all the experiments use auditory signals, so there is no signal modality moderator. For the 200 msec signal and the 400 msec signal models, we have included all four moderators to observe their impact on heterogeneity. The influence that these moderators had on heterogeneity is evaluated with a likelihood ratio test (LRT), which compares the full model (all moderator levels included) to a restricted model with the moderator of interest dropped ($\alpha < .05$). The funnel plots, which were used to evaluate bias, used the full model with moderators; funnel plots generated without moderators included can be found in the Appendix (Appendix A Figure A1).

Evaluation of Bias for Error Rate Effects

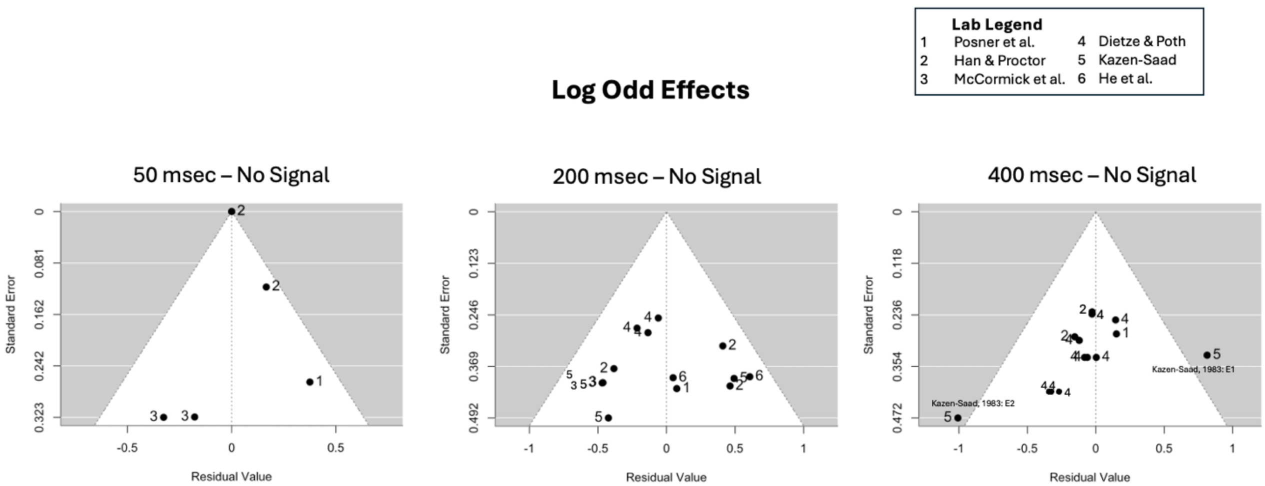
Figure 4. Funnel plots for error rate effects across the different foreperiod conditions. The x-axis represents the residual effect sizes. The line is the mean effect size across all experiments, the white triangle represents the likely values across the various standard error values (y-axis). In non-biased fields, experiments should be arranged symmetrically around the center line. Because many of these papers come from the same researchers, and we want to be able to detect possible ‘biases’ based on this, points are marked based on which author published the study. Experiments that generated unexpected values are labeled.

Within the 50msec foreperiod ER model, block structure (LRT = 5.58, $p = .02$, ΔI^2=39%) and feedback (LRT = 6.59, $p = .01$, ΔI^2=48%) had an influence in reducing heterogeneity, but trials per condition (LRT = 1.14, p =.29, ΔI^2=-24%) did not. In the full model, the confidence interval for residual heterogeneity included a range of values that provided little insight toward the level of heterogeneity across experiments (I^2=24%, CI [0, 98]; *Tau*=.002, CI [0, 7.10]). For the 200msec foreperiod ER model, feedback (LRT = 19.35, $p < .01$, ΔI^2= 80%) and trials per condition (LRT = 6.24, p =.01, ΔI^2= 34%) reduced the residual heterogeneity, while block structure (LRT = 0, $p = .82$, ΔI^2=0.00%) and signal modality (LRT = 3.86, $p = .05$, ΔI^2=8%) did not. In the full model, the confidence interval included a range of values that provided little insight toward the level of heterogeneity across experiments (I^2= 0%, CI [0, 99]; *Tau*=.002, CI [0, 2.16]). For the 400msec ER model, the trials per condition (LRT = 8.40, $p < .01$, ΔI^2=54%) reduced the overall heterogeneity. Block structure (LRT = 1.39, $p = .24$, ΔI^2=0.00%), feedback (LRT = 2.27, $p = .13$, ΔI^2=8%), and signal modality (LRT = .19, $p = .66$, ΔI^2=0.00%) did not have a significant influence in reducing heterogeneity. The full model shows that there is little certainty regarding level of heterogeneity, as the confidence interval covers a wide range of values (I^2= 0%, CI [0, 96]; *Tau*=0, CI [0, 2.64]).

Based on the funnel plots generated for ER models with moderators (Figure 4), there does not appear to be a bias towards larger or smaller values, as they are relatively equally distributed on the left and right side of the plot. Posner et al.'s ER has the largest effect size at the 200 msec and 400 msec foreperiod comparisons study, while also being an outlier in the amount of standard error generated, based on it falling outside of the expected values in the funnel plot (Figure 4). Han and Proctor (2023) have next to no standard error. This is because it is the only 'intermixed' block structure in the 50msec modes, so the residual error value is zero (this can be compared to the funnel plot generated by models without moderators in the Appendix A Figure A1).

Evaluation of Bias for Log-Odd Effects

Figure 5. Funnel plots for log-odd effects across the different foreperiod conditions. The x-axis represents the residual effect sizes. The line is the mean effect size across all experiments, the white triangle represents the likely values across the various standard error values (y-axis). In non-biased fields, experiments should be arranged symmetrically around the center line. Because many of these papers come from the same researchers, and we want to be able to detect possible 'biases' based on this, points are marked based on which author published the study. Experiments that generated unexpected values are labeled.



When comparing moderator influence on heterogeneity for the 50msec model, trials per condition (LRT = 5.44, p =.02, ΔI^2=2%) and feedback (LRT = 9.76, p < .01, ΔI^2=4%) had an influence in reducing residual heterogeneity, while block structure (LRT = .85, p = .36, ΔI^2=-1%) did not. There was evidence of high heterogeneity across experiments, as indicated by a confidence interval range which only includes larger values (I^2= 94%, CI [77, 99]; *Tau*=.37, CI [.17, 2.50]). Within the 200msec foreperiod log-odds model, feedback (LRT = 7.71, p <.01, ΔI^2=1%) had an influence in reducing heterogeneity, while block structure (LRT =.64, p = .42, ΔI^2=-.2%), signal modality (LRT =1.9, p = .17, ΔI^2=1%) and trials per condition (LRT = 1.3, p =.25, ΔI^2=.3%) did not. There was evidence of high heterogeneity across these different experiments (I^2= 97%, CI [93, 99]; *Tau*=.46, CI [.30, .90]). Within the 400msec foreperiod log-odds model, feedback had a small influence on heterogeneity (LRT = 5.5, p = .02, ΔI^2=.02%), but the other moderators did not influence unexplained heterogeneity between experiments (block structure (LRT =1.0, p = .31, ΔI^2=.1%); signal modality (LRT =.01, p = .9, ΔI^2=-.1%) and trials per condition (LRT = 3.5, p =.06, ΔI^2=0%). There are high levels of heterogeneity, as the confidence interval covers only high values (I^2= 99%, CI [99, 99]; *Tau*=.40, CI [.24, 1.15]).

When inspecting the log odd effects within the funnel plot (Figure 5, the effects are distributed evenly on either side of the effect distribution (Figure 5). Posner et al.'s effects are now within the expected outcomes. However, in the 400 msec model, Kazen-Saad's two experiments both have very small (E2) and large (E1) effect size relative to the other experiments, with most experiments being slightly clustered toward higher effect sizes. In their E2 experiment, the visual warning signal remained on until the target was presented. This may have helped with time-estimation, improving the participant's ability to accurately prepare for the target. Additionally, for both E1 and E2, this study had very high accuracy rates (conditions means were between 98.9% and 99.9% accuracy), so it makes sense that this effect ends up being distinct when converted to log-odds.

Evaluation of Bias for Reaction Time Effects

Figure 6. Funnel plots for reaction time effects across the different foreperiod conditions. The x-axis represents the residual effect sizes. The line is the mean effect size across all experiments, the white triangle represents the likely values across the various standard error values (y-axis). In non-biased fields, experiments should be arranged symmetrically around the center line. Because many of these papers come from the same researchers, and we want to be able to detect possible 'biases' based on this, points are marked based on which author published the study. Experiments that generated unexpected values are labeled.

When comparing moderator influence on heterogeneity within the 50msec RT model, trials per condition (LRT = 9.17, p <.01, ΔI^2=38%) had an influence on reducing residual heterogeneity, while feedback (LRT = 2.79, p = .10, ΔI^2=6%) and block structure (LRT = .13, p = .72, ΔI^2=-19%) did not. For the full model, the confidence interval included a full range of values that provided little insight toward the level of heterogeneity across experiments (I^2= 48%, CI [0, 90.3]; *Tau*= 5.38, 95 CI [0, 17.0]). For the 200msec model when comparing moderator influence on heterogeneity, feedback (LRT = 6.42, p =.01, ΔI^2=9%) and trials per condition (LRT = 6.47, p =.01, ΔI^2=8%) had an influence in reducing heterogeneity, while block structure (LRT =.001, p = .93, ΔI^2=-.8%) and signal modality (LRT =.30, p = .58, ΔI^2=-1%) did not. The full model indicates moderate to high heterogeneity (I^2= 77%, CI [47, 95]; *Tau*= 10.22, CI [5.29, 21.0]). For the 400msec model when comparing moderator influence on heterogeneity, trials per condition (LRT = 7.7, p <.01, ΔI^2=54%) had an influence in reducing heterogeneity, while block structure (LRT =.90, p = .34, ΔI^2=-9%), signal modality (LRT =.00, p = .98, ΔI^2=-9%), and feedback (LRT = 1.65, p =.20, ΔI^2=11%) did not. The current analysis indicates no certainty on the level of heterogeneity, as the confidence interval covers a wide range of values (I^2= 9%, CI [0, 86]; *Tau*= 1.3, CI [0, 10]).

For RT (Figure 6), there is a good balance of effect sizes across the different foreperiods (Figure 6). In the 200 msec foreperiod condition, He et al. (2020) falls outside the expected effect. This could possibly be due to only 30% of the trials in a block containing alerting signals, so they were more impactful in generating a reflexive alerting reaction in comparison to other experiments where alerting signals were more frequent/always provided within a block. Additionally, Han and Proctor had two of their experiments which had abnormally small (2022 [no feedback]; 2023 [intermixed foreperiods]), but this was due to them having a level of moderator which was not present across other experiments.

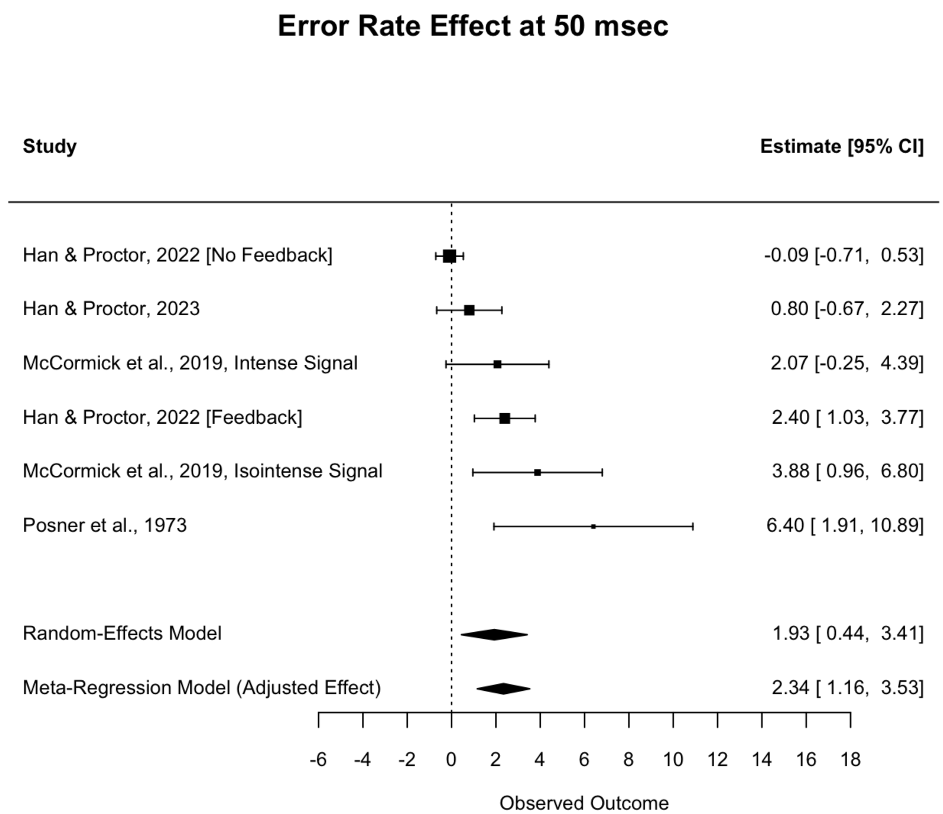
Synthesis of Model Results

Figures 7 to 15 contain the forest plots for ER, RT, and log-odd effects at each foreperiod. Both the Random-Effects model and the Meta-Regression (Adjusted) model are presented. The latter represents a weighted estimate of the effects based on included moderators, while the former is a model that does not involve the influence of moderators. The range of effects for each of these models is often quite similar, but in conditions where the included moderators explained residual heterogeneity, the adjusted model will have a tighter confidence interval. The reported values within text will represent the adjusted model values.

Error Rate

50 msec vs No Signal

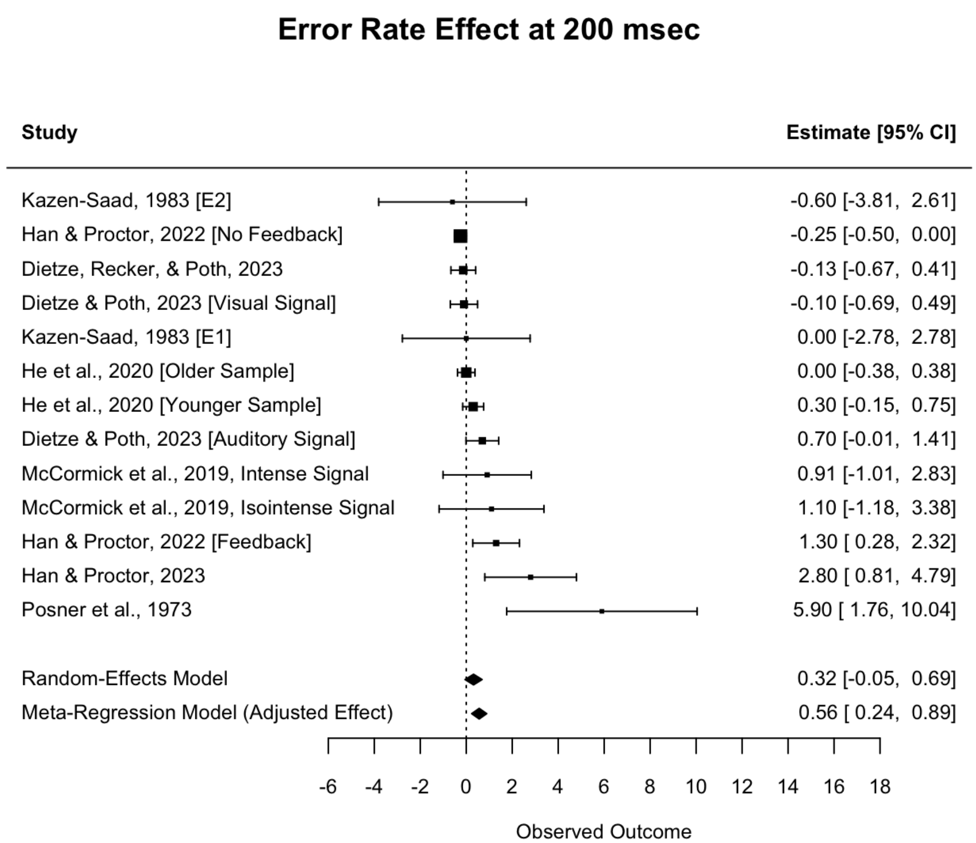
Figure 7. Forest plots for the error rate effect (difference in ER % between signaled trials vs no signal trials) at a foreperiod of 50 msecs. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these six experiments, participants made more mistakes on signaled trials compared to non-signaled trials (effect size = 2.34%, Standard Error (SE) = .61, 95% CI [1.16, 3.53]; see Figure 7). This is a small effect relative to the possible values originally reported in Posner et al. but was captured within the lower bound of their confidence interval. Experiments with RT feedback potentially generated slightly larger ER effects in comparison to experiments without RT feedback (estimate = 2.28%, SE = 1.4, 95% CI [-.48, 5.03]), and experiments with fixed foreperiod block structure had larger ER effects than experiments with intermixed foreperiod block structure (estimate = 2.56%, SE = 1.41, 95% CI [-.20, 5.31]), but the effect in both cases is potentially negligibly small.

200 msec vs No Signal

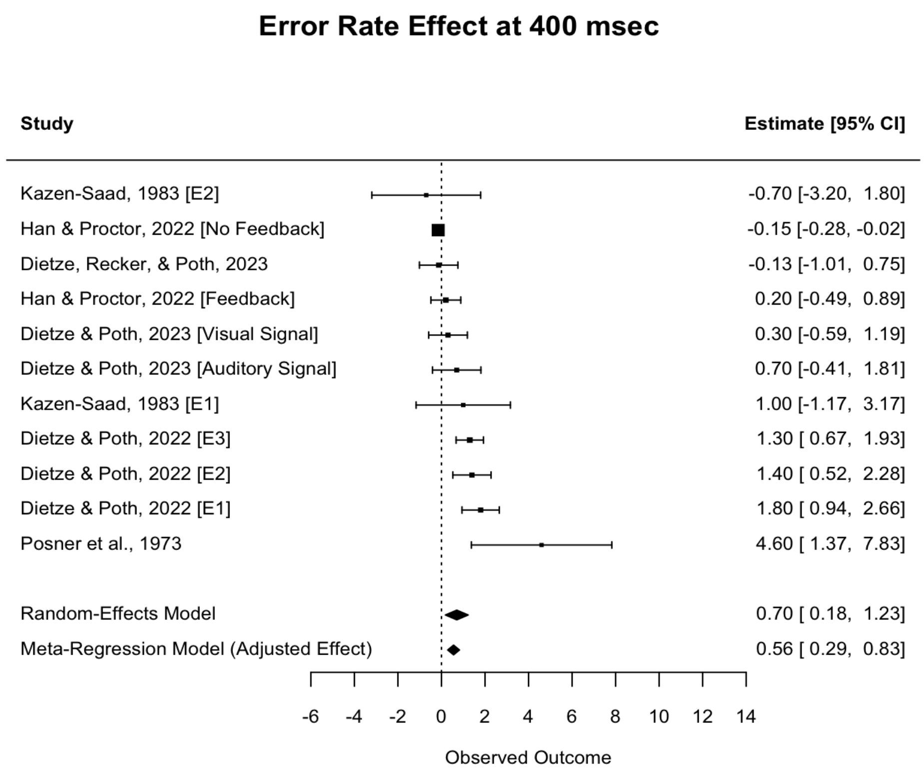
Figure 8. Forest plots for the error rate effect (difference in ER % between signaled trials vs no signal trials) at a foreperiod of 200 msecs. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these 13 experiments, participants made more mistakes on signaled trials compared to non-signaled trials (effect size = .56%, SE = .17, 95% CI [.24, .89]; see Figure 8). This range of small effect sizes falls outside the confidence interval from Posner et al., 1973. Experiments with RT feedback had larger ER effects in comparison to experiments without RT feedback (estimate = 1.75%, SE = .39, 95% CI [.97, 2.54]), and experiments with more trials per condition had very slightly larger effects, but likely not meaningful (estimate = .007, SE = .003, 95% CI [.00, .01]). Auditory signals also generated slightly larger error rate effects compared to visual signals (estimate = .78, SE = .40, 95% CI [.00, 1.57]).

400 msec vs No Signal

Figure 9. Forest plots for the error rate effect (difference in ER % between signaled trials vs no signal trials) at a foreperiod of 400 msecs. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these eleven experiments, participants made more mistakes on signaled trials compared to non-signaled trials (estimate = .56%, SE = .14, 95% CI [.29, .83]; see Figure 9). This range of small effect sizes falls outside Posner et al.'s original study's range. Increasing the trials per condition had a very small effect on ER effect size (estimate = .01%, SE = 0.003, 95% CI [0.00, 0.01]).

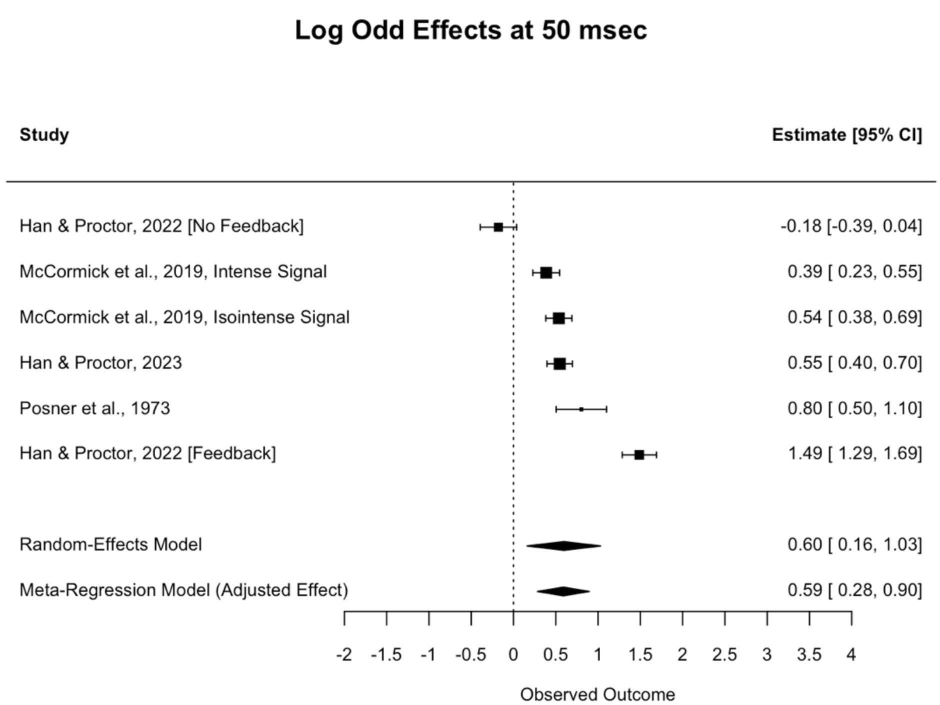
Comparison of ER Effects Across Foreperiods

When comparing ER effects across the different foreperiod conditions, the confidence interval generated by the 50msec model does not overlap with the range of values produced by the 200msec model or the 400msec model. This means that they likely represent different effect sizes, in which the alerting signal at a foreperiod of 50msec generates more error than when there is a 200 or 400msec foreperiod. The 200 and 400 msec effect sizes are nearly identical. Overall, the ER effects are very small, especially compared to Posner et al., 1973.

Log Odds Transformation of Error Rate

50 msec vs No Signal

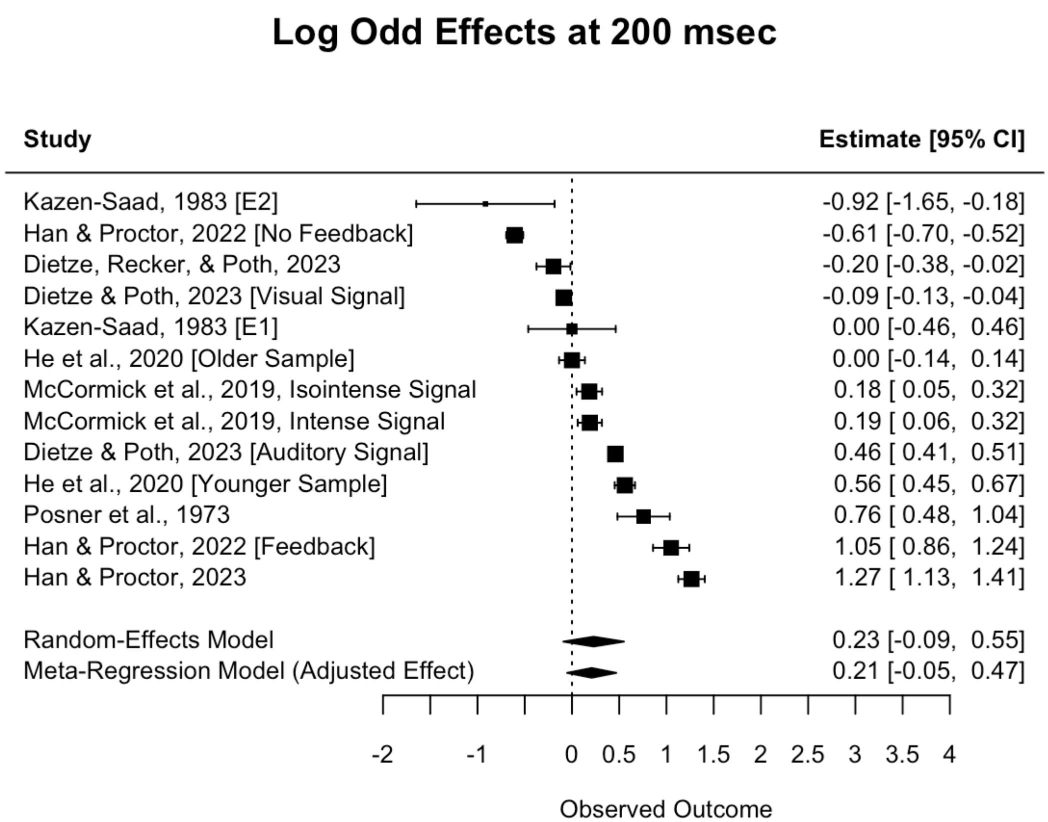
Figure 10. Forest plots for the log odd effect (difference in log-odds between signaled trials vs no signal trial) at a foreperiod of 50 msec. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within the included six experiments, participants made more mistakes on signaled trials compared to non-signaled trials (effect size = .59, SE = .2, 95% CI [.28, .90]; see Figure 10). The range of effect sizes is quite small. Experiments with RT feedback had larger log-odd effects in comparison to experiments without RT feedback (estimate = 1.51, SE = .50, 95% CI [.46, 2.54]).

200 msec vs No Signal

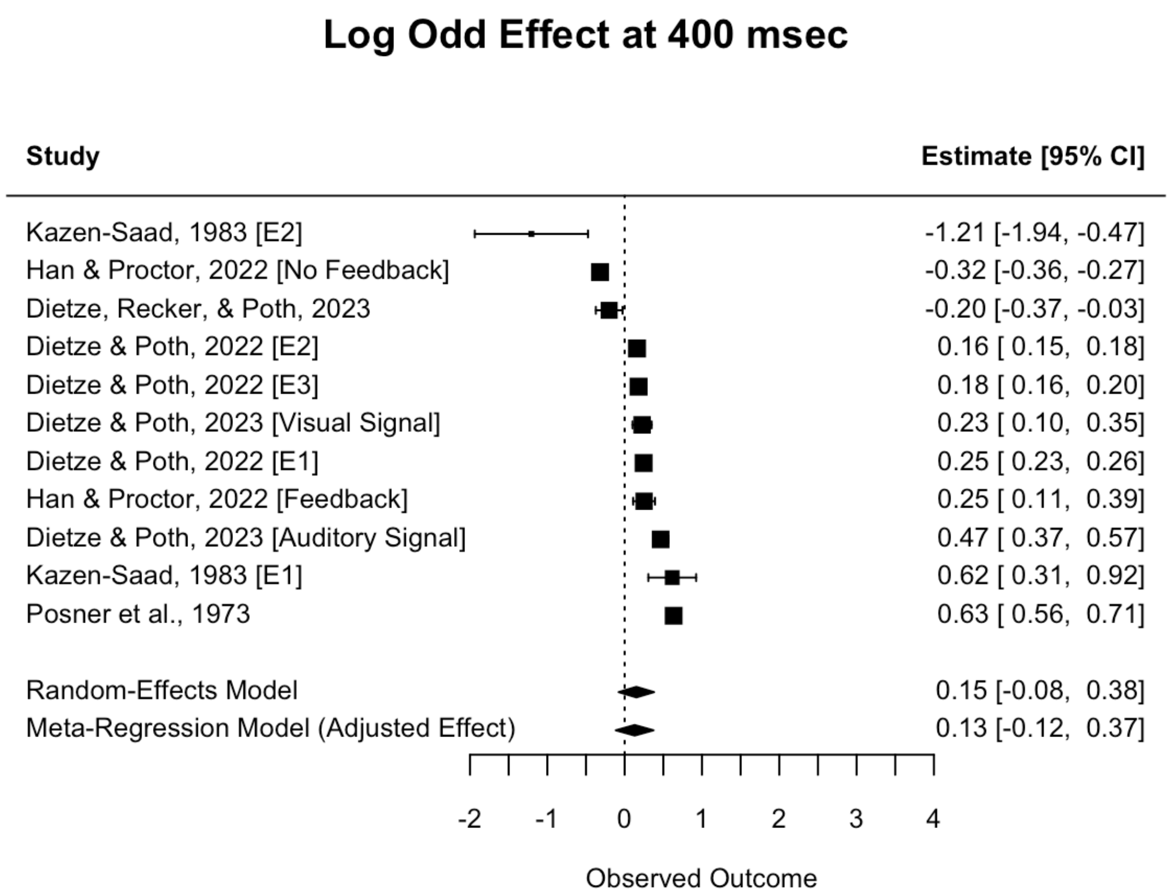
Figure 11. Forest plots for the log odd effect (difference in log-odds between signaled trials vs no signal trial) at a foreperiod of 200 msec. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these 13 experiments, a range of both positive and negative values were captured within the effect of signal on error performance (effect size = .21, SE = .13, 95% CI [-.05, .47]; see Figure 11). Most of the effects captured by the 95% confidence interval are positive values, however, they are small effects. If there is an effect of alerting on accuracy, it is challenging to declare that they are meaningful. Participants had higher log odd effects when provided with RT feedback than when not (0.81, SE = .32, 95% CI [.19, 1.43].

400 msec vs No Signal

Figure 12. Forest plots for the log odd effect (difference in log-odds between signaled trials vs no signal trial) at a foreperiod of 400 msec. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these eleven experiments, both positive and negative values were captured within the effect of signal on error performance (effect size = .13, SE = .12, 95% CI [-.12, .37]; see Figure 12). This indicates uncertainty regarding the presence of an effect of alerting on accuracy.

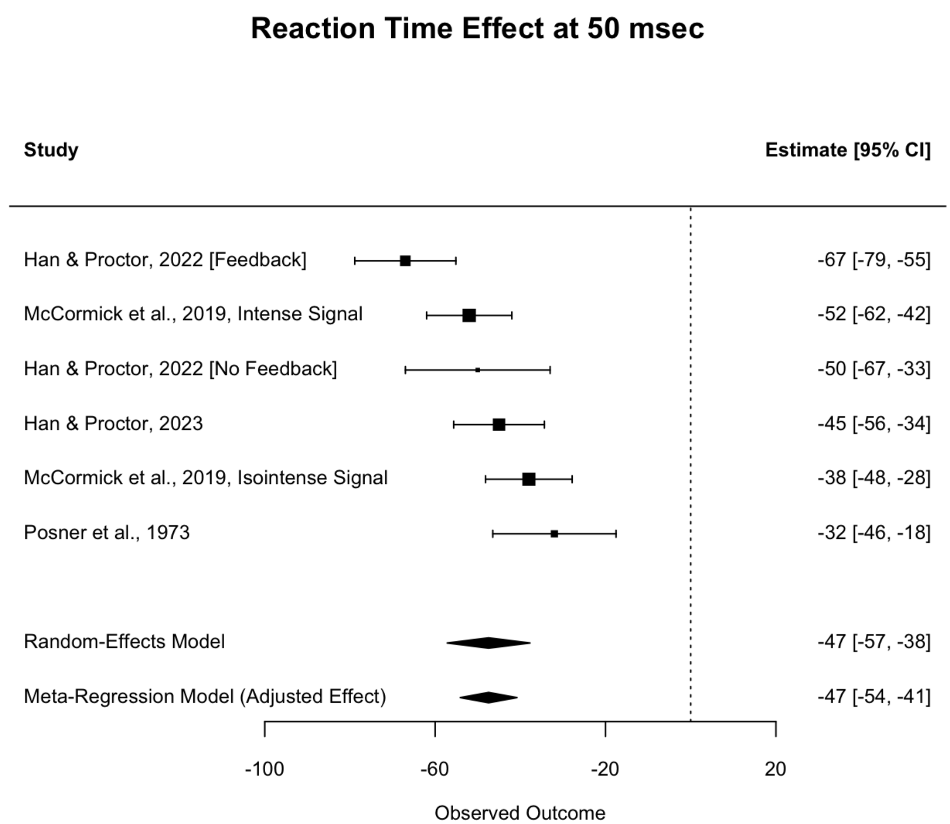
Comparison of Log-Odd Effects Across Foreperiods

When comparing log-odd effects across the different foreperiod conditions, the CI ranges from the adjusted random effect models all overlap with one another. This means that there is a possibility that the effect sizes could be the exact same value, so we cannot distinctly claim that there are differences in accuracy effects between foreperiods. This is distinct from the ER analysis, in which the 200 and 400 msec foreperiod models generated effects that were smaller than the 50msec condition. However, the 50 msec foreperiod condition is the only condition in which the confidence interval contains exclusively positive values. Even so, there is still a possibility of the effect being negligibly small. The transformation did change the magnitude of effects within some of the individual experiments, specifically those with very high (or lower) overall accuracy. However, there is still generally high heterogeneity across these experiments, signaling that there is inconsistency between experiments. The random effects model represents values that are either outside what was presented in Posner et al., 1973, or toward the lowest possible effect sizes.

Reaction Time

50 msec vs No Signal

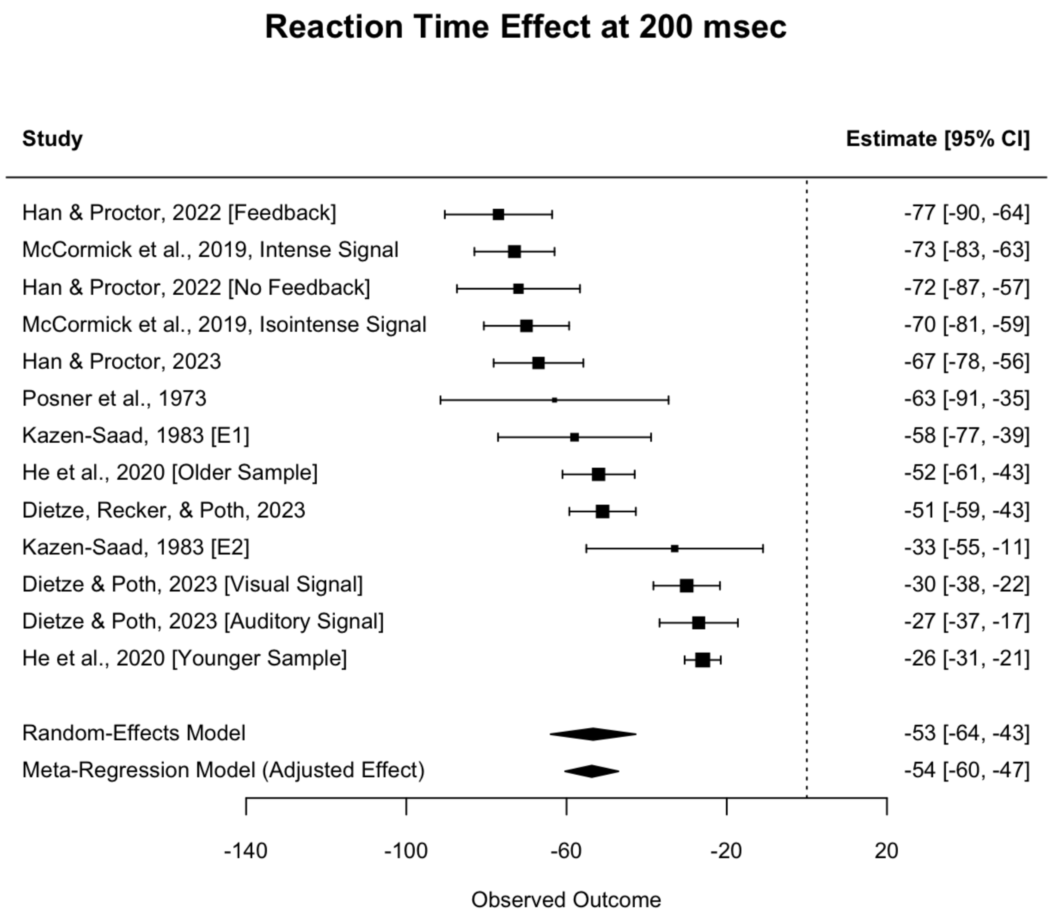
Figure 13. Forest plots for the RT effect (difference in RT between signaled trials vs no signal trial) at a foreperiod of 50 msec. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these six experiments, participants were faster on signaled trials compared to non-signaled trials (effect size = -47 msec, SE = 6.32, 95% CI [-60, -35]; see Figure 13). This is a fairly consistent effect across alerting experiments, and comparable to what was reported in Posner et al. 1973. As the number of trials per condition increased, so did the RT effect size, although by a modest amount (estimate = 1.36, SE = .46, 95% CI [.46, 2.26]).

200 msec vs No Signal

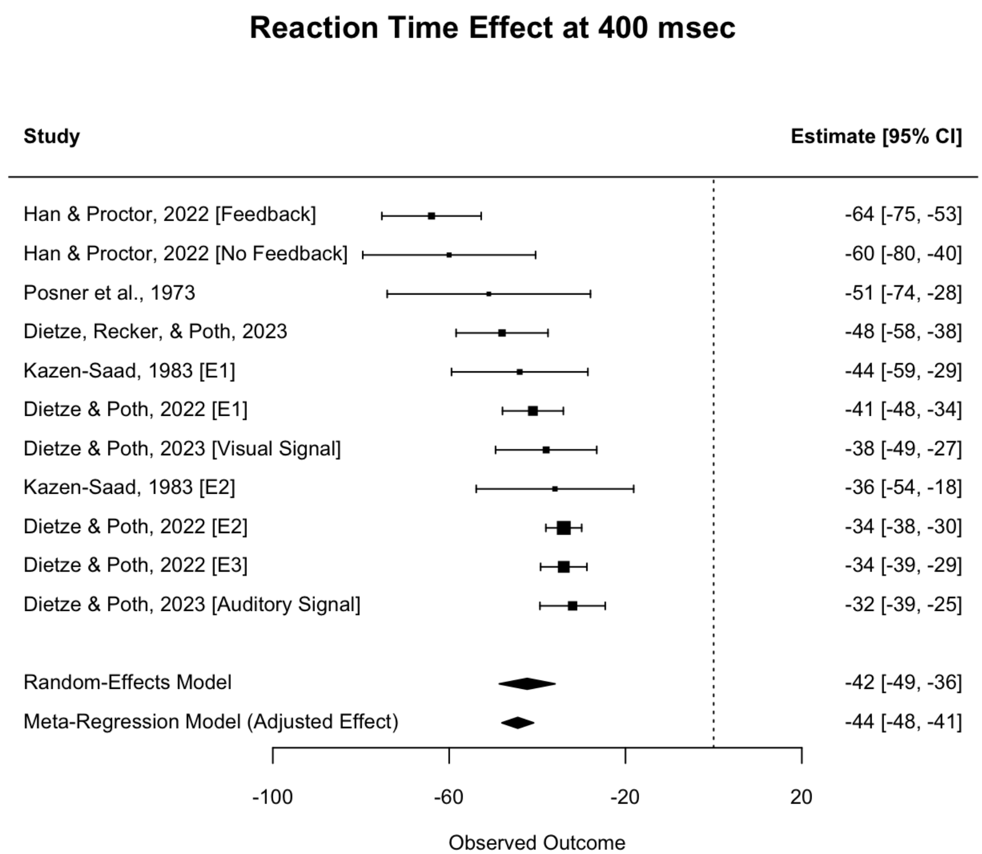
Figure 14. Forest plots for the RT effect (difference in RT between signaled trials vs no signal trial) at a foreperiod of 200 msec. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



Within these 13 experiments, participants were faster on signaled trials compared to non-signaled trials (effect size = -54 msec, SE = 4.1, 95% CI [-62, -46]; see Figure 14). This is a comparable effect size to Posner et al. 1973. Providing RT feedback made RT effects larger, with the effect size ranging between very small and quite large (estimate = 19, SE = 8, 95% CI [3, 35]). The number of trials per condition increased effect size by a small amount (estimate =.24, SE = .10, 95% CI [.03, .44]).

400 msec vs No Signal

Figure 15. Forest plots for the RT effect (difference in RT between signaled trials vs no signal trial) at a foreperiod of 400 msec. Error bars are 95% CIs, while the size of the mean dot represents the study's sample size. The mean and CI generated by a random effects model, and the adjusted effect accounting for moderating variables, are presented at the bottom of the figure.



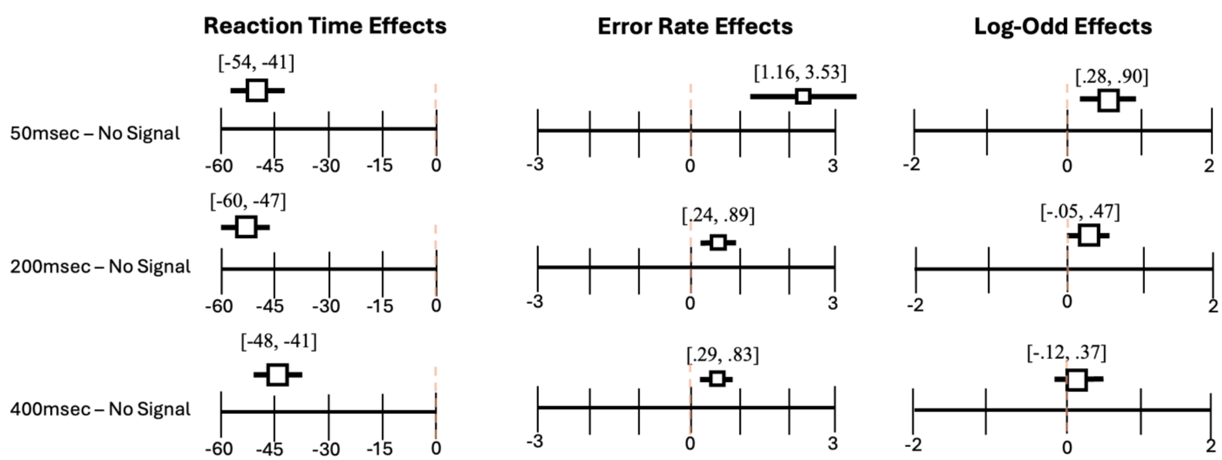
Within these eleven experiments, participants were faster on signaled trials compared to non-signaled trials (effect size = -43 msec, SE = 2.4, 95% CI [-48, -38]; see Figure 15). This is a tight range of possible effect sizes, and comparable to Posner et al., 1973. As the trials per condition increased, so did the RT effect size, by a modest amount (estimate =.09, SE = .03, 95% CI [.03, .15]).

Comparison of RT Effects Across Foreperiods

When contrasting the different RT effects for the foreperiod conditions, they all overlap with one another, so one cannot claim there are differences. This runs counter to the 'U-Shaped' function found in Posner et al. (1973) and, to a lesser degree, the subsequent replications (McCormick et al., 2019; Han & Proctor, 2022), in which 200 msec produced the fastest RTs, with 50 and 400 msec being relatively slower. The confidence intervals for 50 and 200 msec are comparable in width, and the confidence interval around the random effects model for 400 msec is tighter around a slightly smaller effect. The 200 msec foreperiod seems to have the most varying effect size across the 13 experiments.

Summary of Results

Figure 16. A summary of the adjusted effect forest plots for reaction time, error rate, and log-odd across the three foreperiod conditions. Values represent the likely effect sizes for reaction time, error rate, and the log-odd transformation, Error bars are 95% CIs.



RT alerting effects were generally quite large, and all comparable to Posner et al.'s original alerting experiment. When looking at alerting's impact on error rate, all three foreperiod conditions have quite small effect sizes. Additionally, the 200msec and 400msec conditions have a smaller, and equal, ER effect size compared to the 50 msec condition, indicating that participants maintained their speed-advantage for alerted trials while also slightly reducing the overall error rate. When converted to log-odds, which allows us to partially adjust for overall accuracy differences, we can see that the effect sizes are still small, and in the case of the 200 and 400 msec foreperiod, now contain negative values (see Figure 16). This means that alerting generated faster responses on these 2-AFC tasks, while having very little, if any, impact on accuracy.

Discussion

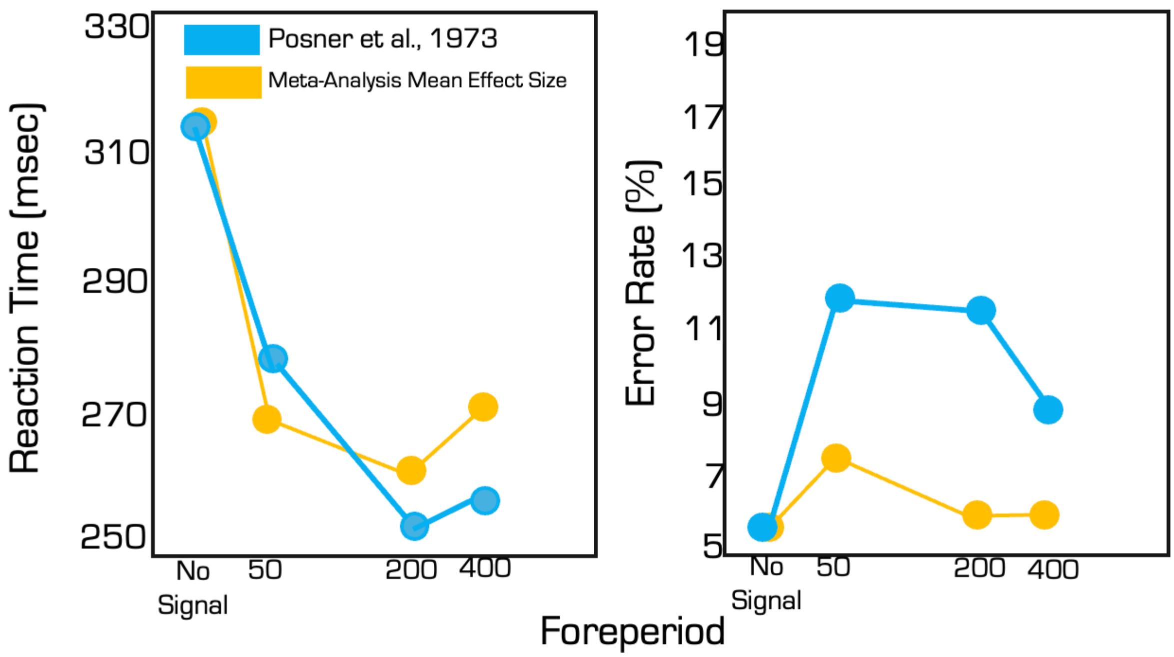
The current meta-analysis aimed to reassess the relationship between speed and accuracy across three foreperiods, using the logic from Posner et al. (1973). Posner's theory (1975), which was drawn from their results, states that alerting shifts an individual's response criterion, so participants are faster at the cost of accuracy[3] without improving the speed at which information is processed. The evidence generated by this meta-analysis will be interpreted in the context of Posner's theory, with consideration to how variations in methodology may have impacted performance outcomes. Specific emphasis is put on contrasting Posner et al's (1973) effects with the confidence intervals generated by the models.

---

[3] Or vice versa. However, all experiments included found the former to be true.

Magnitude of Reaction Time and Error Rate effects in Comparison to Posner et al.

Figure 17. Condition means from Posner et al. compared to the results of this meta-analysis. For the meta-analysis, the effect size for each foreperiod condition was subtracted/added to the RT/ER of Posner et al.'s 'no signal' condition. Only compatible response mapped trials were included.



Although the RT effects were generally large and comparable to Posner et al., it appears that the ER effects in each foreperiod condition had much smaller estimates (see Figure 17). This is especially true for the 200 and 400 msec foreperiod conditions, where the confidence interval for the ER model is completely below the smallest possible values reported in Posner's wide confidence interval (see Figures 2.8 & 2.9). This suggests that the ER effects of alerting at each foreperiod are small, even in the context of large reaction time effects. When considering the log-odds transformation of the error data, which accounts for the overall differences in accuracy between the experiments, there is limited evidence for the presence of an effect of alerting on accuracy. In addition, the confidence interval for 200 and 400 msec conditions contain possible negative effects. The addition of log-odd transformations is helpful in the interpretation of accuracy effects across experiments. Consider the two direct replications recently conducted of Posner et al., and how it changes the interpretation: Posner et al. compared values of 11.6% (warning signal trials) and 5.2% (no warning signal trials), whereas Han and Proctor compared 3.1% and .7%, and McCormick et al. compared 6.5% to 4.4%, respectively. McCormick et al. declared a successful replication within their paper because their ANOVA foreperiod effect on ER was significant, although the relative log-odd effect size we calculated here was half the size of Posner's at 50 msec (.4 vs .9). Meanwhile Han and Proctor declared a lack of a speed-accuracy trade off because their ANOVA foreperiod effect on ER was non-significant, even though the log-odds effect size was much larger than the original study (1.5 vs .9).

Posner et al.'s (1973) ER effect sizes are anomalously large in comparison to the other alerting experiments, which can be observed in the funnel plots (Figure 4). It is not surprising that Posner concluded substantial criterion shifts across the different foreperiod conditions based on this outcome (1973; 1975). One important factor that may have influenced the increased magnitude of errors in Posner et al.'s task was the compatible and incompatible response mappings implemented across blocks. This feature of the task was used to increase the overall error rate within the alerting task, because a prior alerting task failed to observe a significant effect on ER due to high overall accuracy (Posner & Boies, 1971). However, we now know there is an interaction between alerting and response compatibility, in which alerting can activate the direct path between matching stimulus and response features (De Jong et al., 1994). This likely puts additional cognitive load on signaled trials in comparison to non-signaled trials, which means this task manipulation did not evenly apply this increase in difficulty among all the conditions. With this said, the included experiments from McCormick et al. (2019) and Han and

Proctor (2022; 2023) also had response mapping manipulations, and we did not observe anomalous effect sizes. In addition, we only included the 'compatible mapping' trials from these experiments, although one could imagine carry-over cognitive load and task-switching effects within these trials.

In summary, RT effects are consistently large across the three different foreperiods, while the effect of alerting on accuracy, evaluated through log-odd transformations, shows very small, if present at all, accuracy effects. This outcome is a significant deviation from the results of Posner et al., 1973, in which the faster a participant is, the more errors they generate. This novel pattern of results needs to be contrasted using the same logic previously used by Posner (1973; 1975).

Relationship Between Speed and Accuracy: Evidence for a Criterion Shift or Improved Information Processing?

Under Posner's theory of alerting (1975), alerting generates faster responses without any increase in processing speed, resulting in less information being available to the participant at the time of response. This means that when RTs are reduced via a warning signal, ERs should increase. However, across the three foreperiod conditions in which substantial reaction time effects were observed, the effect on accuracy was either quite small, or non-existent, as indicated by both ERs and the transformed log-odd values. Additionally, when we specifically contrasted RTs and ERs between our three foreperiod conditions, we observed a consistent RT effect, but that ER effect at the 200 or 400msec foreperiods is smaller than the 50 msec condition. The maintenance of the RT benefit with an improving ER supports the premise that alerting has some enhancement to the efficiency of information processing, and these effects are not just exclusively shifting the response criterion to trade-off speed for accuracy. It is worth noting that even though Posner et al (1973) originally described the alerting effect as a criterion shift, this does not fully explain the pattern of data that they themselves observe. Within their data, RT is faster in the 200msec foreperiod condition in comparison to the 50 msec foreperiod condition, and instead of this resulting in a higher ER for the 200msec condition, it is numerically reduced. This improvement to both aspects of performance indicates improvements in information processing at this signal-target interval.

With this said, it is possible that alerting generates a criterion shift in which participants trade-off accuracy at the betterment of speed, in addition to increasing the efficiency of information processing. In our meta-analysis, the highest error rates appear to be at a foreperiod of 50 msec. Han and Proctor (2022), in their replication of Posner et al. (1973), indicate that a shift in criterion may only occur at the shortest foreperiod durations due to the increased influence of automatic alerting mechanisms. Lawrence and Klein (2013) isolated the influence of voluntary (endogenous) and automatic (exogenous) modes of temporal attention (a more general term for alerting). Endogenous temporal attention involves voluntary preparation for an upcoming stimulus based on learned (or cued) temporal distributions, whereas exogenous temporal attention is an automatic response to salient stimuli, like a warning signal. The two modes of temporal attention have distinct time-courses, and therefore have varying levels of influence on performance depending on foreperiod condition. Peak exogenous alerting occurs around 50 to 80 msec after a warning stimulus (Denison, Carrasco, and Heeger, 2021), while peak endogenous alerting occurs around 400msec (McCormick, Redden, and Klein, 2023). Lawrence and Klein propose that the combination of endogenous and exogenous temporal attention, around very short foreperiod durations between 50 and 100 msec, generates SATs (2013; page 568). Together, this evidence suggests that the peak of the exogenous mode of temporal attention is driving a criterion shift, and the endogenous mode is driving a criterion shift of speed for more accuracy. More research is required to better separate the dynamic influence of endogenous and exogenous modes at various foreperiod durations.

The consideration of how endogenous and exogenous modes of temporal attention differently impact performance is salient in relation to Klein's recent replication (2023) of Posner's et al.'s pattern through a reanalysis of Los and Schut (2008). In this reanalysis, the four foreperiods (50, 100, 350, and 400 msec) with the slowest RTs, and the four foreperiods (150, 200, 250, and 300 msec) with the fastest RTs were combined to compare the combined mean ERs. Klein found that the average ER of the fastest SOAs was larger than the average ER of the slowest, replicating the SAT outcome of the seminal study by Posner et al. (1973). However, based on what we know about the two modes of alerting, the errors occurring at the 50 msec foreperiod have a distinct source in contrast with errors occurring at the 400 msec foreperiod. With this said, Klein indicates that the outcome of the reanalysis mirrors the overall pattern of performance from Posner et al. (1973), but this does not necessarily mean that these changes in performance are being generated exclusively by a shift in response criterion.

Moderating Factors

This experiment included either direct replications of Posner et al. (in the case of McCormick et al, 2019; Han & Proctor, 2022) or near-direct replications, so it was surprising to obtain such high heterogeneity measures across our various effects. There are a relatively small number of experiments within some of our meta-analyses, so an outlier study can have a stronger bias on the amount of heterogeneity calculated (the I^2 value). Papers that contributed multiple experiments also impacted the heterogeneity, as the size of effects from experiments in the same study tended to cluster together even when using different sets of participants, likely due to minor experimenter-specific distinctions in methodology or sampling. Three of our moderators- trials per condition, RT feedback, and block structure- did account for some of the heterogeneity. Trials per condition affects the amount of variance between experiments, since studies that have more trials for each participant will have a more accurate representation of their average performance, so this is a rather straightforward moderating factor. The other moderators, along with other considerations, are discussed below.

RT Feedback

In the 50 msec condition, RT feedback impacted ER effect size, but did not impact RT. However, only one out of the six experiments did not provide feedback, so we exercise caution in generalizing this comparison. However, at 200 msecs, where there is a more balanced representation of experiments providing RT feedback or not, there was a moderating effect of feedback on RT, ER, and Log-Odd effects in the expected direction: larger ER and RT effects (Hines, 1979). This indicates that RT feedback shifts the response criterion at a cost to accuracy. However, in the 400 msec condition, there was no difference in speed or accuracy performance based on RT feedback. It is possible that sufficient volitional preparation after a signal can help control for errors when increasing speed. This is the duration that is most commonly used in temporal cueing experiments, as a measure of endogenous temporal attention, and these experiments often display overall improvements in performance (as opposed to a trade-off between speed and accuracy; Nobre & van Ede, 2018). Han and Proctor (2022) explicitly manipulated RT feedback across their two included experiments, and showed that the warning signal condition improved both speed and accuracy in comparison to the no signal condition when no RT feedback was provided (in contrast with the speed-accuracy trade-off observed when it was). These two experiments both produced relatively larger, and comparable RT effects. It is still puzzling why participants were still able to maintain similar RT effects when they did not have feedback, while also reducing ER. Experiments from Dietze and Poth (2022; 2023; also Dietze, Recker, and Poth, 2023), which did not provide RT feedback, had overall smaller RT effects than other experiments included in this analysis.

Block Structure

The only model that detected a difference in ER performance for block structure levels was the 50 msec foreperiod analysis. It is unwise to generalize this significant difference, since there was only one study with an 'intermixed' condition, so this difference could have been influenced by other methodological factors from that one study. There was otherwise no impact of having predictable foreperiod durations vs intermixed foreperiods within a block. This is surprising, as we would anticipate that experiments using a fixed foreperiod design would generally have lower ER effects (and possibly faster RTs), considering that the consistent duration allows participants to volitionally prepare for upcoming targets. In the case of Han and Proctor, which intermixed foreperiods within a block, it is not surprising that varying the foreperiod did not impact this component of preparation, considering the range of the foreperiods was less than 200msec. These foreperiods are similar enough that participants could maintain an increased state of vigilance once hearing a warning signal. Additionally, 200 msec is the shortest duration that participants can volitionally prepare for in a 2-AFC task (McCormick, Redden, & Klein, 2023). However, in the case of Dietze and Poth (2023), in which a non-aging distribution of foreperiods was used (ranging from 100 to 1500 msec foreperiods), this represents a purer version of the reflexive alerting mode. This is because there is much more uncertainty on the timing between a signal and a target, so participants cannot rely on volitional preparation, resulting in performance differences coming from a participants reflexive reaction to the salient warning stimulus. Based on Lawrence and Klein's findings (2013), we would expect to observe performance contrasts between this study by Dietze and Poth (2023) and the other experiments that involved components of volitional preparation.

Signal Modality

Although past research has indicated distinctions in the time-course of alerting via auditory and visual warning signals (Bertelson, & Tisseyre, 1969; see introduction of Dietze & Poth, 2023 for a review), an effect was only observed in the 200 msec foreperiod for error rate, and it was rather small. In addition, there was no effect on speed, or when ER was transformed to log-odd values. One of the experiments included in the analysis, Dietze and Poth, 2023, did make an explicit comparison of visual and auditory alerting signals across a variety of foreperiods, and concluded that in general, both stimulus types are relatively equal in the alerting response they generate.

Other Considerations

There are other possible distinctions in methodology that could impact performance across experiments. While we did control for the difference in overall ER by transforming values to log-odd, which did impact interpretations, this was not controlled for within RT analysis. It is worth considering that the overall speed could impact the size of these effects across the different foreperiod conditions. However, for our current meta-analysis, involving RT transformations would have added additional complexity and affected the overall power available. Instead, this is something that may be explored empirically in future experimental paradigms. Additionally, we saw some clustering of performance effects based on the lab in which a study was conducted. While this clustering could be related to deviations in methods- such as how experimenters emphasize speed or accuracy within the task- it is possible this is also due to sampling differences. A focus for future alerting research should be to expand beyond WEIRD[4] samples to obtain a more representative picture of how alerting impacts information processing in the average human (Henrich, Heine, & Norenzayan, 2010. However, any slight clustering of effects based on which author was involved was never overly concerning, as the magnitude of influence over the effect was within reason.

Limitations

While the homogeneity of methodological procedures was an asset for this meta-analysis in generating meaningful estimates of effects, as it ensured our comparisons of mean values were more directly comparable, such homogeneity could also be considered a limitation for generalizability. Based on the outcome of this analysis, we now better understand how alerting operates under these very stringent conditions, and we can update past theories on how alerting impacts information processing. However, we are limited in understanding how the alerting process is impacted by many other mediating mechanisms which are often involved when we elicit alerting to interact within dynamic and noisy tasks in our daily lives. Research indicates that there are a number of reflexive and volitional mechanisms that distinctly impact performance outcomes depending on task demands, and that often operate in tandem (Nobre & van Ede, 2023). Future meta-analyses should consider analyzing the effect of alerting on information processing across distinct methodologies.

Conclusion

Based on the outcome of the current meta-analysis, it appears that alerting increases the efficiency of information processing to some degree, contrary to Posner's theory (1975) that it solely impacts the response criterion. Large RT effects were generated by alerting signals, but

[4] Western, Educated, Industrialized, Rich, and Democratic

there was limited evidence of a meaningful impact on ER, especially in the longer 200 and 400msec foreperiod conditions. The behavioural effects occurring at the 50 msec foreperiod are likely more influenced by exogenous alerting mechanisms than the later 200 and 400 msec foreperiods, which rely more on endogenous alerting mechanisms. To our knowledge, this is the first meta-analysis to be conducted on this topic in the 50 years since the publication of Posner et al. (1973), and our divergent result emphasizes the importance of revising theories as new evidence is generated. Future research should explore how endogenous and exogenous modes of alerting differently impact performance at different foreperiods, determine the degree to which there may be changes in response criteria, and ascertain how this is influenced.

Declarations

Funding

This project was supported by a Nova Scotia Graduate Scholarship award.

Conflicts of Interest

There are no conflicts of interest to report.

Ethics Approval

Not applicable.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Availability of Data and Materials

Data and analysis materials are available at the link provided in the open practices statement.

Code Availability

Not applicable.

References

Bertelson, P., & Tisseyre, F. (1969). The time-course of preparation: Confirmatory results with visual and auditory warning signals. Acta Psychologica, 30, 145-154.

Los, S. A., & Schut, M. L. (2008). The Effective Time Course Of Preparation. Cognitive Psychology, 57(1), 20-55.

De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. Journal of Experimental Psychology: Human Perception and Performance, 20, 731–750

Denison, R. N., Carrasco, M., & Heeger, D. J. (2021). A dynamic normalization model of temporal attention. *Nature Human Behaviour*, *5*(12), 1674–1685. https://doi.org/10.1038/s41562-021-01129-1

Dietze, N., & Poth, C. H. (2022). Phasic Alertness is Unaffected by the Attentional Set for Orienting. Journal of Cognition, 5(1).

Dietze, N., Recker, L., & Poth, C. H. (2023). Warning signals only support the first action in a sequence. *Cognitive Research: Principles and Implications*, *8*(1), 29. https://doi.org/10.1186/s41235-023-00484-z

Dietze, N., & Poth, C. H. (2022). Phasic Alertness is Unaffected by the Attentional Set for Orienting. *Journal of Cognition*, *5*(1), 46. https://doi.org/10.5334/joc.242

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*(4), 447-456.

Han, T., & Proctor, R. W. (2022). Effects of a neutral warning signal on spatial two-choice reactions. Quarterly Journal of Experimental Psychology, 75(4), 754-764.

Han, T., & Proctor, R. W. (2023). Effects of a neutral warning signal under increased temporal uncertainty. *Memory & Cognition*, *51*(6), 1346–1357. https://doi.org/10.3758/s13421-023-01404-8

He, M., Heindel, W. C., Nassar, M. R., Siefert, E. M., & Festa, E. K. (2020). Age-related changes in the functional integrity of the phasic alerting system: A pupillometric investigation. *Neurobiology of Aging*, *91*, 136–147. https://doi.org/10.1016/j.neurobiolaging.2020.02.025

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of memory and language, 59(4), 434-446.

Kazen-Saad. (1983). The Interaction of Phasic Alertness and Pathway Activation: Relationships to Specific Task Preparation*, Dissertation, The Pennsylvania State University, University Park.*

Klein, R. M. (2023). Assessing the predictions from Posner's theory of phasic alertness using data from Los and Schut (2008). *Memory & Cognition*. https://doi.org/10.3758/s13421-023-01438-y

Lawrence, M. A., & Klein, R. M. (2013). Isolating exogenous and endogenous modes of temporal attention. Journal of Experimental Psychology: General, 142(2), 560.

McCormick, C. R., Redden, R. S., & Klein, R. M. (2023). How does exogenous alerting impact endogenous preparation on a temporal cueing task. *Visual Cognition*, *31*(5), 363–379. https://doi.org/10.1080/13506285.2023.2250530

McCormick, C. R., Redden, R. S., Hurst, A. J., & Klein, R. M. (2019). On the selection of endogenous and exogenous signals. Royal Society open science, 6(11), 190134.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. Psychological bulletin, 89(1), 133.

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34–48. https://doi.org/10.1038/nrn.2017.141

Nobre, A. C., & Van Ede, F. (2023). Attention in flux. *Neuron*, *111*(7), 971–986. https://doi.org/10.1016/j.neuron.2023.02.032

Posner, M. I. (1975). Psychobiology of attention. *Handbook of psychobiology*, 441-480.

Posner, M. I. (2008). Measuring alertness. *Annals of the New York Academy of Sciences*, *1129*(1), 193-199.

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, *78*(5), 391–408. https://doi.org/10.1037/h0031333

Posner, M. I., Klein, R., Summers, J., & Buggie, S. (1973). On the selection of signals. Memory & Cognition, 1(1), 2–12. https://doi.org/10.3758/BF03198062

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. Psychological review, 83(2), 157.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25-42.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual review of neuroscience*, *35*, 73-89.

Xuan, B., Mackie, M. A., Spagna, A., Wu, T., Tian, Y., Hof, P. R., & Fan, J. (2016). The activation of interactive attentional networks. *Neuroimage*, *129*, 308-3

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48. https://doi.org/10.18637/jss.v036.i03

Appendix

Appendix Table 1. Experiment design features

Study Name	Data extracted from	N	Median Age	Total Trials	TPC	Signal Modality	Tone (dB)	Block Structure	RT Feedback
Posner et al, 1973	Paper	9	NR	360	40	Auditory	NR	Constant	YES
Han & Proctor, 2022: No Feedback	Raw	36	NR	180	15	Auditory	80dB	Constant	NO
Han & Proctor, 2022: Feedback	Raw	36	NR	180	15	Auditory	80dB	Constant	YES
McCormick et al, 2019: Intense	Raw	24	20.2 (17 - 30)	256	32	Auditory	NR	Constant	YES
McCormick et al, 2019: Isointense	Raw	24	20.2 (17 - 30)	256	32	Auditory	NR	Constant	YES
Dietze, Recker, & Poth, 2023: E2	Raw	34	25 (19-41) & 23 (18-30)	210	23	Auditory	70dB	Intermixed	NO
Dietze & Poth, 2023: Visual Signal	Raw	7	27 (21-41)	52500	145	Visual	36dB-64dB	Intermixed	NO
Dietze & Poth, 2023: Auditory Signal	Raw	7	27 (21 - 41)	52500	152	Auditory	NA	Intermixed	NO
Kazen-Saad, 1983: E1	Paper	20	NR	180	60	Visual	NA	Constant	NO
Kazen-Saad, 1983: E2	Paper	15	NR	180	60	Visual	NA	Constant	NO
Dietze & Poth, 2022: E1	Raw	41	24(19-42)	404	202	Visual	NA	Constant	NO
Dietze & Poth, 2022: E2	Raw	42	25.5(16-60)	404	202	Visual	NA	Constant	NO
Dietze & Poth, 2022: E3	Raw	71	23(18-35)	404	202	Visual	NA	Constant	NO
He et al, 2020: Old Sample	Table	25	75	202	60	Auditory	53dB	Constant	NO
He et al, 2020: Young Sample	Table	25	20	202	60	Auditory	53dB	Constant	NO
Han & Proctor, 2023: E1	Raw	36	NR	240	30	Auditory	80dB	Intermixed	YES

Study Name	Data Extraction Notes
Posner et al, 1973	Means are extracted from Figure 3, and the effect SD calculated from the reported F value. Only included compatible trials, which is accounted for in total trials. Mean RTs from Table 1.
Han & Proctor, 2022: No Feedback	Data extracted from table included in appendix. Only compatible condition included, which is accounted for in total trials. 80dBa
Han & Proctor, 2022: Feedback	Data extracted from table included in appendix. Only compatible condition included, which is accounted for in total trials. 80dBa
McCormick et al, 2019: Intense	Mean and SD pulled from raw data (available on OSF). Only compatible conditions included, which is accounted for in total trials
McCormick et al, 2019: Isointense	Mean and SD pulled from raw data (available on OSF). Only compatible conditions included, which is accounted for in total trials
Dietze, Recker, & Poth, 2023: E2	Means and SD were extracted from the data.
Dietze & Poth, 2023: Visual Signal	Means and SDs were extracted from data. The no alerting condition was randomly split for each participant to make independent comparisons between visual and alerting, so there was no data overlap.
Dietze & Poth, 2023: Auditory Signal	Means and SDs were extracted from data. The no alerting condition was randomly split for each participant to make independent comparisons between visual and alerting, so there was no data overlap.
Kazen-Saad, 1983: E1	Effect SD for ER & RT was made equal to PKSB. This was a conservative choice, and made sence given similarity of sample & methods. Only used the 'Novel' condition.
Kazen-Saad, 1983: E2	Effect SD for ER & RT was made equal to PKSB. This was a conservative choice, and made sence given similarity of sample & methods. Only used the 'Novel' condition.
Dietze & Poth, 2022: E1	Means and SDs extracted from raw data.
Dietze & Poth, 2022: E2	Means and SDs extracted from raw data.
Dietze & Poth, 2022: E3	Means and SDs extracted from raw data.
He et al, 2020: Old Sample	Means and SD pulled from table. Effect SD for ER was gemerated by choosing a conservative correlation value (.6) based on prior studies and using the condition variances.
He et al, 2020: Young Sample	Means and SD pulled from table. Effect SD for ER was gemerated by choosing a conservative correlation value (.6) based on prior studies and using the condition variances.
Han & Proctor, 2023: E1	Only compatible trials were used. Means and SD pulled from table

Appendix Table 2. Distinctions in foreperiods across included experiments

Study Name	SOA Notes
Posner et al, 1973	Warning tone of 50 msec followed after constant intervals of 50. 200. 400. (SOA of 100, 250, 450)
Han & Proctor, 2022: No Feedback	50 msec warning tone followed by 50, 200, or 400 (SOA of 100, 250, 450)
Han & Proctor, 2022: Feedback	50 msec warning tone followed by 50, 200, or 400 (SOA of 100, 250, 450)
McCormick et al, 2019: Intense	100 msec marning tone. SOAs are 100 and 250.
McCormick et al, 2019: Isointense	100 msec marning tone. SOAs are 100 and 250.
Dietze, Recker, & Poth, 2023: E2	50 msec warning tone. The 200 msec foerperiod condition is actually 247; the 400 msec foreperiod condition is actually 341 and 353 (but mostly 353)
Dietze & Poth, 2023: Visual Signal	The 200 msec foreperiod condition is made up of the foreperiods 153, 200, and 247. The 400 foreperiod is made up of 388 and 435.
Dietze & Poth, 2023: Auditory Signal	The 200 msec foreperiod condition is made up of the foreperiods 153, 200, and 247. The 400 foreperiod is made up of 388 and 435.
Kazen-Saad, 1983: E1	200 was actually 150 SOA and 400 was 350 SOA.
Kazen-Saad, 1983: E2	200 was actually 150 SOA and 400 was 350 SOA.
Dietze & Poth, 2022: E1	400 msec forperiod was actually a 500 msec SOA.
Dietze & Poth, 2022: E2	400 msec forperiod was actually a 500 msec SOA.
Dietze & Poth, 2022: E3	400 msec forperiod was actually a 500 msec SOA.
He et al, 2020: Old Sample	100 msec signal. SOA of 150 (treated as 200msec)
He et al, 2020: Young Sample	100 msec signal. SOA of 150 (treated as 200msec)
Han & Proctor, 2023: E1	50 msec tone, followed by 50 200 or 400 (SOA of 100, 250, 450).

Appendix Figure 1. Funnel plots run without moderators in model. The x-axis represents the effect sizes. The line is the mean effect size across all experiments, the white triangle represents the likely values across the various standard error values (y-axis). In non-biased fields, experiments should be arranged symmetrically around the center line. Because many of these papers come from the same researchers, and we want to be able to detect possible ‘biases’ based on this, points are marked based on which author published the study. Experiments that generated larger mean effect values are labeled.