Understanding the causes of adapting, and failing to adapt, to time pressure in a complex multi-stimulus environment.

Hector Palada[1], Andrew Neal[1], Rachel Tay[1], & Andrew Heathcote[2]

[1]The University of Queensland, Australia

[2]The Universities of Tasmania and Newcastle, Australia

Author Note

Hector Palada, Andrew Neal, Rachel Tay, School of Psychology, The University of Queensland; Andrew Heathcote, Schools of Medicine and Psychology, The Universities of Tasmania and Newcastle.

Hector Palada was supported by an Australian Government Research Training Program Scholarship. This research was funded by the Australian Government through the Australian Research Council (DP160101891). We would like to thank Daniel White for his programming and technical support of the UAV task. We thank the reviewers for their helpful comments during the review process.

Correspondence concerning this article should be addressed to Hector Palada, School of Psychology, The University of Queensland, St Lucia, QLD 4072, Australia. E-mail: hector.palada@uqconnect.edu.au.

Abstract

We examined how people respond to time pressure factors in a complex, multi-stimulus environment. In Study 1, we manipulated time pressure by varying information load via stimulus complexity and the number of stimuli. In Study 2, we replaced the complexity manipulation with deadline – that is the time available to classify stimuli presented within a trial. We identified several ways that people can adapt to time pressure: Increasing the rate of information processing via effort or arousal, changing strategy by lowering response caution, and adjusting response bias. We tested these mechanisms using the LBA model of choice and response time (Brown & Heathcote, 2008). Whereas stimulus complexity influenced the quality of choice information, the number of stimuli influenced response caution, and deadline pressures caused a failure of encoding that was only partially compensated for by increased effort or arousal. Our results reveal that, rather than having a common response, people adapt, and fail to adapt, to the different time pressure factors in different ways.

Public Significance Statement

The paper shows that individuals use a range of cognitive strategies in response to time pressure in a multi-decision context. Such strategies include lowering the amount of information needed to make a decision and increasing effort. We found that the use of a given strategy depends on how time pressure is varied, whether it be the time available to make decisions or the number of decisions required. We found that these strategies can be ineffective as time pressure becomes excessive, and as a result, performance suffers, sometimes catastrophically.

Keywords: Response time modeling, linear ballistic accumulator, task demands, workload & time pressure

There are many industries that require people to work under high levels of time pressure. Examples include aviation, emergency response, medicine, finance, and defence. Excessive time pressure is a risk, because it can impair performance, producing increases in response times and error rates, which can increase the risk of accidents and incidents (Parasuraman, Sheridan, & Wickens, 2008). For this reason, there has been a long history of research attempting to identify the critical "redline" – the point at which the person does not have sufficient capacity to meet further demands, and thus performance deteriorates (Hart & Wickens, 2010). However, it has proven extremely difficult to identify a clear redline in most occupations, because people can manage their workload to ensure that performance remains within acceptable bounds (Loft, Sanderson, Neal, & Mooij, 2007). For example, people are thought to adapt to changes in time pressure by increasing effort, changing strategy or adjusting task priorities (Hendy, Liao, & Milgram, 1997; Neal, Ballard, & Vancouver, 2017; Raby & Wickens, 1994). Despite over 60 years of research into workload, the field lacks a formal model that explains the mechanisms underlying adaptation to time pressure. As a result, it is difficult to know how a person will adapt to changes in time pressure, and to identify the point at which performance is likely to deteriorate.

Formal psychological models provide an explicit, mathematical description of the latent processes thought to be responsible for observed behaviours (Lewandowsky & Farrell, 2010). Within the workload literature, researchers typically draw inferences about the way that people adapt to changes in time pressure by observing changes in performance, or from self-report measures (Vidulich & Tsang, 2012). However, people may adapt to the same set of demands in multiple ways. As a result, verbally specified models are often ambiguous, and it is not clear whether the latent processes described in the verbal model are able to account for the data (Busemeyer & Diederich, 2010). Furthermore, if people can adapt in different ways, then the point at which performance will start to deteriorate will vary. As a result, there is substantial variability in findings across studies, and it is hard to assess whether the same set of latent processes can account for the different patterns of behaviours across different studies. It is difficult for a field to make progress under these conditions.

Our aim in this paper is to understand the mechanisms by which people adapt to changes in time pressure when performing a target detection task. We apply a cognitive accumulate-to-threshold decision-making model, the linear ballistic accumulator (LBA, Brown & Heathcote, 2008), to quantify these effects. We manipulated time pressure across two studies. In Study 1 we varied the amount of information that needs to be processed. In Study 2 we varied both the amount of information and the amount of time that is available to process it. We fit the LBA to participants' performance and examined how the parameters of the model responded to the manipulations of time pressure. First, we review previous literature on time pressure effects, and then discuss how the mechanisms identified can be mapped to the parameters of the LBA.

Adaptation to Time Pressure

Resource based accounts of workload assume that the humans have limited information processing capacity (Kahneman, 1973). Capacity limits are explained using the concept of resources. Resources are a hypothetical construct, representing units or channels that process information (Navon & Gopher, 1979). The rate at which information is processed for a given task is determined by the amount of resources allocated to the task. The time required to complete a task is determined by the amount of information to be processed (bits of information) and the rate of processing (bits per second, Hendy et al., 1997). Time pressure is determined by the ratio of the time required to complete a task to the time available. An increase in time pressure can be caused by an increase in the amount of information to be processed within the time available, or a reduction of time available to process a given amount of information.

The critical redline of workload is the point at which supply can no longer meet demand, and is the breakpoint of performance, as the individual enters the “overload region”, such that they have no capacity to meet further demands and performance deteriorates (Hart & Wickens, 2010). The redline represents the point at which time required exceeds time available. If a redline can be identified, then measures can be put in place to ensure that workload does not cross that line (Young, Brookhuis, Wickens, & Hancock, 2015). However, surprisingly little progress has been made in identifying the elusive redline. Studies that have attempted to identify the overload region typically use self-report workload to measure cutoffs (e.g., Colle & Reid, 2005), yet self-reported workload is often dissociated from performance (Vidulich & Wickens, 1986; Appl Erg). Furthermore, objective performance measures often respond in different ways to workload manipulations, depending on the strategy that the person uses. Differences in strategy add to the fuzziness of the redline concept, because it means that there is no single point at which performance will degrade (Young et al., 2015).

In principle, there are four ways in which people can adapt to changes in task demands: By increasing the rate of information processing to meet task demands; reducing the amount of information that is processed; prioritizing a subset of responses over others (e.g., by adjusting their response bias); or by extending the decision deadline or re-ordering the tasks (Hendy et al., 1997; Loft, Bolland, Humphreys, & Neal 2009). Extending the deadline or re-ordering the tasks is only possible if the person has discretion over the timing and order of the tasks they perform. In the current paper, we examine tasks in which deadlines are imposed externally, and so we focus on the first three mechanisms.

Rate of Information Processing

Humphreys and Revelle (1984) argue that there are two mechanisms that regulate the rate of information processing. The first is arousal. As task demands increase, more resources are thought to become available, which speeds up the rate of information processing. This is thought to be a relatively automatic response. The second is effort. If the demand for resources exceeds the supply, the person may reallocate resources from lower priority tasks, or from off-task activities, to the primary task to meet processing demands. Effort is thought to be a volitional response that is under motivational control. An individual will only exert greater effort if they are sufficiently motivated to maintain a desired level of performance (Hockey, 2013; Kool, McGuire, Rosen, & Botvinick, 2010; Neal et al., 2017). Increases in arousal and effort are thought to have the same effect, namely to increase the rate of information processing.

There is surprisingly little in the way of direct evidence to support the assumption that people can change their rate of information processing in response to increases in time pressure. There is evidence, however, from dual-task paradigms that information-processing resources are limited, based on an inverse relationship between primary task demands and secondary task performance. Resource based accounts explain this finding by assuming that both tasks draw on a common resource pool, and that as additional resources are allocated to meet information processing demands on the primary task, fewer resources are available for the secondary task, and thus performance degrades (Navon & Gopher, 1979; Norman & Bobrow, 1976).

Given that resources can never be directly observed, the dual-task paradigm provides indirect evidence for changes in rates of information processing. Although these inferences may be reasonable in tightly controlled experimental paradigms where it is possible to rule out alternative explanations, it is much more difficult to draw inferences about the availability and allocation of resources in complex and demanding tasks. For example, some manipulations of task demand may make the task too difficult for increased effort or arousal to compensate, potentially bringing into play alternative adaptations (Neal et al., 2017). It is, therefore, difficult to assess whether the concepts of resources and information processing capacity are necessary or sufficient to account for a given set of behavioral data. Other sources of evidence suggesting changes in rates of information with task demands come from physiological measures, including cardiovascular, eye and brain activity (Zijlstra, 1993), and self-report (Earle, Hockey, Earle, & Clough, 2015). It is thought that these measures are indicators of either effort or arousal, because they correlate with task demands, however, once again it is unclear whether they reflect changes in the rate of information processing, or whether other mechanisms are involved.

Information Reduction

Hendy et al. (1997) argue that people can reduce the amount of information to be processed by changing their strategy for performing the task. This can be done in several ways, for example, by a change in the depth of processing (Donkin, Little, & Houpt, 2014) or by using simpler heuristics (Gigerenzer & Gaissmaier, 2011). In signal detection tasks, the primary way in which the load of information processing can be reduced is by trading accuracy for speed. The inverse relationship between speed and accuracy is referred to as a speed-accuracy tradeoff. The tradeoff between speed and accuracy have been observed in response to deadlines, incentive structures and response signals (e.g., Förster, Higgins, & Bianco, 2003; Pachella & Pew, 1968; Rae, Heathcote, Donkin, Averell, & Brown, 2014), and is pervasive across experimental and applied literature (Standage, Wang, Heitz, & Simen, 2015), as well as in industry (Drury, 1994). Unless the individual is highly motivated to protect accuracy, they may change strategy in preference to increasing effort, as the latter is an aversive experience (Neal et al., 2017; Kool et al., 2010).

Response Bias

A response bias is a strategic preference for one response over the alternatives. For example, in classifying targets and non-targets, an individual may favour responding “target” compared to responding “non-target”. Traditionally, it has been thought that individuals adjust their bias in response to payoffs and base rates. However, there is evidence that under some circumstances, people can adjust their bias in response to workload. Loft et al. (2009) presented air traffic controllers with pairs of aircraft that were on crossing tracks, and asked them to indicate whether they would intervene to assure separation between the aircraft (e.g., by changing the level of one of the aircraft). The false alarm rate was higher when controllers were under higher levels of workload, suggesting they applied a larger safety margin. Loft et al. (2009) argued that this shift in response bias was motivated by the desire to ensure safety over accuracy.

Formally Modelling Adaption

The arguments presented above suggest a range of different ways that people can adapt to changes in task demands. It is difficult to make inferences regarding these mechanisms from behavioral data alone, as there could be a combination of mechanisms being used. These effects may be obscured by changes in difficulty as demand increases, or by a bias for one response over another, making it difficult to infer how people adapt to changes in task demands from behavioral data alone. Formal models allow us to make such inferences, as they quantify the latent psychological processes hypothesised to be responsible for a set of behavioral data.

Systems factorial technology (SFT; Townsend & Nozawa, 1995) represents one approach to this problem. SFT was developed to examine the cognitive mechanisms underlying the processing of multiple sources of information in specialized designs, such as double-factorial paradigms (see Townsend & Eidels, 2011, for a review of formal capacity measures). The response time distributions measured in these paradigms allow SFT to derive a capacity coefficient, quantifying changes in rates of information processing as the amount of information to be processed increases. Using the capacity coefficient, individuals can be classified as having limited, unlimited, or super capacity if information processing rates decrease, remain unchanged, or increase with information load, respectively. For example, Eidels, Donkin, Brown and Heathcote (2010) found that while some individuals increased their rates of information processing when they had to process multiattribute stimuli compared to a single stimulus property, others' processing was constant, or decreased with added information load.

Although SFT was developed for high accuracy tasks, recent developments allow for the model to account for variations in accuracy (Townsend & Altieri, 2012). Nonetheless, SFT is limited in that it requires specialized and highly controlled designs that can be challenging to adapt to complex applied tasks. Furthermore, SFT does not differentiate between the different mechanisms that might underlie changes in capacity, such as changes in the rate of information processing and changes in strategy (but see Eidels et al., 2010 and Bushmakin, Eidels & Heathcote, 2017, for LBA-based approaches to modeling these factors in SFT paradigms). To address this problem, we need a cognitive model that explains how these different adaptive responses influence task performance. In the current paper, we use an accumulate-to-threshold model of decision making, the LBA (Brown & Heathcote, 2008), to identify these different mechanisms.

Linear Ballistic Accumulator Model

Accumulate-to-threshold models of decision-making have provided detailed accounts of the cognitive processes that underlie simple choices. Although the models vary in architecture and complexity, they share a common assumption, namely that the decision-maker samples evidence from the environment until a threshold is reached, triggering a response. These models are able to account for accuracy, the full distribution of correct and error response times, as well as their complex interactions. This is achieved by first partitioning response times into decision time and non-decision time. The latter accounts for the combined times to complete stimulus encoding and response production processes, and is assumed to have a constant value. A change in non-decision time causes a shift in the entire response time distribution without changing the variance or shape of the response time distribution. Decision time and response choices are accommodated by parameters reflecting the input and operation of the evidence accumulation process. The rate of evidence accumulation is determined by the difficulty of the task and the degree of attention allocated towards that information (i.e., effort or arousal). The threshold parameter indexes response caution – the amount of evidence required to make a response, which underlies the tradeoff between speed and accuracy.

The LBA (Brown & Heathcote, 2008; see Figure 1) is a mathematically tractable decision-making model that specifies linear, independent accumulators corresponding to each alternative response (e.g., “target” and “non-target” accumulators). For a given stimulus, each accumulator has its own normal distribution of evidence accumulation rates, with the mean rate v, and standard deviation s_v. The normal distribution is truncated at zero, such that only positive drift values can be randomly sampled from the rate distributions. The start-point of accumulators is assumed to vary as a uniform distribution with a lower boundary of zero and an upper boundary, A. The model includes a threshold, b, for each response alternative, with the selected response being determined by the accumulator that reaches its threshold first. The distance from the top of the start point distribution to the threshold is denoted B. Finally, the model includes a non-decision time parameter, t_{er}, reflecting sum of times for stimulus encoding and response production.

The LBA accounts for changes in difficulty and speed of information processing via rate parameters. The accumulator for the correct response alternative is referred to as the “matching” accumulator, because the response that is produced by the accumulator “matches” the stimulus being presented. The accumulator for the incorrect response is referred to as "mismatching" accumulator, because the response produced by the accumulator does not match the stimulus being presented. For example, when a target stimulus is presented, the target accumulator is the matching accumulator, while the non-target accumulator is the mismatching accumulator. In this case, the matching accumulator will produce a correct acceptance if it reaches the threshold before the mismatching accumulator. When a non-target stimulus is presented, the non-target accumulator is the matching accumulator, while the target accumulator is the mismatching accumulator. In this case, the matching accumulator will produce a correct rejection if it reaches the threshold before the mismatching accumulator.

An increase in the rate of information processing that reflects heightened effort or arousal could be indexed by an increase in the drift rate for both the matching and mismatching accumulator. The mean drift rate has been shown to converge with SFT's capacity measure (Eidels et al., 2010). A change in the difficulty of the choice, by contrast, affects the quality of information that the person is able to extract, which is indexed by the difference between the matching and mismatching rates. When performance is at chance, the drift rate for the matching accumulator equals that of the mismatching accumulator, as the two accumulators have an equal chance of triggering a response assuming an equal threshold. As it becomes easier to discriminate between the correct and incorrect response, the difference in rate between the matching and mismatching accumulator increases; specifically, the mean of the matching accumulator should be greater than the mismatching accumulator. As noted earlier, it is possible that people may compensate for an increase in difficulty by applying more effort (Neal et al., 2017). This could be reflected in greater attention to diagnostic stimulus attributes (i.e., attributes relevant to the choice), which would produce an increase in information quality, and hence produce a greater difference between match and mismatch rates. We use rate estimates to determine whether different time pressure manipulations differentially affect the overall rate and quality of information processing.

The LBA accounts for strategic changes in the amount of evidence accumulated and in response bias via the threshold parameter. Figure 1 shows how adjusting the threshold parameter can account for speed-accuracy tradeoffs. Since the start point parameter is randomly sampled from a uniform distribution, on some trials the non-target accumulator will start closer to the threshold compared to the target accumulator, and vice-versa. In the case shown in Figure 1, the non-target accumulator has a greater start point compared to the target accumulator. With an equal threshold for both accumulators, the initial greater start-point allows the non-target accumulator to reach its threshold first, resulting in a non-target response regardless of the non-target accumulator having a lower rate of evidence accumulation. However, the greater start point for the non-target accumulator can be overcome by setting a higher threshold. Specifically, since the target accumulator rate is greater than the non-target accumulator, indicated by the steeper line, a sufficiently high threshold would result in a target response. Setting a higher threshold overcomes the differences in start points, which are assumed to occur randomly from trial to trial, and thus increases accuracy. However, this comes at the expense of a greater response time, as it takes longer for the accumulator to reach the threshold.

Threshold parameters also account for response biases, by allowing threshold values to differ between response alternatives. All else being equal, having a lower threshold for one response compared another means that the accumulator has lesser distance to travel. Therefore, the accumulator is more likely to reach its lower threshold first, and will trigger the corresponding response more frequently. Thus, the favoured response will occur more quickly and more frequently than the response with a higher threshold.

The LBA and other accumulate-to-threshold models are able to separate the different effects of latent cognitive processes that are conflated in observed data. The term "selective influence" describes the idea that the effect of a manipulation is captured by a single model parameter. For example, changes in difficulty are commonly reflected in the quality of evidence that can be extracted from the stimulus but have no influence on response caution and response bias (i.e., thresholds), at least when the changes in difficulty are not predictable. Selective influence has practical as well as theoretical implications. For example, if difficulty affects only rates there is little point attempting to ameliorate its effects with interventions targeting response strategies. However, studies using relatively simple choice tasks suggest that task demands may sometimes influence several cognitive processes. For example, when increases in difficulty are predicable participants often try to compensate by increasing thresholds. Therefore, we test for such failures of selective influence in circumstances where they might reasonably occur, such as when time pressure manipulations are predictable.

Accumulate-to-threshold models have been used previously to test selective influence in a range of relatively simple laboratory paradigms, such as perceptual discrimination, lexical decision and recognition memory tasks. For example, Rae et al. (2014) found that in all of these tasks participants reduced their threshold when instructed to emphasize speed over accuracy, as is conventionally assumed. However, contravening these assumptions, they also observed an increase in the rate of evidence accumulation in the accuracy condition, concluding that prioritizing accuracy results in an increase in the quality of information processed. Dambacher and Hübner (2015) found similar effects with a deadline manipulation, such that thresold and rates declined with tighter deadlines. They also observed that non-decision time decreased when participants had tighter deadlines.

Palada and colleagues (2016) extended this approach to a more complex task, testing whether the LBA was capable of accounting for performance on a simulation involving a navigation and target classification task that is more representative of the types of tasks found in applied settings. They found that the LBA was able to provide an accurate characterization of all aspects of performance even though multiple stimuli were simultaneously present in the display, and participants were free to decide how they allocated their attention amongst tasks and stimuli. The study included manipulations of time pressure, specifically the number of stimuli requiring classification, and the perceptual difficulty of stimuli. In line with conventional assumptions, time pressure selectively influenced threshold, with a decline in response caution under greater load, but did not influence the rate of information processing or response bias. Again, in line with conventional assumptions, their manipulation of perceptual difficulty selectively influenced rates, with poorer discrimination under greater difficulty; as difficulty increased, the difference between the rates of the matching and mismatching accumulators decreased.

There were several shortcomings in Palada et al. (2016) that limit the inferences that can be made about how individuals adapt to time pressure. Firstly, participants could control the order in which they attended to stimuli, and so could potentially switch attention back and forth between stimuli before making a classification on any one. In contrast, the modelling assumed that stimuli were processed one at a time. Palada et al. provided evidence that the modelling assumptions was fulfilled by scoring response time beginning from the last response, but noted that there was still some potential inaccuracy. Such inaccuracies would occur if processing of the next stimulus began during the period that the response to the last stimulus was being produced, if processing was interrupted by the secondary task which alarmed when performance degraded, and if processing of a stimulus was interrupted because it exited the screen, resulting in an overestimation of the subsequent response time as it would include the partial processing time of the previous stimulus. In the present studies, we took a more controlled approach, enforcing serial processing by requiring participants to unmask and decide about each stimulus sequentially. This allowed response time to be unambiguously defined from the point at which a stimulus is unmasked. We also removed the secondary task.

A second limitation is that Palada et al. (2016) used a range of randomly varying stimulus configurations that differed markedly with respect to the difficulty of the target vs. non-target classification. The hypothesis that the heterogeneity of stimulus difficulty would be accounted for by the rate variability parameter was supported, with target variability being greater than non-target variability. The heterogeneity of target difficulty effect was further explored in a post-hoc analysis by specifying “easy” target and “hard” target subtypes. The variability effects dropped out of the selected model, and the model accounted for the differences in easy and hard target effects via the mean rate parameter. However, the conclusions made from the post-hoc modeling were limited, as it lacked power due to insufficient observations within cells after having broken down targets into subtypes. In the present study, stimulus difficulty is balanced and its effects are explicitly modeled. We expect stimulus difficulty to be captured by a change in inputs to the decision process (i.e., a rate effect), though we also examine whether individuals adapt to difficulty by changes in response caution (i.e., a threshold effect) as well.

The third limitation is that Palada et al. (2016) only manipulated one factor affecting time pressure, namely the number of stimuli to be processed. A range of factors influence time pressure, including the number of stimuli, stimulus complexity, and the time available. It is not known whether people respond to these factors in the same way, suggesting a single mechanism mediating all time pressure effects, or whether the factors have independent selective influences, or some combination of the two. Consistent with there being more than one mechanism, the factors determining time pressure have been shown to have independent effects on perceived time pressure (Hendy et al., 1997; Loft et al., 2007).

In our studies, we manipulated three time pressure factors. In our first experiment, we re-examined the effects of the complexity and number of stimuli using our more controlled design, both in terms of attention and stimulus difficulty. In our second experiment, we examined how individuals adapt to changes in the number of stimuli and deadline. This examination of a broad range of factors aimed to provide a more detailed understanding of how individuals adapt to time pressure.

Experiments and Model Variants

Both experiments used a dynamic target-detection task. Participants had to classify ships as targets or non-targets as they moved across the screen. We manipulated time pressure by varying the number of stimuli to be processed, the complexity of the stimulus, and the time available to make the response, and used model selection to examine how these variables influenced the parameters of the model. For both studies, we tested a set of alternative models describing the different ways in which the experimental variables could influence threshold, rate and non-decision time parameters. In doing so, we could also test whether the time pressure manipulations had similar or different effects on adaptations to time pressure. We identified the best account of the observed data by using model selection techniques that factor in both the simplicity of the model and its goodness-of-fit.

In fitting the LBA to the observed behaviours, we used the model selection strategies outlined in Donkin et al. (2011). The model was fit using maximum likelihood estimation. Having fit several alternative models that vary in their parameterizations, we selected the model that had the best tradeoff between simplicity and goodness-of-fit via Akaike Information Criterion (AIC), where AIC = D + 2p. The term D (-2 times the maximized log-likelihood) quantifies model deviance, and is a measure of model misfit analogous to sums of squares errors for a Gaussian model. The term 2p penalizes models with more parameters, so unless adding a parameter decreases the deviance by more than two, AIC rejects the more complex model. Having determined the AIC selected models, we examined their goodness-of-fit graphically to determine if they account for the effects of the experimental manipulations error rates and on response time distributions. We examined both error and correct response times, quantifying fast, median and slow responses as the 10th, 50th and 90th percentiles of the response time distribution. Finally, we analyzed the estimated parameter values to examine the effects of our time pressure manipulations on decision-making processes, and how individuals adapted to time pressure.

The set of model variants were generated from a highly flexible "top" model, which will necessarily fit the data best. The simpler models introduce restricted parameterizations, with the simplest model having the same parameter values across all experimental factors except for the flexibility to perform better than chance (by allowing a rate difference between match and mismatch accumulators). As such, the simplest model estimates a single value for the parameters B, A, s_v, t_{er}, whereas two values are estimated for v – one corresponding to each accumulator. The simplest model was fit first and thereafter, the best-fitting values informs the optimization algorithm start-point for the next model variant where one extra factor is introduced. This method continues up until the top model is fitted, with more complex models fit from many different start points, and reduces the chance of the fitting procedure finding sub-optimal solutions (Donkin et al., 2011; for examples, see Heathcote & Love, 2012, Palada et al., 2016, and Rae et al., 2014).

The architecture of the LBA introduces two accumulator-related factors, a response (*R*) factor and a match (*M*) factor. The *R* factor accounts for the alternative responses, in our experiment being either a "target" or "non-target" response, and enables the estimation of response bias by allowing for a lower threshold value for the favored response (e.g., a lower threshold for the target response would indicate a "target" bias). The *M* factor accounts for the match or mismatch between accumulator and stimulus (e.g., if a non-target stimulus is presented, then the match level corresponds to the non-target accumulator and the mismatch level corresponds to the target accumulator). Above chance performance is accommodated by allowing the rate (*v*) to be greater for the match than mismatch accumulator. Doing so means that the matching (i.e., correct) accumulator has a greater chance of sampling a higher rate of evidence accumulation than the mismatching (i.e., incorrect) accumulator and so that the matching accumulator is more likely to reach its threshold first and trigger a correct response. The difference between the matching and mismatching accumulator's rates, when standardized by the rate standard deviation, is analogous to *d'* in signal detection theory.

There were two experimental factors common to both studies: stimulus type (*S*) and number of stimuli (*NS*). A third factor was complexity (*C*) in Study 1 and deadline (*D*) in Study 2. Using the model selection method, there is an exponential increase in the number of possible model variants that can be generated from the top model as a function of the sum of the number of factors (*NF*) mapped to each model parameter, for a total of 2^{NF} models (Donkin et al., 2011). The computational demands of model fitting procedures meant that it was not feasible to specify a top model where all model parameters could vary by all experimental factors. We outline the assumptions that allowed us to reduce the complexity, and therefore computational demands, of fitting the set of LBA model variants.

Firstly, it is circular to argue that the stimulus factors related to the choice (i.e., whether the ship is a target or non-target) adjusts the accumulator-related processes, as these stimulus properties are only known once an accumulator reaches its threshold. Hence, there is a strong argument that accumulator-related parameters (*A* and *B*) should not be influenced by stimulus factors that are not known at to the start of a trial. We also assumed that non-decision times did not differ between stimulus types, corresponding with the assumption that encoding and response production times do not differ between targets and non-targets. Further, we assumed that time pressure manipulations would not influence rate variability (sv), though we did allow rate variability to be a function of the match factor (for examples using the same approach see Palada et al., 2016 and Strickland, Heathcote, Remington, & Loft, 2017). We estimated a single value for the start point parameter (A) for all conditions. Although this assumption is not necessary, the *A* parameter is hard to estimate and rarely been found to account for experimental effects (for an exception see Heathcote & Hayes, 2012).

The top models for each study are outlined in Table 1, with both having 67 parameters. One accumulator parameter must have a fixed value to make the model identifiable (Donkin, Brown, & Heathcote, 2009); we fixed s_v for the non-target mismatching accumulator to one. With these limitations, 512 models were fit to each participant for each study.

Study 1: Information Load

The first study examines how individuals adapt to time pressure factors that determine information load – that is the amount of information to be processed. We manipulated two factors that are representative of applied situations: The complexity of the stimuli, and the number of stimuli requiring classification. The complexity factor determines the amount of information that needs processing to classify a single stimulus, whereas the number of stimuli acts as a multiplier and determines the total required amount of information processing within the time available. In line with Palada et al. (2016), we hypothesise that participants will respond to an increase in the number of stimuli by lowering their threshold. We further hypothesise that complexity will influence the inputs to the decision process, i.e., a rate effect, as complexity is a property of the stimulus. Because complexity is unknown at the start of the trial it seems likely that it will not affect thresholds. However, the decisions in this task are relatively slow, so it is possible that this may enable at least a limited change in response caution or response bias; we used rigorous model selection procedures to test these possibilities, as well as potential effects of complexity on stimulus encoding and hence non-decision time.

Method

Participants

Fifty-eight undergraduate students from the University of Queensland participated in the study in return for course credit. The sample included 30 females and 28 males, with a mean age of 19.87 years (*SD* = 5.10).

Experimental Task

The target detection task presented a UAV camera ocean view with multiple ships synchronously passing the camera's field of view (see Figure 2). The ocean view included five ship lanes, and ships were positioned on adjacent lanes. Within discrete trials, ships transitioned into the screen, entering from the right and moving leftwards across the screen at a constant speed of 247 pixels/s, making responding possible over a 9s period (from when the front of the 300-pixel length ship appeared on the right of the screen of width 1920 pixels until the front of the ship had disappeared on the left). There was a 1s interval between trials.

Peripheral ships were masked, so that at any given point in time all resources were allocated to the focal ship requiring classification. Ships were unmasked beginning from the bottom lane, and working upwards; once the participant had classified the visible ship, the next ship lane positioned above was unmasked, and the classified ship's lane was masked. Arrows indicated the position of ships within masked ship lanes; classified ship's arrows were filled with the corresponding classification colour, whereas unclassified ships had an unfilled arrow with a yellow border. Response boxes travelled synchronously above each ship. Participants made a target or non-target classification by clicking on the green or red portion of the response box, respectively. Participants were unable to reclassify ships.

Representative of applied contexts involving surveillance footage, the simulation included noise via dynamic grey-scale noise overlay, and individual ships were obscured by a fog overlay of constant opacity (65%).

Experimental Design

The experiment had three within-subject factors producing 18 conditions: stimulus type (target and non-target), complexity (3-, 5-, and 7-feature rule), and number of stimuli per trial (2, 3, and 4 ships). Stimulus type (target vs. non-target) was defined by a classification rule that specified the number of features that needed to by present for the ship to be a target. The features included an anchor, crane, fishing line, life boat, mast, smokestack and flag.

Complexity and number of ship factors created nine time pressure conditions. There were 30 trials at each level of time pressure, with trials from each level blocked to avoid the convolution of complexity rules. The order of presentation of the nine time pressure blocks was randomized for each participant. The levels for the manipulations of complexity and number of stimuli were calibrated based on pilot data from the same population to ensure substantial difference in performance across experimental conditions, and to avoid floor or ceiling effects.

Complexity was manipulated by varying the number of features in the decision rule. The features had fixed positions on the ship, and were either present or absent. Participants had to count the total number of features on the ship. In the least complex condition, participants used a three-feature rule. Up to three features could be present on the ship, and the rule specified that any two of the three possible features needed to be present for the ship to be a target. In the intermediate condition, participants used a five-feature rule. Up to five features could be present; any three of the five possible features needed to be present for the ship to be a target. In the most complex condition, participants used a seven-feature rule. Up to seven features could be present; any four out of the seven possible features needed to be present for the ship to be a target.

The number of possible ship configurations (i.e., the presence and absence of features) for rule is a function of 2^n, where n represents the maximum number of features. As such, the 3-, 5-, and 7-feature rules produced, respectively, 8, 32, and 128 possible ship configurations, half of which were consistent with a target definition. For each trial, the simulation randomly selected a target or not target, and then randomly selected one of the possible corresponding ship configurations.

Two stimulus sets were created to control for potential heterogeneity in the difficulty of identifying individual features. In set A, the 3-feature rule included a crane, mast and flag, while the 5-feature rule also included a fishing rod and life boat. In set B, the 3-feature rule included an anchor, crane and life boat, while the 5-feature rule also included a mast and smoke stack. The 7-feature rule was consistent across both sets, and included all features. Use of the two stimulus sets were counterbalanced across participants.

Procedure.

Initially, participants viewed an audio-visual presentation of task instructions. A 5-minute training phase was completed, exposing participants to all nine levels of time pressure. Participants were first presented with a block of trials using the 3-feature rule; three trials were completed with 2 stimuli, followed by three trials with 3 stimuli, finishing with three trials with 4 stimuli. The same order was then completed using the 5-feature rule, and then the 7-feature rule. Before each block, participants were informed of which classification rule they should use. During training participants could refer to a handout outlining the classification rules. In training, participants received feedback for each response; a green tick for correct or a red cross for incorrect responses travelled behind classified ships. After training, the experimenter asked if the participant understood the task. When necessary, clarifications were made by the experimenter. Once the participant stated that they understood the task requirements, the classification rules handout was removed, and the participant commenced the experimental phase of the study. In attempts to reduce shifts in decision-making strategies as the participants progressed through the experimental trials, feedback was not provided. Overall, the experiment lasted an hour. Ethical approval for this study was granted by the School of Psychology at the University of Queensland, Australia.

Results

Initially, we examined experimental effects on manifest behaviour. Stimulus set was entered as a control variable, and is not included in the interpretation of results. We analyzed non-response rates and accuracy using a generalized linear mixed-model (GLMM) specifying a binomial distribution with a probit link function, whereas response time measures were analysed with GLMM specifying a Gaussian distribution. Analyses were conducted using the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2014), and inferences made via Wald χ^2 tests with type III sums of squares as implemented by the *car* package (Fox & Weisberg, 2011).

Non-Response Rates

Mean non-response rates were 2.01% (SD = 5.34%). Due to having insufficient observations for model fitting, we removed one participant with 39% missing data. Remaining participants had 10% missing data or less, leaving a sample of 57 participants. There was a significant three-way interaction between complexity, number of stimuli, and stimulus type on non-responses (see Figure 4; $\chi^2(4) = 10.86$, $p = .028$). Non-response rates were similar across conditions, except for the four-ship conditions, where complexity had a stronger effect for non-targets compared to targets. Given the small percentage of non-responses (with the highest mean value of 3.15% across conditions; see Figure 4), it seems unlikely that this interaction would have a substantial effect on the remaining results.

Although all other interactions and main effects did not approach significance (*ps* > .145), Figure 4 shows a trend for an increased in non-response rates with an increasing in the number of ships ($p = .150$).

Accuracy and Response Times

Response times outside of 0.5s and 6s were removed from statistical analyses and model fits; 0.52% of data was censored, with at most 4.11% of data removed from a given participant. Consistent with the usual practice with simple choice tasks, the lower cutoff was selected because lesser response times would be insufficient to make the decision required in our task. The upper cutoff was quite long to account for the slower and more variable performance in our complex task, with only a small percentage of slow responses being removed.

We analyzed four response time-based measures (mean and standard deviation for correct and error responses) and error rates. Predictors included within-subject experimental factors (stimulus type, number of stimuli and complexity). Table 2 summarises the results for all measures. There was a three-way interaction involving error rates; complexity increased errors, however this effect only occurred for targets, and became stronger as the number of stimuli increased (see left panel in Figure 5).

Correct and incorrect responses became faster and less variable as the number of stimuli increased. However, for correct response time there was an interaction between complexity and number of stimuli, such that response times slowed with increased complexity, and this effect was stronger when there were fewer ships in a trial (see right panel in Figure 5). In summary, a complex speed accuracy tradeoff was observed in response to number of stimuli and complexity. For non-targets, responses became faster and slightly less accurate as the number of stimuli increased. For targets, this speed-accuracy tradeoff in response to the number of ships was more pronounced as complexity increased. Response time variability decreased with a greater number of ships for correct (0.93, 0.81, 0.71) and incorrect (0.92, 0.81, 0.72) responses.

Model Selection and Fits

The fits for the top model and the AIC-selected model is shown in Table 3. The AIC-selected model did not fit significantly worse than the top model, $\chi^2(2052) = 759$, $p > .999$, and had less than half the number of parameters. As the AIC selected variant provided the best account of the data in terms of goodness-of-fit and number of parameters, we focus on it in further analyses.

The model accounted for the effects of complexity via rate (v) and number of stimuli via threshold (B). Both factors also influenced non-decision time (t_{er}). Differences in responses to target and non-target stimuli were accounted for by mean rate (v) and rate standard deviation (s_v), as well as by differences in threshold (i.e., a response bias).

Figure 6 shows the goodness-of-fit for the selected model. The model provides an accurate account of the trends in accuracy and response time, capturing the difference in accuracy between target and non-targets, as well as the strong effect of number of stimuli on response times. There is some systematic underestimation of response times at the .5 and .9 quantiles for the 2-ship condition, but the “top” model did not provide a better fit. The underestimation could have occurred because in the 2-ship condition participants occasionally took advantage of the lack of time pressure to recheck the ship configuration (i.e., accumulate evidence for a second time), thus slowing responses. However, this effect appears relatively minor, and so we now consider how the AIC model parameter estimates explain the observed experimental effects

Experimental Effects on Parameters

We analyzed model parameters with GLMMs assuming Gaussian distributions. The “target” threshold was significantly lower than the “non-target” threshold (1.91 vs. 2.09, respectively), $\chi^2(1) = 4.80$, $p = .028$, indicating a target response bias. Threshold decreased as the number of stimuli increased (2.82, 1.93, 1.24), $\chi^2(2) = 151.97$, $p < .001$, and this effect was additive with the response factor.

There was a three-way interaction between complexity, match factor and stimulus type on rate, $\chi^2(2) = 9.23$, $p = .010$. Figure 7 reveals that, in trials where targets were presented, the difference in rate for the matching (i.e., correct) accumulator and mismatching (i.e., incorrect) accumulator became smaller as complexity increased, thus increasing the chances of the mismatching accumulator sampling a higher rate and reaching its threshold first, resulting in an incorrect response. For non-targets, complexity had little effect on the difference in rates between accumulators, with the matching accumulator having a higher rate than the mismatching accumulator, regardless of complexity. As the difference between the matching and matching accumulators is analogous to discriminability, the three-way interaction suggests that an increase in the complexity of the decision rule made it harder to identify targets, but had relatively little effect on participants' ability to identify non-targets.

There was a trend for drift variability being greater for the mismatching (1.00) than the matching accumulator (0.81), though this did not achieve significance ($p = .078$). The main effect of stimulus type and its interaction with the match factor were non-significant (*ps* > .614). Finally, we found a significant increase in non-decision times with an increase in the number of stimuli (0.35s, 0.39s, 0.46s), $\chi^2(2) = 20.04$, $p < .001$; the main effect of complexity and its interaction with number of stimuli were non-significant (*p*s > .816).

Discussion

In this study, we tested the effects of information load on adapting to time pressure. In line with Palada et al. (2016), we varied the number of stimuli requiring classification.

Additionally, we manipulated stimulus complexity. Both manipulations changed the level of time pressure, but had different effects on the parameters of the LBA.

Our manipulation of the number of stimuli, although analogous to Palada et al. (2016), used trials with a predetermined number of ships, rather than a probabilistic manipulation to draw the number of ships presented a 3-minute period. Regardless of this difference, like Palada et al., we found that participants did not respond to the number of stimuli by changing their rate of information processing. Rather, the number of stimuli selectively influenced threshold; participants strategically lowered their threshold when more ships were present, prioritizing the speed over the accuracy of their choices to reduce non-responses. However, this strategy was not completely effective, as evidenced by the trend for non-responses rates to increase with the number of stimuli (see Figure 4).

Contrary to Palada et al. (2016), who did not observe changes in non-decision times, we found that non-decision times slowed with a greater number of ships. It seems likely that the locus of the divergence in findings is stimulus encoding time, rather than response production time, because stimulus features differed across the studies, whilst response modalities were identical. However, the evidence for this effect was weak, as we explain below when describing additional analyses.

Increasing the complexity of the stimuli made it harder to identify targets, but had no effect on non-target accuracy. The model explained these findings by the difference between matching and mismatching rates. The difference was much less for targets than non-targets overall, and for targets it decreased as complexity increased, whereas the difference was constant for non-targets.

Participants appear to have attempted to compensate for the difficulty of targets with a bias to respond target (i.e., setting a lower threshold for targets than non-targets), however, threshold and response bias did not vary as a function of complexity. Although the lower target threshold would have slightly improved performance on target trials, as the target accumulator had lesser distance to travel, this was overshadowed by the rate effects. Specifically, the difference between accumulator rates for non-target trials means that it was unlikely for the mismatching (i.e., target) accumulator to reach its lower threshold first, resulting in infrequent errors on non-target trials. Note, however, that additional analyses described below suggests that there is overall weak evidence for this target bias. We found no evidence that participants adapted to changes in complexity by adjusting their overall rate of information processing, as the main effect of complexity on rate was non-significant.

Consistent with Palada et al.'s (2016) supplementary analysis, which accounted for easy and hard targets, our controlled manipulation of complexity did not influence drift variability. Therefore, the demands of complexity were homogenous between target and non-targets. The stimulus factor was, however, selected by the AIC model as affecting rate variability, though the main effect of stimulus type on rate variability was non-significant. This suggests that, although on average there was no main effect, there were individual differences, with the selected model suggesting that targets sometimes had more or less variability in rates than non-targets.

Given that ships transitioned into the screen, participants could not immediately process all ship features on the first visible ship during the early period of response time. Thus, the stimulus encoding time for the first ship may have been slowed compared to subsequent ships. To test this possibility, we fit the AIC selected model and introduced a new ship order factor (first ship vs. subsequent ships) and allowed it to influence non-decision time. Introducing the ship order effect significantly improved model fit compared to the original model, $\chi^2(513) = 11894$, $p < .001$. Non-decision time was significantly greater for the first ship compared to subsequent ships (0.97s vs 0.27s, $\chi^2(2) = 251.35$, $p < .001$). The parameter estimate for the first ship condition is in line with the time it took for the first ship to transition into the screen, and the estimate for subsequent ships condition is in line with conventional studies when all stimulus features are presented simultaneously. The model provided the same explanation for the number of ships on threshold, and complexity on discriminability, though the target response bias and the effect of number of ships on non-decision time were no longer significant.

In this study, our modeling suggests that participants partially responded to information load. Stimulus complexity influenced rate, particularly the quality of information for targets. There was some evidence that participants responded to this effect by adopting a target response bias; however, as the difference in response thresholds was only minor, and did not vary with complexity, target accuracy declined considerably with complexity. This suggests that participants were approaching the redline as they did not effectively compensate for the increase in complexity. In contrast, participants responded to the number of stimuli by adjusting their threshold. In doing so, there was only a minor drop in performance as additional ships were presented, and thus participants avoided the redline by adjusting their strategy.

Study 2: Information Load and Time Available

The previous study examined how people adapt to changes in information load caused by changes in the complexity and number of stimuli. In Study 2, we kept complexity constant, using the medium difficulty (5-feature) classification rule, and added a deadline manipulation (i.e., the time available to complete a trial) to the manipulation of the number of ships in each trial. Given previous research suggesting that deadline influences threshold (Dambacher & Hübner, 2015), we were interested in whether an interaction would occur between number of stimuli and deadline on threshold, or whether we would find a selective influence of each factor. Hence, we factorially manipulated the number of stimuli in a trial, using the same three levels as in the Study 1, and trial deadline, so that the time available to classify the ships within a trial was either less than, equal to, or greater than the time available in Study 1. In line with previous research, we hypothesised that both time pressure factors would influence threshold, such that people would respond to greater time pressure by lowering their threshold in a dynamic environment.

Study 1 blocked the presentation of time pressure conditions to avoid participants experiencing confusion about which classification rule to follow on a given trial. However, applied decision making often takes place in an uncertain and dynamic environment. For example, air traffic controllers have to be able to adapt to changes in time pressure caused by changes in traffic, weather, schedules, and unexpected events (e.g., airspace closures or aircraft emergencies). For this reason, we explored a more realistic scenario where time pressure conditions varied randomly between trials. Note that participants could become aware of the level of time pressure before they responded on a trial because of the speed at which ships transitioned into the screen, where all features on the ships were not fully visible until about 1.0-1.5s into the trial. Hence, as in Study 1, participants could in theory make different strategic adjustments for different time pressure conditions, although this may have required more flexibility than under the blocked conditions of Study 1.

Method

Participants

Sixty-three undergraduate students from the University of Queensland completed the study in exchange for course credit. The group included 30 females and 28 males. Mean age was 19.49 years (*SD* = 3.87).

Experimental Task, Design and Procedure

Study 2 used the same target detection task as in Study 1. There were 18 within-subject experimental conditions: stimulus type (target and non-target), deadline (6, 9, and 12s), and number of stimuli (2, 3, and 4 ships). The conditions were also calibrated based on pilot data using a sample from the same population. Thirty trials were completed for each of the nine conditions that were created by crossing deadline and number of stimuli. The presentation of conditions was randomized.

The stimulus type and number of stimuli manipulations were identical to Study 1. We manipulated deadline by varying the speed at which ships travelled (6s: 370 pixels/s, 9s: 247 pixels/s, and 12s: 185 pixels/s), and thus the time available to classify the ship. The 5-feature rule from Study 1 was used across all conditions, and we included the same between-person ship configuration manipulation.

The same procedure was used as Study 1. The 5-minute training phase consisted of nine blocks, starting with a 12s deadline and three trials for 2, 3, and then 4 ship conditions; this sequence was then repeated for the 9s deadline and then the 6s deadline. This fixed order was adopted to make it easier for participants to cope with the more demanding short-deadline conditions. Thereafter, participants completed experimental trials. Ethical approval for this study was granted by the School of Psychology at the University of Queensland, Australia.

Results

We used corresponding censoring and statistical procedures in Study 1. Mean non-response rates were 3.25% (SD = 3.86%). We removed three participants with non-response rates greater than 10% (11%, 20%, 22%), resulting in a final sample of 60 participants. Non-response rates increased with number of stimuli in a trial (0.74%, 1.33%, 4.29%; $\chi^2(2)$ = 95.45, p < .001), and shorter response deadlines (6.72%, 0.76%, 0.07%; $\chi^2(2)$ = 11.14, p = .004). The remaining main effects and interactions were non-significant (ps > .290).

Accuracy and Response Times

We used the same criterion and reasoning from Study 1 to censor response times outside the range of 0.5s and 6s; 0.45% of the data was removed, with at most 0.99% removed from a given participant. Table 4 summarises the results for behavioural measures and experimental effects. There were significant interactions between stimulus type, number of stimuli and deadline on error rates and mean correct response time (see Figure 8). Error rates increased with tighter deadline, and this effect was stronger for targets than non-targets. The interaction between stimulus type and deadline became weaker as the number of stimuli increased. Figure 8 also reveals chance accuracy for targets classified under a 6s deadline across all number-of-stimulus conditions, whereas non-target accuracy remained well above chance.

Correct response times generally sped up with tighter deadlines and an increase in the number of stimuli. Trends were similar for targets and non-targets; however, for target ships there was little difference in response times between 9s and 12s deadlines. Participants took more time for non-target decisions in the 12s than 9s condition, yet improved very little in accuracy. In contrast, they did not appear to make use of the additional time available in the 12s condition, yet target accuracy was much better than in the 9s condition (see Figure 8).

Model Selection and Fits

Table 5 shows the fits for the top model and AIC-selected model. The reduction in fit for the AIC-selected relative to the top model was not significant, $\chi^2(2700) = 818$, $p > .999$. The AIC-selected model has less than a quarter of the number of parameters compared to the top model. Hence, we selected the AIC variant as providing the most parsimonious account of the data, and focus on this model in further analyses.

Consistent with Study 1, the model attributed the effects of number of stimuli to the threshold parameter (B). Unlike Study 1, but consistent with Palada et al. (2016), number of stimuli did not affect non-decision time (t_{er}). Deadline effects were explained by both rate (v) and non-decision time (t_{er}), with no effect on threshold.

The goodness-of-fit of the selected model is shown in Figure 9. The model provides an excellent account of accuracy, with little misfit and capturing the complex three-way interaction on accuracy described above, and the chance accuracy for targets under the 6s deadline. The fit for correct and error response time captures the trends across experimental conditions. As in Study 1, there was a systematic underestimation of slower response times for the 2-ship condition, although in Study 2 this was less pronounced and restricted to the 0.9 quantile. Like Study 1, participants may have rechecked their decision prior to making an overt response, and hence extended response times at the tail end of the distribution. Again, the “top” model fit did not improve this underestimation.

Experimental Effects on Parameters

We analyzed model parameters using the same procedure as Study 1. Consistent with Study 1, threshold declined with an increase in the number of stimuli (2.71, 2.10, 1.69), $\chi^2(2) = 348.84$, $p < .001$. For rate, all main effects of deadline, match factor and stimulus type were significant, as well as their interactions ($ps < .001$). We examined the main effect of deadline on rates, which would reflect strategic changes in effort or arousal. We found a significant main effect of deadline, such that mean rates increased with tighter deadlines (2.09, 1.64 and 1.19). However, deadline also interacted with the match factor and stimulus-type, producing a three-way interaction (see Figure 10). The difference between the matching and mismatching accumulator rates, which reflects discriminability, decreased as the deadline became tighter, and this effect was stronger for targets than non-targets.

Figure 10 shows an apparent decrease in match rates with tighter deadline for targets, whereas in all other cases an increasing effect was found. To examine the reliability of this pattern, we broke down the three-way interaction by examining the two-way interactions between deadline and accumulator type on rates for each stimulus type. Both two-way interactions were significant (target: $\chi^2(2) = 285.14$, $p < .001$; non-target: $\chi^2(2) = 40.34$, $p < .001$). The two-way interactions were further broken down by examining the simple effects of accumulator type. Deadline had the opposite effect on the match and mismatch rates for targets; while the target match rate decreased with tighter deadlines, the target mismatch rate increased (*p*s < .001). For non-targets, match rates and mismatch rates increased with tighter deadlines, with the latter having a stronger positive effect (*p*s < .001).

There was a significant interaction between stimulus type and match factor on rate variability, $\chi^2(1) = 4.93$, $p = .026$. Figure 10 shows that rate variability was greater for the mismatching than matching accumulator, and this effect was stronger for non-target stimuli. Finally, we found a significant decrease in non-decision time with shorter deadlines, $\chi^2(2) = 222.29$, $p < .001$, and effect that was particularly marked for the 6s condition (0.14s), which was less than half the non-decision time for the 9s and 12s conditions (0.30s and 0.40s).

Discussion

In this study, we examined how people adapt to dynamic changes in both information load and deadline. We did this by randomly varying the number of stimuli and the time available to classify all stimuli presented within a trial. Initially, we were interested in whether the time pressure factors would both selectively influence threshold. Instead, we found distinct selective influence for both factors. Deadline selectively influenced rate, whereas, in line with Study 1 and Palada et al. (2016), the number of stimuli selectively influenced threshold. We examine these effects in turn.

As in Study 1, participants strategically lowered their threshold in response to a greater number of stimuli in a trial to speed their decision and thus minimize non-responses; again, this strategy was not completely effective as non-response rate increased with a greater number of ships. Participants also did not respond to the number of stimuli by changing their rate of information processing or by adopting a bias. Unlike Study 1, where time pressure conditions were blocked, and Palada et al. (2016), who probabilistically varied the number of ships in a 3-minute period, Study 2 randomized the presentation of time pressure conditions. A novel finding using this design is that participants could flexibly adjust their threshold in response to the number of stimuli with relatively little notice. Though threshold adjustments are an effortful process, participants appear to have devoted resources to strategically adjust their response caution given the time pressure demands for each trial. This highlights a good deal of flexibility in adapting to time pressure. However, the same sort of flexibility did not apply to the deadline manipulation.

Participants struggled to adapt to tighter deadlines. While responses were faster, they became less accurate as the deadline became tighter. Indeed, it appears that the time pressure in the 6s condition was sufficient to cross participants' redline, as accuracy declined drastically to the point of chance. However, this was only evident for targets. Participants were able to maintain a high degree of accuracy for non-targets under a 6s deadline, even though they were responding to non-targets as rapidly as to targets. These suggest that there is no single redline at which performance begins to deteriorate.

Our modeling provided detailed insight into the strong effect of deadline, revealing that it was not due to a non-target response bias or a change in threshold. Rather, the modelling suggests that it was a consequence of two opposing processes that differentially affected the drift rates for the matching and mismatching accumulators. On one hand, the increase in the mean rate as deadlines became tighter suggests that participants tried to compensate for the shorter deadline by applying more effort. Assuming a constant discriminability (i.e., the difference between the matching and mismatching accumulator), an increase in effort or arousal should produce an increase in the drift rate for both matching and mismatching accumulators, producing faster responses without any change in accuracy. However, the increase for the matching accumulator was weaker than for mismatching accumulator, suggesting that there was an opposing process at work. The decrease in the difference between the matching and mismatching rates as the deadline became tighter suggests that the quality of evidence that could be accumulated decreased as participants had less time available, and therefore discriminability declined. This effect was much stronger for targets than for non-targets. For non-targets, the increase in effort was sufficient to compensate for a modest reduction in quality of information, as indicated by the fact that the drift rate for the matching accumulator increased as the deadline became tighter. For targets, however, the increase in effort was not sufficient to compensate a more substantial reduction in quality of information, as indicated by the fact that the drift rate for the matching accumulator decreased as the deadline became tighter.

Finally, tighter deadlines speeded non-decision times. One possibility is that this was a result of participants attempting to adapt to the extreme difficulties experienced with responding to targets, by becoming more efficient in motor planning and effectively executing a response when less time was available. Another possibility is that they rushed their encoding process, particularly for the shortest deadline where non-decision time was very fast. This rushed encoding might, in turn, be expected to result in poorer quality evidence being available from the stimulus.

As for Study 1, we examined the effect of ship ordering (first ship vs. subsequent ships) on non-decision time by re-fitting the AIC selected model and allowing the ship order factor to influence non-decision time. The more complex model significantly improved fit over the AIC selected model, $\chi^2(180) = 18189$, $p < .001$, and retained the effect of number of ships on threshold and the effects of deadline on rates. Consistent with the additional analyses in Study 1, the non-decision time of ships presented first were significantly longer than subsequently presented ships (0.75s vs 0.15s, respectively; $\chi^2(1) = 147.06$, $p < .001$). There was also an interaction between deadline and ship order, $\chi^2(2) = 9.67$, $p = .008$. As shown in Figure 11, the decrease in non-decision time with tighter deadlines was considerably stronger for the first ship compared to other ships, though simple effects slope analyses indicated that both downward trends were significant (*p*s < .001). The general decline in non-decision times is consistent with the possibility that participants rushed encoding time with tighter deadlines. The stronger decline for ships presented first could be accounted for by the delay in the ability to encode features on the ship as it transitioned into the screen.

Our modeling suggests that participants adapted to changes in information load and deadline in different ways. Like Study 1, participants adjusted their threshold in response to the number of ships, which partially attenuated the effects of the number of ships on performance, though non-responses increased with more ships. Nonetheless, this strategy was successful in avoiding the redline. In contrast, deadline had a strong effect on the quality of information, particularly for targets. This effect was due the compromise of stimulus encoding. While participants responded to deadline by exerting effort, this response was unsuccessful, as performance declined drastically, and thus participants crossed the redline in the 6s condition.

General Discussion

Across two experiments we manipulated time pressure via stimulus complexity, the number of stimuli, and deadline. Our aim was to examine how individuals adapt to these changes in time pressure. Our review of the literature identified several ways that people can adapt to increases in time pressure: Increasing the rate of information processing, prioritizing speed over accuracy, and adopting a response bias. It is difficult to infer how people adapt to time pressure based on the analysis of behavioural measures such as accuracy or mean response time alone, because these different responses can interact in complex ways, and the effects may be obscured by changes in difficulty. We addressed this problem using the LBA (Brown & Heathcote, 2008) to model accuracy and all aspects of response time distribution simultaneously. The LBA conceptualizes the rate and quality of information processing via rate parameters for correct and incorrect responses, which reflects changes in effort and/or arousal; speed-accuracy tradeoffs via the threshold parameter, which reflects changes in response caution; and response biases via the difference in thresholds between response alternatives, which reflects a strategic preference towards a response choice.

To date, only one study has applied evidence accumulation modeling to a complex task involving a time pressure manipulation. Palada et al. (2016) found that the LBA provided a good description of accuracy and the time between decisions, and that an increase in the number of stimuli led to a decrease in response caution. However, the study had several potential shortcomings, including allowing participants to switch attention between stimuli and tasks and unbalanced heterogeneity of stimulus difficulty, and examined the effects of only one time pressure manipulation. We addressed these shortcomings using a more controlled design, so that participants had to make decisions serially, and we used a balanced manipulation of stimulus difficulty. We also examined the effects of three different time pressure factors.

We used thorough model selection techniques outlined by Donkin et al. (2011) to examine whether the same psychological mechanisms mediated the effects of different time pressure factors. We found clear support for this being not the case, with each of the three factors that we manipulated to change time pressure having a distinctive profile of effects. In line with Palada et al. (2016), in both studies, we found that the number of stimuli selectively influenced a strategic factor, the threshold amount of information required for a response, with no effect on factors related to the rate of information processing and the quality of the evidence obtained from the stimuli.

Study 1 replicated the post-hoc analyses reported by Palada et al., (2016), finding that stimulus complexity selectively influenced the quality of evidence. Complexity had no effect on the overall rate of information processing. There was also no effect of complexity on thresholds, just as would be expected in simple choice tasks with quick responses. We had speculated that, given the much slower decisions in our task, participants may have been able to adjust their thresholds in an attempt to compensate the greater difficulty of decisions with more complex stimuli, but this does not appear to have been the case. We did find a lower threshold for target than non-target responses, which was likely a strategic attempt to compensate for the greater overall difficulty of target stimuli. Once again, however, this threshold-mediated effect did not change with stimulus complexity.

In Study 2, we found that deadline influenced the rate of information processing and the quality of evidence, but had no effect on threshold. Our findings are contrary to Dambacher and Hübner (2015), who examined the effects of deadlines on decision processes in a simple choice task. Dambacher and Hübner (2015) found that threshold declined with tighter deadlines. Further, while we observed an increase in mean rate with tighter deadlines, they observed a decrease in rate. In our study, deadline had a strong effect on the quality of evidence, particularly for targets. Dambacher and Hübner (2015) used a relatively simple flanker task, with the greatest error rate being 22.5%. This decline in performance may have been insufficient to motivate participants to protect accuracy by increasing their effort. Comparing this to our study, with targets approaching near-chance performance, it appears that this decline in performance may have been sufficient to motivate participants to protect their accuracy by exerting effort in response to tighter deadlines. We suggest that participants in our study were unwilling to compound the effects of deadline on accuracy by also reducing their response threshold. The decline in rate observed in Dambacher and Hübner, which reflects degradation in the quality of evidence being extracted from the stimulus, is consistent with our observed three-way interaction on rates showing poorer discriminability as deadlines became tighter.

A potential reason for the large increase in the rate of accumulation of evidence for a non-target response under extreme time pressure when in fact the stimulus was a target is that participants failed to encode all of the ships features, and then took the corresponding lack of evidence for the presence of target features as supporting the stimulus being a non-target. Further, it is possible that target features are encoded more slowly than non-target features so that reduced encoding time is associated with an increased amount of non-target evidence relative to target evidence. This explanation is supported by the very short non-decision time estimate for the 6s condition, which was less than half the non-decision time for the 9s condition. This short encoding time is consistent with the potential for failures in the encoding process, which in turn causes the quality of target evidence to be poor. The corresponding lack of difference between the evidence accumulation rates for target and non-target accumulations is then ultimately responsible for the observed chance accuracy with target stimuli at the shortest deadline.

These findings not only provide an example of how the LBA can provide insights into the processes underlying the redline phenomenon, but also illustrates why the concept of the redline has proven so elusive (Hart & Wickens, 2010). The redline refers to the point at which demands exceed resources, and performance deteriorates. However, there are multiple ways in which individuals can respond to time pressure: Increasing effort, changing strategy, or adjusting response bias. The use of such strategy is determined by the motives and goals of the individual, and suggests that there is not a common redline for workload, as the point at which performance declines depends on how individuals respond to demands. Without an explicit model that can quantify such strategies, it is difficult to explain why performance does or does not decline as workload increases.

The LBA can also be used to identify the effectiveness, or failures, of the strategies used to protect performance in high workload situations. For example, where performance does not suffer in high workload situations, the LBA can be used to explain how performance was protected. In line with this, the LBA could also be used to predict when performance will degrade considerably in a given situation with the use of a specific cognitive strategy or processes. In our case, it appears that performance deteriorated when the deadline was decreased to 6s, because participants did not allow sufficient time to encode the stimuli, and increases in effort were not sufficient to compensate for the resulting decline in the quality of stimulus information.

Accumulate-to-threshold models suggest that there is no discrete point at which performance deteriorates. However, these models can be exploited to identify individual differences in the way that people respond to task demands, and to predict how performance will change as demands continue to change into the future. However, future work is necessary to identify whether accumulate-to-threshold models can explain performance decrements that are associated with underload, when the person has to maintain attention on task for extended periods of time (Young et al., 2015). There is promising work in similar areas; for example, McVay and Kane (2012) found that mind wandering was reflected by increases in the LBA's drift variability parameter (*sv*).

Although the work successfully extends accumulate-to-threshold models of simple, static perceptual decisions to a more complex setting (see also Palada et al., 2016), more work is required to address even more realistic and applied areas. For example, it is likely there would be a need to account for the possibility of repeated attention switching during the course of a decision (for an initial approach in a simple task see Holmes, Trueblood, & Heathcote, 2016). Nonetheless, our work illustrates the benefits of applying the LBA to understand human performance in dynamic, multi-stimulus environments. The LBA could be used in practice to identify how operators manage their workload, which has implications for system design, processes and procedures, such as identifying cases where task support is required. Similarly, the model could be used to identify how technologies and interfaces influence decision-making processes. The model could also be used for predictive purposes, such as simulating human performance under novel or unsafe conditions – as we show, this approach appears viable even in situations where the operator may be overloaded.

While it may be interesting to apply other successful accumulate-to-threshold models to complex tasks, such as the diffusion model (Ratcliff & McKoon, 2008), the accumulator class of models, particularly the LBA, show more promise in complex applied tasks because of the versatility in model architecture. For example, Strickland et al. (submitted) applied the LBA to understand the processes underlying prospective memory, and included a task with more than two possible responses; Trueblood, Brown & Heathcote (2014) extended the LBA to account for stimuli with multiple attributes; Bushmakin et al. (2017) extended the LBA to account for choices contingent on the combination of attributes; and the LBA has been extended to account for cases where the evidence changes during the course of the decision (Holmes et al., 2016). In addition to the current paper, all of these previous studies illustrate the ability of the LBA to account for complex decisions which could be extended to complex environments.

Our results also suggest a novel type of speed-accuray tradeoff, where increased speed associated with decreased non-decision time also results in decreased accuracy. These findings have broader application for evidence accumulation models, providing a link between the quality of evidence with the time spent encoding a stimulus. They suggest that evidence accumulation rates are determined not only by stimulus factors outside participants' control, but also by stategic factors related to the time, and perhaps attentional resources, devoted to extracting evidence from a stimulus that is relevant to the requried decision.

This novel type of encoding-mediated speed-accuracy tradeoff has very different applied implications from the traditional threhold-mediated mechanism, and has the potential to provide important insights into the causes of the redline phenomenon (Young et al., 2015), and hence how it can be dealt with when it occurs in practice. In the present paradigm, for example, our results suggest that if more accurate target detection were critical to successful performance, the most effective adaptive strategies would encourage participants to devote more time to encoding target features.

In summary, workload is not a direct function of task demands. Rather, the effects of task demands depend on the way that the individual adapts to those demands (Loft et al., 2007; Parasuraman & Hancock, 2001; Young et al., 2015). Such volitional and effortful behaviours are motivated by the need maintain or achieve a certain level of performance as task demands increase, and to keep workload manageable. Without the use of adaptive behaviours in response to increased demands, performance will necessarily degrade as the individual approaches the redline. Further, the effectiveness of a given response partially determines how quickly the individual will reach the redline (e.g., Sperandio, 1978). Our studies illustrate the importance of accounting for how individuals adapt to task demands, and provide evidence that they are sensitive to the distinct aspects of time pressure, using a variety of mechanisms to adapt to different factors. They also show that investigating adaptive behaviour using cognitive models can further our understanding of how individuals cope, and fail to cope, with the effects of time pressure, and provide novel insights into how to improve human performance under high and even excessive workload.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7), 1-23.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cogn Psychol, 57*(3), 153-178. doi:10.1016/j.cogpsych.2007.12.002

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling.* Los Angeles : Sage.

Bushmakin, M.A., Eidels, A. & Heathcote, A. (2017). Breaking the rules in perceptual information integration. *Cognitive Psychology*, *95*. doi: 10.1016/j.cogpsych.2017.03.001

Colle, H.A., & Reid, G.B., 2005. Estimating a mental workload redline in a simulated air-to-ground combat mission. *Int. J. Aviat. Psychol., 15* (4). doi: 10.1207/s15327108ijap1504_1

Dambacher, M., & Hübner, R. (2015). Time pressure affects the efficiency of perceptual processing in decisions under conflict. *Psychological Research, 79*(1), 83-94. doi: 10.1007/s00426-014-0542-z

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: rethinking the scaling problem. *Psychon Bull Rev, 16*(6), 1129-1135. doi:10.3758/PBR.16.6.1129

Donkin, C., Little, D. R., & Houpt, J. W. (2014). Assessing the speed--accuracy trade-off effect on the capacity of information processing. *J Exp Psychol Hum Percept Perform, 40*(3), 1183-1202. doi:10.1037/a0035947

Drury, C. G. (1994). The speed-accuracy trade-off in industry. *Ergonomics, 37*(4), 747-763.

Earle, F., Hockey, B., Earle, K., & Clough, P. (2015). Separating the effects of task load and task motivation on the effort-fatigue relationship. *Motivation and Emotion, 39*(4), 467-476. doi:10.1007/s11031-015-9481-2

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychon Bull Rev, 17*(6), 763-771. doi:10.3758/PBR.17.6.763

Förster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes, 90*(1), 148-164. doi:10.1016/S0749-5978(02)00509-5

Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal, 5*(1), 39-52.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology, 62*, 451-482.

Hart, S. G., & C. D. Wickens (2010). Cognitive Workload. In NASA Human Integration Design Handbook (HIDH). Washington, DC: NASA. Retrieved from: www.nasa.gov/sites/default/files/atoms/files/human_integration_design_handbook_revision_1.pdf

Heathcote, A., & Love, J. (2012). Linear Deterministic Accumulator Models of Simple Choice. *Frontiers in Psychology, 3*, 292. doi:10.3389/fpsyg.2012.00292

Heathcote, A., & Hayes, B. (2012). Diffusion versus linear ballistic accumulation: different models for response time with different conclusions about psychological mechanisms? *Can J Exp Psychol, 66*(2), 125-136. doi:10.1037/a0028189

Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information- processing load. *Human Factors, 39*(1), 30-47. doi:10.1518/001872097778940597

Hockey, R. (2013). *The psychology of fatigue work, effort, and control / Robert Hockey*: Cambridge : Cambridge University Press.

Holmes, W. R., Trueblood, J. S., Heathcote, A. (2016). A new framework for modeling decisions about changing information: The Piecewise Linear Ballistic Accumulator model. *Cognitive Psychology*, 85, 1–29. doi: 10.1016/j.cogpsych.2015.11.002

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review*, *91*(2), 153-184. doi: 10.1037/0033-295X.91.2.153

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of experimental psychology. General, 139*(4), 665. doi:10.1037/a0020198

Lewandowsky, S., & Farrell, S. (2010). *Computational Modeling in Cognition: Principles and Practice*. Thousand Oaks, CA: Sage.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. *J Exp Psychol Appl, 15*(2), 106-124. doi:10.1037/a0016118

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: critical review and broader implications. *Hum Factors, 49*(3), 376-399.

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “d'oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 525–549. http://doi.org/10.1037/a0025896

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*(3), 214-255. doi:10.1037/0033-295X.86.3.214

Neal, A., Ballard, T., & Vancouver, J. B. (2017). Dynamic Self-Regulation and Multiple-Goal Pursuit. *Annual Review of Organizational Psychology and Organizational Behavior, 4*(1).

Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review, 83*(6), 508-510. doi:10.1037/0033-295X.83.6.508

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, *4*(2), 61–64. doi 10.20982/tqmp.04.2.p061

Palada, H., Neal, A., Vuckovic, A., Martin, R., Samuels, K., & Heathcote, A. (2016). Evidence accumulation in a complex task: Making choices about concurrent multiattribute stimuli under time pressure. *J Exp Psychol Appl, 22*(1), 1-23. doi:10.1037/xap0000074

Pachella, R. G., & Pew, R. W. (1968). Speed-Accuracy Tradeoff in Reaction Time: Effect of Discrete Criterion Times. *Journal of Experimental Psychology, 76*(1), 19-24. doi:10.1037/h0021275

Parasuraman, R., & P. A. Hancock (2001). Adaptive Control of Mental Workload. In P. A. Hancock, and P. A. Desmond (Eds.), *Stress, Workload and Fatigue*, 305–320. Mahwah, NJ: Lawrence Erlbaum.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160. doi:10.1518/155534308x284417

Raby, M., & Wickens, C. D. (1994). Strategic Workload Management and Decision Biases in Aviation. *The International Journal of Aviation Psychology, 4*(3), 211-240. doi:10.1207/s15327108ijap0403_2

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. *J Exp Psychol Learn Mem Cogn, 40*(5), 1226-1243. doi:10.1037/a0036801

Sperandio, J. C. (1978). The regulation of working methods as a function of work-load among air traffic controllers. *Ergonomics, 21*(3), 195-202. doi:10.1080/00140137808931713

Standage, D., Wang, D. H., Heitz, R. P., & Simen, P. (2015). Toward a unified view of the speed-accuracy trade-off. *Front Neurosci, 9*, 139. doi:10.3389/fnins.2015.00139

Strickland, L., Heathcote, A., Remington, R. W., & Loft, S. (2017). Accumulating evidence about what prospective memory costs actually reveal. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *43*(1), 1616-1629. doi: 10.1037/xlm0000400

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2017). Racing to Remember: A Theory of Decision Control in Event-Based Prospective Memory. Manuscript submitted for publication.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal Properties of Elementary Perception: An Investigation of Parallel, Serial, and Coactive Theories. *Journal of Mathematical Psychology, 39*(4), 321-359. doi:10.1006/jmps.1995.1033

Townsend, J. T., & Altieri, N. (2012). An accuracy-response time capacity assessment function that measures performance against standard parallel predictions. *Psychol Rev, 119*(3), 500-516. doi:10.1037/a0028448

Townsend, J. T., & Eidels, A. (2011). Workload capacity spaces: a unified methodology for response time measures of efficiency as workload is varied. *Psychon Bull Rev, 18*(4), 659-681. doi:10.3758/s13423-011-0106-9

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. Psychological Review, 121, 179–205. http://dx.doi.org/10.1037/a0036137

Vidulich, M. A. and C. D. Wickens (1986). Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments. *Applied Ergonomics*, *17*(4), 291-296. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/15676597

Vidulich, M. A., & Tsang, P. S. (2012). Mental Workload and Situation Awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (Vol. 4th Ed., pp. 243-273). Hoboken, NJ: John Wiley & Sons.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics, 58*(1), 1-17.

Zijlstra, F. R. H. (1993*). Efficiency in work behaviour: A design approach for modern tools.* Delft, The Netherlands: Delft University Press.

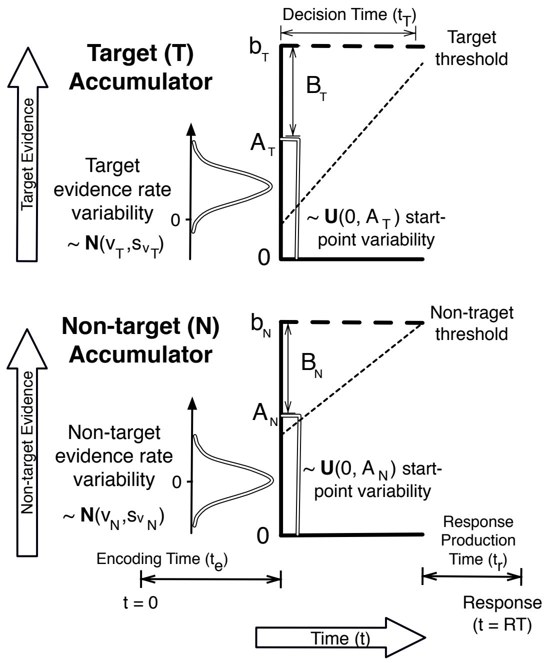

Figure 1. The standard LBA model applied to a target-detection task.

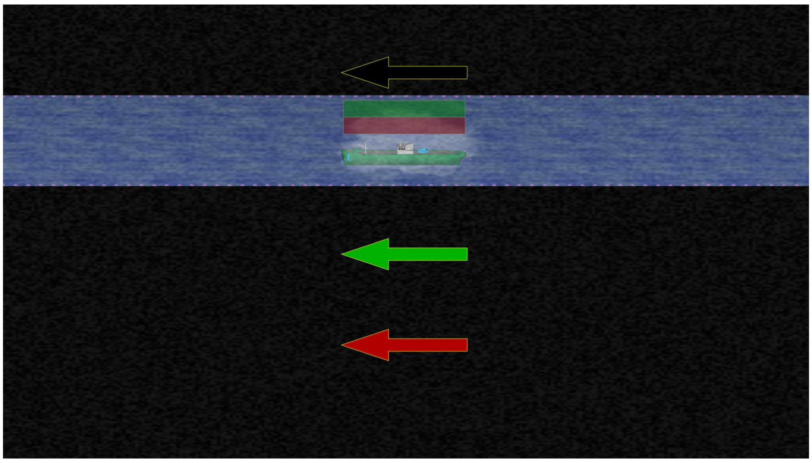

Figure 2. Screenshot of UAV task, with a non-target (red arrow) and target (green arrow) responses, a ship requiring classification in the unmasked lane, and a masked unclassified ship indicated with an outlined, unfilled arrow. *Note.* Noise and fog overlay are reduced for illustrative purposes.

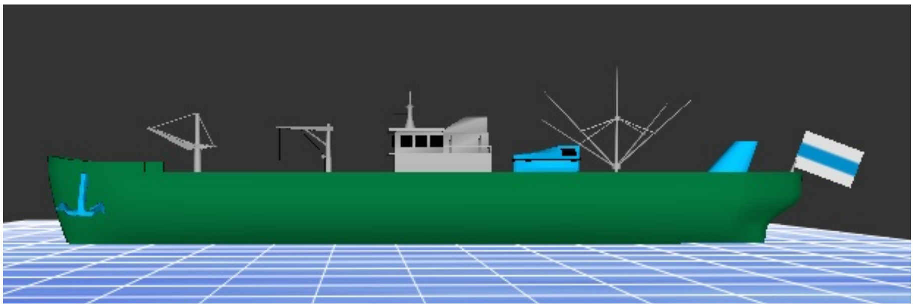

Figure 3. An example target ship with seven features. *Note.* The superstructure in the centre of the deck was present on all ships.

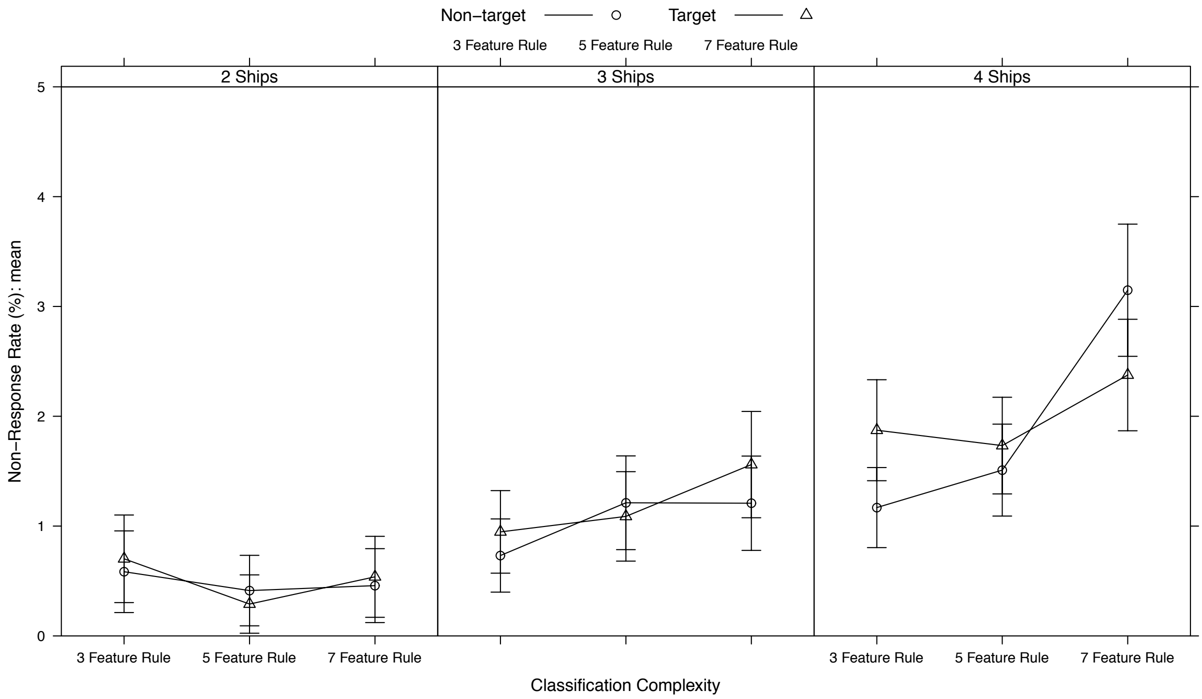

Figure 4. Study 1: Interaction between stimulus, difficulty, and ships in trial on non-response rates, with 95% within-subject confidence intervals (Morey, 2008).

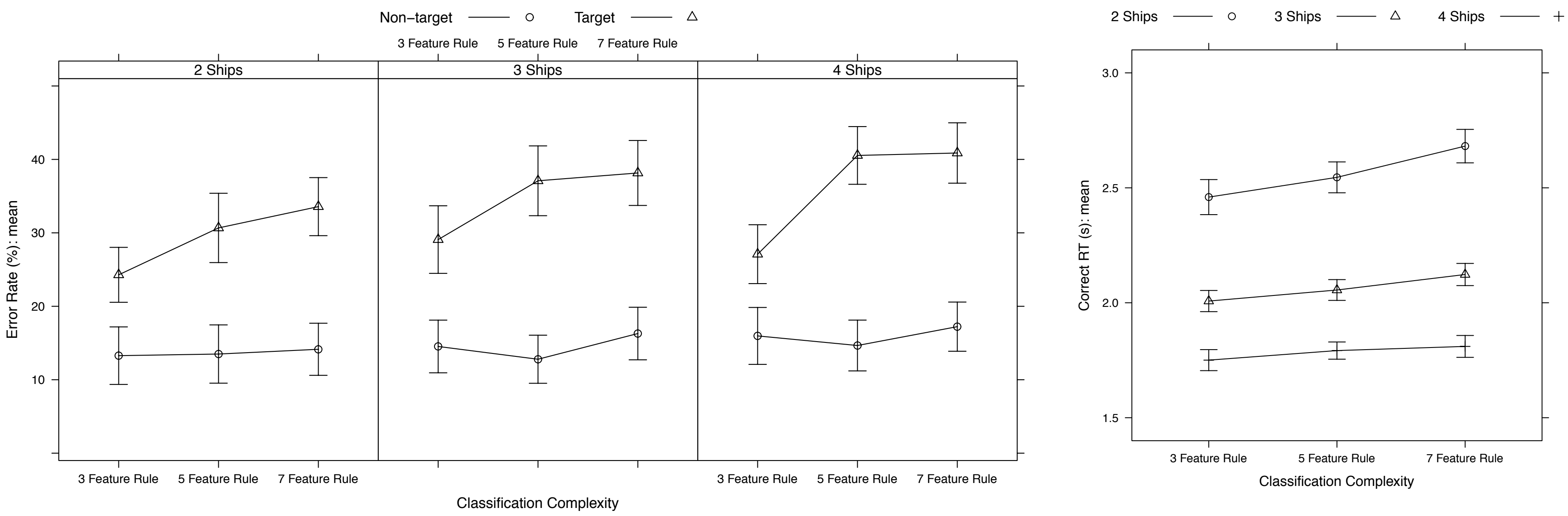

Figure 5. Study 1: The three-way interaction between stimulus type, complexity and number of stimuli on errors rates (left panel) and the two-way interaction excluding stimulus type on mean correct response time (right panel), with 95% within-subject confidence intervals (Morey, 2008).

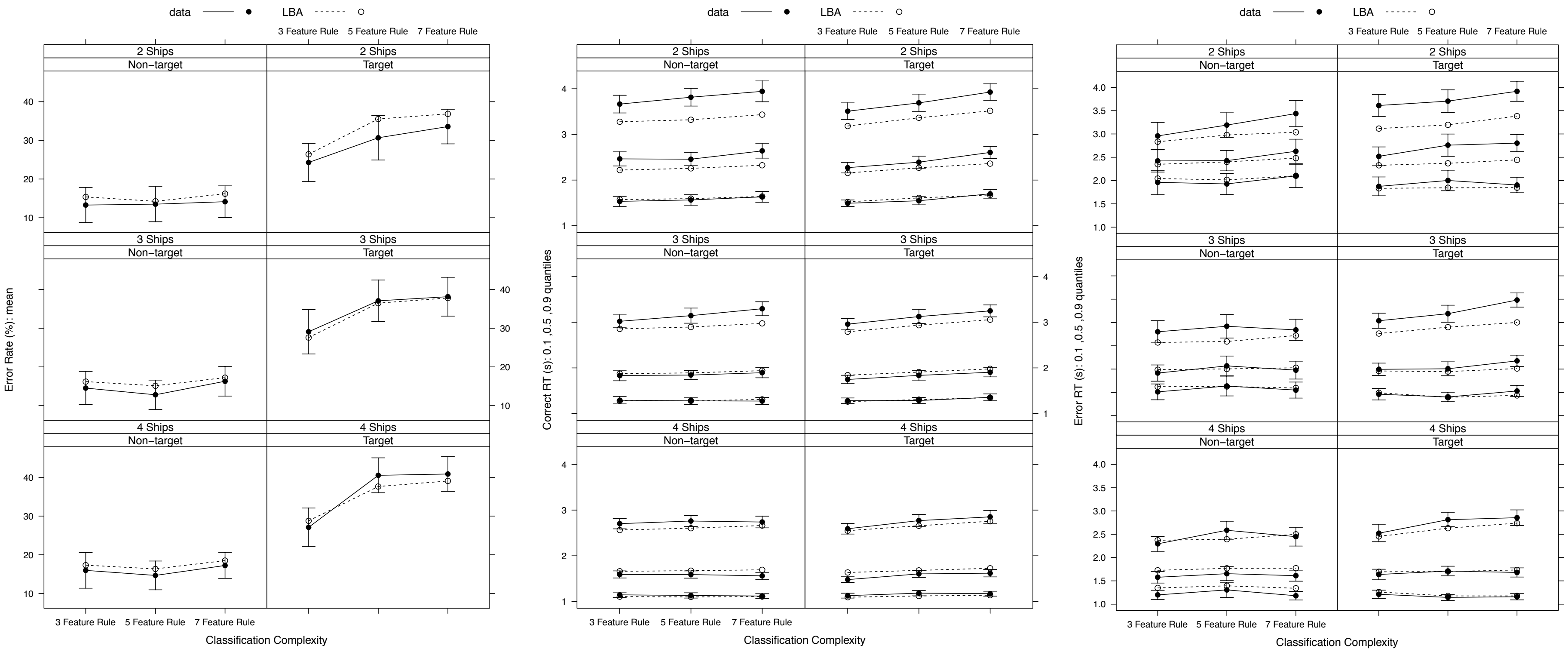

Figure 6. Study 1: Fits of the AIC-selected LBA model to error rates, correct, and error response time distribution data with 95% confidence intervals.

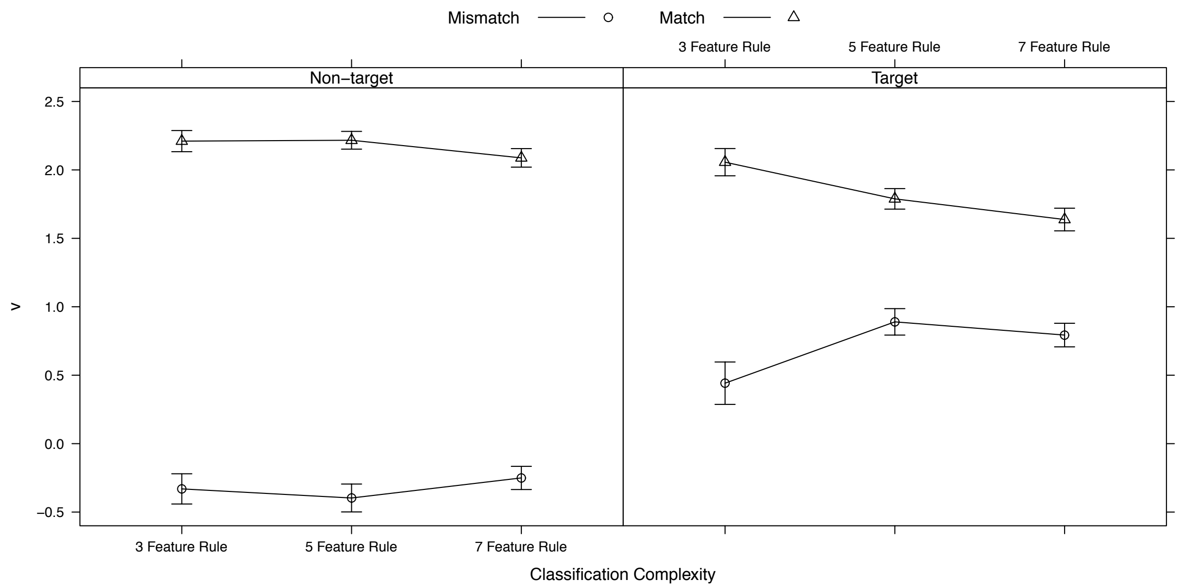

Figure 7. Study 1: Mean rate estimates for the AIC selected model, with within-subject standard errors (Morey, 2008).

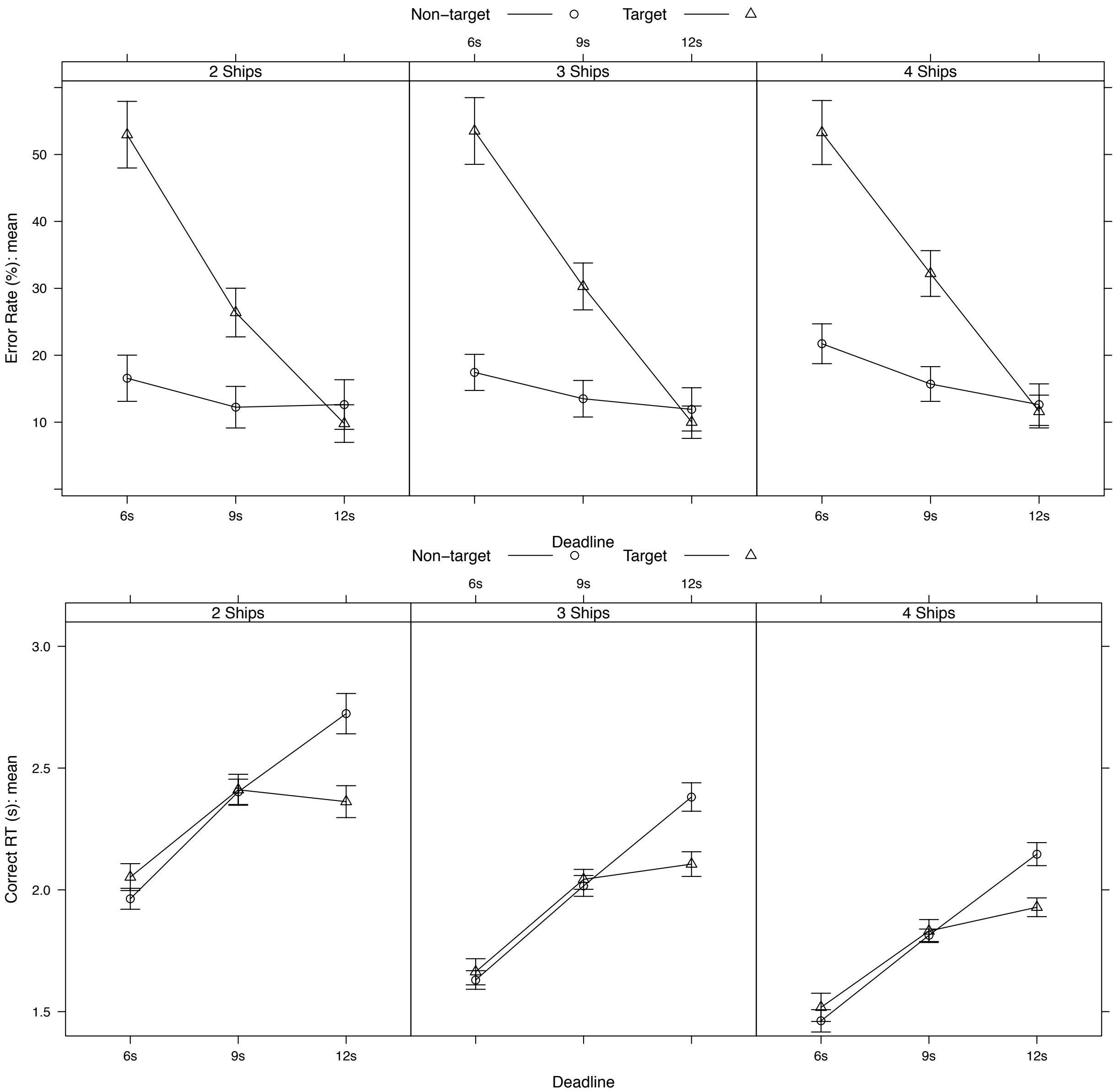

Figure 8. Study 2: Plots of the joint effects of stimulus type, deadline and number of stimuli on error rates (top row) and correct response time (bottom row). Error bars show 95% within-subject confidence intervals (Morey, 2008).

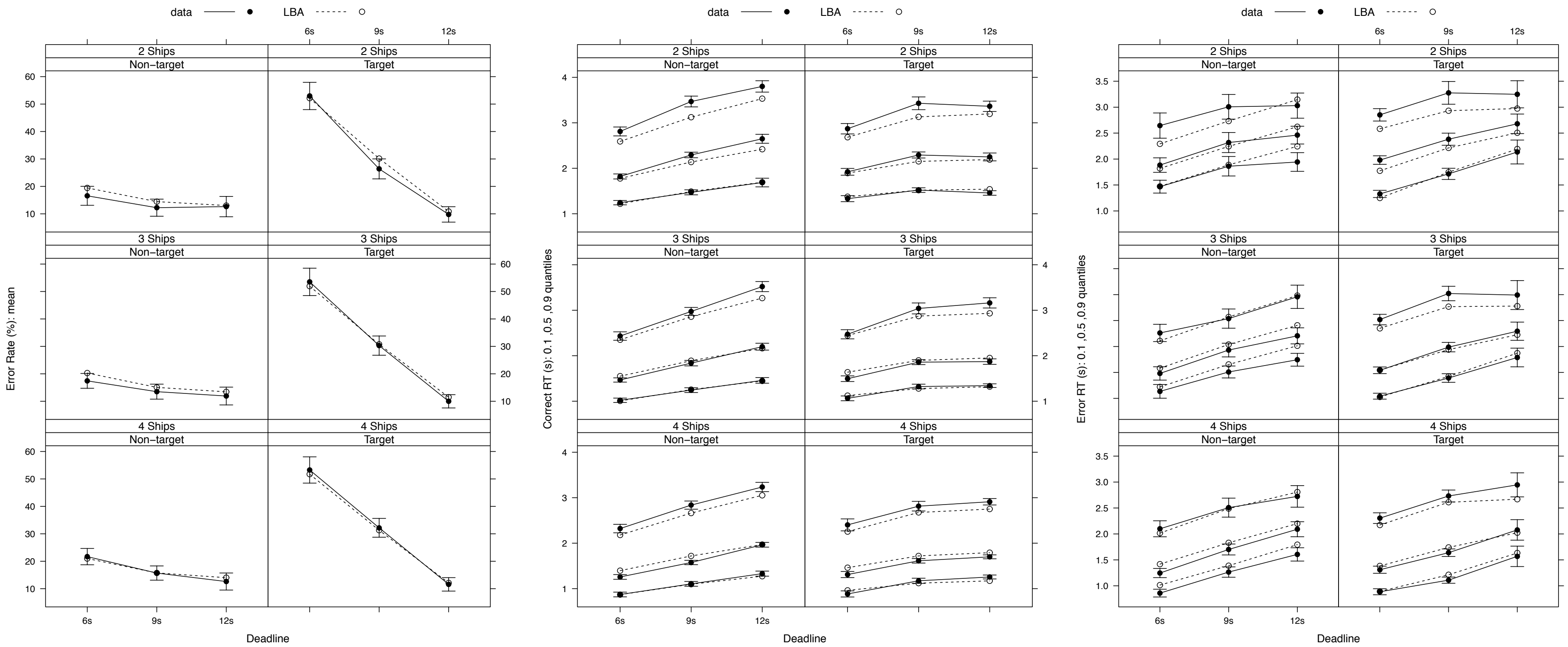

Figure 9. Study 2: Fits of the AIC-selected LBA model to error rates, correct, and error response time distribution data with 95% confidence intervals.

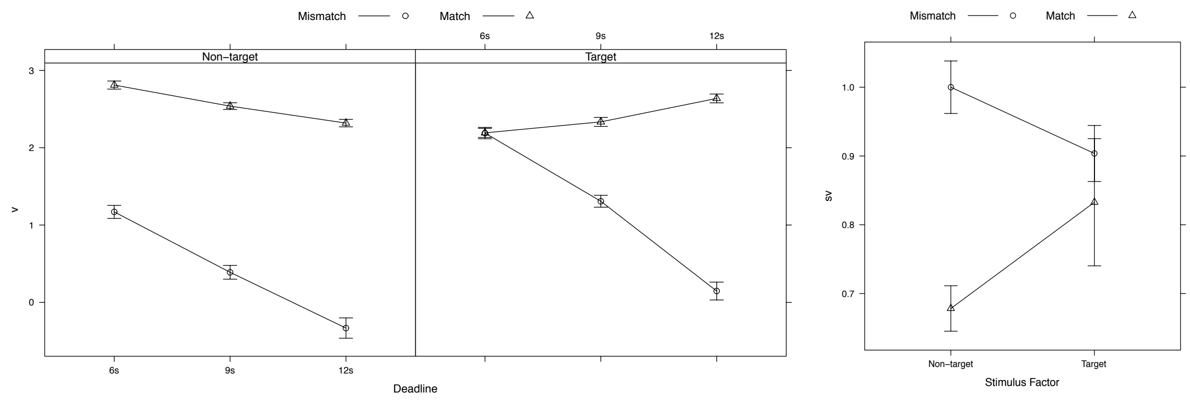

Figure 10. Study 2: Mean rate (left graph) and rate variability (right graph) estimates for the AIC selected models, with within-subject standard errors (Morey, 2008).

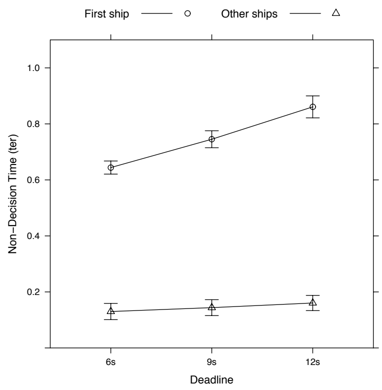

Figure 11. Study 2: Non-decision time estimates for the additional model fit, having the parametrization of the AIC selected model, though allowing non-decision time to also be influenced by ship order factor, with within-subject standard errors (Morey, 2008).

Table 1.
LBA top model parameterization for Study 1 and Study 2, and the number of parameters (NP). Experimental factors are stimulus (S), ships in trial (NS), complexity (C) and deadline (D). Accumulator factors are match (M) and response (R). Total refers the total number of parameters in the model.

		LBA Parameters					
		Start point variability (A)	Threshold (B)	Rate (v)	Rate variability (s_v)	Non-decision time (t_{er})	Total NP
Study 1	Factors	-	C, NS, R	C, NS, M, S	M, S	C, NS	
	NF	1	18	36	3	9	67
Study 2	Factors	-	D, NS, R	D, NS, M, S	M, S	D, NS	
	NF	1	18	36	3	9	67

Table 2.
Study 1 analysis of variance results for conventional moment-based performance measures.

		Error Rate		Mean Correct RT		Mean Error RT		SD Correct RT		SD Error RT	
Effect	*df*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
S	1	80.78	<.001	5.70	.017	2.45	.117	1.61	.204	0.95	.331
NS	2	18.14	<.001	440.46	<.001	67.99	<.001	13.59	.001	9.73	.008
C	2	5.11	.078	20.41	<.001	1.96	.375	5.04	.081	5.93	.052
S·NS	2	5.69	.058	2.79	.248	0.46	.795	1.94	.379	1.10	.576
S·C	2	1.63	.442	4.91	.086	0.73	.696	3.04	.219	4.20	.123
NS·C	4	4.39	.356	13.20	.010	5.38	.250	3.11	.540	7.32	.120
S·NS·C	4	14.09	.007	3.06	.548	5.34	.254	3.50	.478	5.21	.267

Note. S=stimulus type, NS=number of stimuli in trial, C=complexity, RT=response time. Ship configuration has been omitted from the table since it served as a control variable.

Table 3.

Study 1 LBA fits summed over participants. AIC is the Akaike Information Criterion (AIC), p is the number of parameters per participant. Model factors are number of stimuli in trial (NS), complexity (C), stimulus (S), match (M), and response (R).

	LBA Parameters				Model Fit		
Variant	Threshold (B)	Rate (v)	Rate variability (s_v)	Non-decision time (t_{er})	p	Deviance	AIC
Top	C, NS, R	C, NS, M, S	M, S	C, NS	3819	129126	136764
AIC	NS, R	C, M, S	M, S	C, NS	1767	129885	136511

Table 4.
Study 2 analysis of variance results for conventional moment-based performance measures.

		Error Rate		Mean Correct RT		Mean Error RT		SD Correct RT		SD Error RT	
Effect	*df*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
S	1	339.24	<.001	2.67	.102	0.20	.655	1.60	.206	3.76	.053
NS	2	17.51	<.001	224.69	<.001	98.13	<.001	2.19	.334	14.71	.001
D	2	7.99	.018	411.56	<.001	37.26	<.001	31.41	<.001	0.05	.976
S·NS	2	7.03	.030	1.65	.438	3.45	.178	0.95	.621	5.01	.082
S·D	2	227.89	<.001	86.02	<.001	0.49	.783	5.67	.059	6.07	.048
NS·D	4	10.69	.030	7.95	.093	11.93	.018	2.86	.581	8.94	.063
S·NS·D	4	11.23	.024	15.43	.004	2.05	.726	0.39	.983	6.40	.171

Note. S=stimulus type, NS=number of stimuli in trial, D=deadline, RT=response time. Ship configuration has been omitted from the table since it served as a control variable.

Table 5.

Study 2 LBA fits summed over participants. AIC is the Akaike Information Criterion (AIC), p is the number of parameters per participant. Model factors are number of stimuli in trial (NS), deadline (D), stimulus (S), match (M), and response (R).

	LBA Parameters				Model Fit		
Variant	Threshold (B)	Rate (v)	Rate variability (s_v)	Non-decision time (t_{er})	p	Deviance	AIC
Top	D, NS, R	D, NS, M, S	M, S	D, NS	4020	127103	135143
AIC	NS	D, M, S	M, S	D	1320	127921	130561