2020 ■ Vol. 16 ■ no. 2

An LBA account of decisions in the multiple object tracking task

Reilly J. Innes [a] and Caroline L. Kuhne [a]

[a]School of Psychology, University of Newcastle, Australia

Abstract ■ Decision making is a vital aspect of our everyday functioning, from simple perceptual demands to more complex and meaningful decisions. The strategy adopted to make such decisions is often viewed as balancing elements of speed and caution, i.e. making fast or careful decisions. Using sequential sampling models to analyse decision making data can allow us to tease apart strategic differences, such as being more or less cautious, from processing differences, which would otherwise be indistinguishable in behavioural data. Our study used a multiple object tracking task where student participants and a highly skilled military group were compared on their ability to track several items at once. Using a mathematical model of decision making (the linear ballistic accumulator), we show the underpinnings of how two groups differ in performance. Results showed a large difference between the groups on accuracy, with the Royal Australian Air Force (RAAF) group outperforming students. An interaction effect was observed between groups and level of difficulty in response times, where RAAF response times slowed at a greater rate than the student group as difficulty increased. Model results indicated that the RAAF personnel were more cautious in their decisions than students, and had faster processing in some conditions. Our study shows the strength of sequential sampling models, as well as providing a first attempt at fitting a sequential sampling model to data from a multiple object tracking task.

Keywords ■ Cognitive modelling ; Multiple object tracking ; Decision modelling; Linear Ballistic Accumulator.

reilly.innes@uon.edu.au

10.20982/tqmp.16.2.p175

Acting Editor ■ Cheng-Ta Yang (Department of Psychology, National Cheng Kung University)

Reviewers ■ Ingmar Visser (University of Amsterdam)

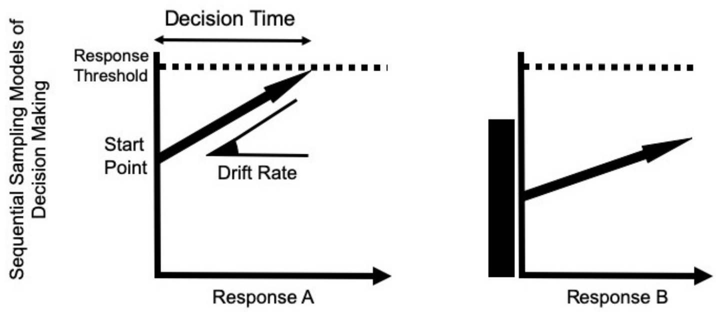
■ and two more.

Introduction

The practice of cognitive psychometrics has contributed greatly to our understanding of human decision making (Batchelder, 2010). Sequential sampling models are a prime example of this discipline, in which latent psychological processes are represented by parameters of the quantitative model (Donkin, Averell, Brown, & Heathcote, 2009). How these parameters change in different decision making contexts allows inferences to be made about the influence of specific cognitive processes across tasks and groups. For instance, these models have been able to show parameter shifts in simple perceptual tasks (Brown & Heathcote, 2008; Ratcliff, Gomez, & McKoon, 2004), as well as in more complex tasks (e.g., Hawkins et al., 2014; Ho et al., 2014). Parameter shifts include the strategies undertaken by participants, such as being more fast or cautious in responding (Rae, Heathcote, Donkin, Averell, & Brown, 2014; Evans, Rae, Bushmakin, Rubin, & Brown, 2017), the speed in which different groups process information (Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Ratcliff, Thapar, & McKoon, 2010), or a combination of these factors. Further examples of sequential sampling models have accounted for a range of decision-making data across age groups (Forstmann et al., 2011), personality groups (Evans et al., 2017), clinical groups (e.g., depression and schizophrenia; Dillon et al., 2015; Heathcote et al., 2015), and groups differing in blood alcohol levels (van Ravenzwaaij, Dutilh, & Wagenmakers, 2012).

Sequential sampling models posit that before a decision is made, evidence in favour of competing responses is sequentially sampled from the environment until a decision criterion (known as the threshold) is reached, triggering a decision response. When the threshold is set high, more

Figure 1 ■ The LBA model of decision-making which explains the trade off between speed and caution. A basic framework for a simple two choice paradigm is shown. The two accumulators "race" by accumulating evidence for either decision before reaching the response threshold. The first accumulator to reach the threshold triggers the associated decision.



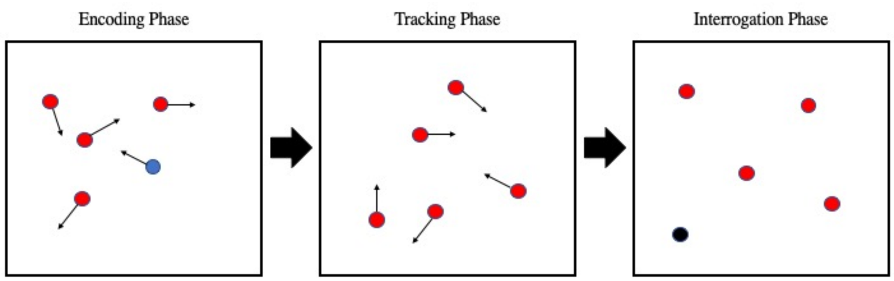
evidence must be sampled from the environment, resulting in slower responses. When it is set low, the decision will be quicker. These models also account for the time taken to perceive the information, the time taken to make a response and any bias an individual may have towards a particular choice. There are a variety of decision making models, however, one of the most commonly used is the linear ballistic accumulator model (LBA), illustrated in Figure 1. The LBA simplifies the processes of sampling by approximating the sampling as linear, where two separate accumulators represent separate response choices.

We often observe differences and interactions in response times (and accuracy) between experimental conditions and groups, however, we do not always have insight into what may lead to these differences. Sequential sampling models can assist in explaining underlying psychological processes through parameter estimation, and further, indicate that different individuals may use different decision strategies. The key success of sequential sampling models is that they allow us to analyse both response time and accuracy data together, as well as using the entirety of the response time distribution. This allows us to predict the precise quantitative relationship between speed and accuracy whilst also specifying how much a person's error rate varies as a function of millisecond changes in response time (Heitz & Schall, 2012). However, the greatest benefit of these models is that they allow for the comparison of psychological theories about the use of decision strategies in different contexts and between distinct groups. Hypotheses about how specific cognitive processes should work can be translated into specific predictions about how the parameters of the model will be affected by different experimental manipulations. We can then directly compare alternate models with differing parameterisations, as we would directly compare hypotheses. As outlined above, it is evident that different groups may show varying response times and accuracy, however, there are benefits in considering the multiple components that contribute to the decision process and explain these response time and accuracy differences (Starns & Ratcliff, 2010; Forstmann et al., 2011; Heathcote et al., 2015). These components can include the strategy someone chooses to take (for example sacrificing accuracy for speed), varying processing speeds between individuals or certain biases towards particular responses.

Due to the benefits of sequential sampling models, this mathematical technique has been applied to a range of decision making tasks. These tasks include standard cognitive paradigms such as go/no-go tasks (Gomez, Ratcliff, & Perea, 2007), absolute identification tasks (Brown, Marley, Donkin, & Heathcote, 2008), as well as more complex tasks such as discrete choice experiments (Hawkins et al., 2014), recognition memory paradigms (Ratcliff & Starns, 2009; Ratcliff, Thapar, & McKoon, 2011; Rae et al., 2014; Osth, Jansson, Dennis, & Heathcote, 2018) and unmanned aerial vehicle operation (Palada et al., 2016). One task which has been broadly used to model perceptual processes is the multiple object tracking (MOT) task, however, there has not yet been an attempt to understand the underlying decision process.

The MOT task involves participants tracking a number of objects within a display of distractor objects (tracking period) before deciding which objects were the targets. In this decision (interrogation) phase, they are either asked to identify the targets themselves, or they are asked sequentially whether a certain object was originally a target. An example of this task can be seen in Figure 2. Pylyshyn and Storm (1988) originally proposed the experiment to test how many items could be tracked simultaneously. Accuracy is the main dependent variable in the MOT, with accu-

Figure 2 ■ Illustration of the MOT task. The target dot is shown in blue during the encoding phase. All of the dots move randomly about the display. After a short amount of time, the target changes colour to be the same as the distractors. In the tracking phase, dots move around randomly (and can cross over) for an extended amount of time. Finally, in the interrogation phase, motion of the dots stops. Participants are then asked to sequentially classify highlighted dots (shown in black) as targets or non-targets.



racy generally decreasing as the number of objects to track increases. This is to be expected, as researchers observe that small subsets of objects are easily tracked over the course of each trial (generally around 15 seconds of tracking), whereas larger subsets of targets may lead to the loss of tracking among the distractor items (Pylyshyn & Storm, 1988).

Pylyshyn and Storm (1988) initially found that people could track up to eight items simultaneously, with the MOT task providing a new method to study object based attention. Cavanagh and Alvarez (2005) note the importance of the Pylyshyn and Storm (1988) study as a means of providing evidence against many attentional theories at the time, which posited a single focus view. In the MOT, individuals did not have to directly look at each individual stimulus, but rather, could track the motion of several items whilst focusing on a central reference point. Further studies of the MOT have focused on the task purely from a perceptual standpoint, opting to analyze eye-tracking data and proposing models of possible mechanisms used to track multiple objects (Scholl & Pylyshyn, 1999; Cavanagh & Alvarez, 2005; Drew, McCollough, Horowitz, & Vogel, 2009). These mechanisms can be thought of as either 'blob' tracking (where targets can merge and split; Haritaoglu, Harwood, & Davis, 1998), be viewed in a minimum amount of separation framework (Rasmussen & Hager, 1998), or by incorporating 3D geometric elements to distinguish between targets (Koller, Weber, & Malik, 1994). These studies present strong arguments and robust models of tracking, however, the decision making element of the task is yet to be formally investigated or modelled.

Relying on MOT accuracy alone may be misleading because it gives no insight into latent processes we know are involved in decision making. For example, some individuals may be more cautious when responding, leading to slower response times but increases in accuracy. Applying a sequential sampling model to MOT decision processes, such as the LBA, could serve to identify underlying strategic differences between individuals as well as inherent differences in ability. Factors such as the total number of objects an individual can track, tracking ability and decision caution may vary between individuals, impacting results in obscure ways. Using a sequential sampling model would allow us to move beyond assessing just accuracy, by allowing us to capture the shape of the response time distribution across both correct and error responses in the interrogation phase of the MOT.

Similar principles that have been used to model decisions in recognition memory tasks with the LBA and Diffusion models (e.g., Ratcliff & Starns, 2009; Ratcliff et al., 2011; Dube, Starns, Rotello, & Ratcliff, 2012; Rae et al., 2014; Ratcliff & McKoon, 2015; Aschenbrenner, Balota, Gordon, Ratcliff, & Morris, 2016; Osth, Bora, Dennis, & Heathcote, 2017; Osth et al., 2018) could be used to model decision processes in the MOT. The MOT decision phase, which follows the tracking phase, could be modelled by evidence accumulation, analogous to how the recognition phase in a memory task, which occurs after the study phase, is modelled by evidence accumulation. Similar to modelling decisions made in recognition memory experiments, we assume that the decision process only commences when the stimuli are directly interrogated, with participants sampling evidence from an internal representation. The MOT is a complex task which involves multiple components and, much like recognition memory modelling research, we do not attempt to model the entire task. Instead, the LBA can be used to gain useful insights into strategies and processes occurring during the decision making component of the

MOT.

With sequential sampling methods often enabling a deeper understanding of the decision making process and underlying strategy, we proposed an experiment which tested two divergent groups on varying levels of MOT. A large proportion of modelling studies that compare across groups have tended to focus on cognitive deficits associated with certain groups (e.g., aging populations; Forstmann et al., 2011; , schizophrenia; Heathcote et al., 2015; and depression; Dillon et al., 2015) rather than a "proficiency". However, our study aimed to compare a student control group to a group of highly trained Royal Australian Air Force (RAAF) personnel as a means of understanding potential differences in cognitive decision making processes. The RAAF personnel were recruited as part of a selection program for a role which required candidates to cope with significant mental demands over an extended period of time, such as war zone air traffic control. The tracking and workload component of the MOT task therefore appears to resemble a situation in which combat controllers should perform well. To be eligible for the role, candidates must have skills in reconnaissance, assault zone control, clearance of an airfield, meteorology observation, military tactics and first completed a battery of cognitive and physical tests. Due to the sensitive nature of this position, this battery of tests is kept confidential.

We proposed a simple MOT task with three levels of difficulty (indexed by the number of objects to track). We compared results across two distinct groups; psychology undergraduate students and RAAF personnel – combat controller program candidates. The RAAF group completed the task as part of a training and selection program. Selected candidates for the program are given intense specialised training over an 18 month period. Further, combat controller personnel are highly sought after, and the position is lucrative within military contexts. Consequently, we expected the RAAF group to value accuracy over speed and be more cautious, as errors may have a negative impact on their selection. Due to the element of potential selection, we expected that this would increase motivation in the RAAF group (comparatively to the student group who had no incentive to perform), and further, due to this motivation to perform well, it was possible that the RAAF group would pay closer attention to the task, which could lead to an overall higher drift rate (Smith & Ratcliff, 2009; Nunez, Vandekerckhove, & Srinivasan, 2017). Therefore, we hypothesized that the RAAF group would have higher accuracy and lower response time in the MOT compared to students. Finally, we hypothesized that divergent decision strategies would emerge in the parameters underpinning this behavioural data, with the RAAF group setting higher thresholds and displaying higher drift rates across difficulty conditions.

Method

Participants

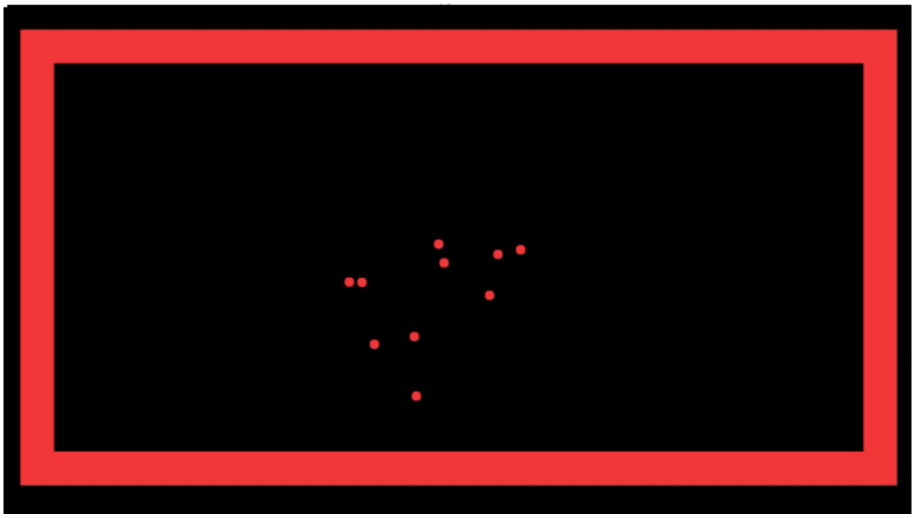
Two groups completed the task. The RAAF group was comprised of 39 Royal Australian Air Force personnel (all of whom were male) who were selected to train for a combat controller course. The RAAF group was tested as part of the combat controller course selection procedures. The student group was comprised of two sub-groups who completed alternate testing sessions; an online session which consisted of 64 University of Newcastle undergraduate students who completed the task in their own time online; and an in lab session which consisted of 28 undergraduate students from the University of Newcastle who completed the task at the same time at the University campus. Three student participants were excluded from the analysis due to computer-based errors in data.

Tasks

The MOT was displayed on a computer in front of the seated participants. All participants completed the same task. The MOT required participants to track multiple moving objects (coloured dots) within a circular display of 150 pixels (corresponding to a visual angle of 2°) for a short period of time (15s). The task had three phases; the encoding phase, the tracking phase and the interrogation phase (shown in Figure 2). In the encoding phase the target dots were blue, and all other dots were red. After three seconds of movement, the target dots turned red and the tracking phase began. All dots moved within the display area at a frame rate of 15 frames per second for the duration of a trial and participants were required to track the movement of the indicated target dots. In the interrogation phase, movement stopped and five dots were highlighted at random (one at a time). For each of those five dots, the participants were required to indicate whether each had been a tracked target, using the keyboard (either the "P" and "O" keys, or the "Q" and "W" keys, depending on handedness). Participants were given feedback following the completion of the test phase, for example, "Good work! For this trial you identified 3 out of 5 dots correctly".

There were three levels of difficulty in the MOT; 0, 1 or 4 dots to track. This was manipulated within-subjects. Across all difficulty levels, there was a total of ten dots on screen (comprised of the number of targets and distractor objects). The dots were circular, with a diameter of 14 pixels, and were restricted by the display area so that they were always in fovea. The dots' motion did not follow any uniform direction as it was randomly sampled across each frame, but could only change direction by up to 15° during

Figure 3 ■ Illustration of the concurrent display of the MOT (red dots) and DRT (red border) stimuli. The DRT component flashed on intermittently and the red dots moved around the display area randomly throughout a trial. The display area was a 150 pixel circle which did not overlap with the DRT stimulus.



each frame. This meant that the dots motion was somewhat autocorrelated from one frame to the next, however, the path of the object could not be anticipated. The dots could spatially overlap. If the motion of a dot was about to take it off the edge of the display, it was reflected, so that dots appear to bounce off the sides of the display.

A detection response task (DRT) was also carried out during the tracking phase of the MOT. For the purposes of the current study, the DRT data was not analyzed. DRT implementation generally adhered to ISO 17488 (2016) guidelines. The DRT only occurred during the tracking phase of the MOT, where a red frame appeared around the display, as shown in Figure 3. Participants were asked to respond to each elicitation of the stimulus using the keyboard (either “T” or “Y” keys, depending on handedness). The red frame stayed on screen for 1 second or until the participant responded to it (whichever occurred first). Time between successive DRT stimuli was randomly distributed between 2-4 seconds.

Procedure

Participants completed the task on a computer, with results recorded online. Participants were given instructions on screen which first introduced the DRT procedure and then the MOT procedure.

Participants completed a practice block followed by nine test phase blocks. Each block consisted of ten trials of the MOT, with the exception of the initial practice block which only consisted of three trials. Difficulty was presented sequentially in each block, with the sequential order randomised between subjects. Within each test block, all of the trials used the same number of dots to be tracked: either 0, 1 or 4. In the practice block, participants tracked 2 dots. Each of these three levels of MOT were used for three blocks, giving a total of 30 MOT trials for each difficulty (number of dots to track). Within each trial, participants made five decisions for the MOT task, giving a total of 150 decisions per difficulty. Participants were given breaks between blocks, and the total time taken to complete the experiment was between 1-1.5 hours. Response time and accuracy were recorded.

Analysis

Analysis for the experiment included inferential Bayesian analysis of differences between conditions and groups for accuracy and response time in the MOT task. We used JASP to conduct the Bayesian ANOVAs and t-tests (JASP Team, 2019). Following this, the LBA was separately fit to the decision making data of participants in each group to separate the between-groups performance differences from strategy differences. In the current task, the LBA model assumes that each decision is a race between evidence accumulators for responding “non-target” or “target” to each interrogated dot. Each accumulator gathers evidence until a threshold is hit, whereby the participants make that decision. The sequential sampling process is simple and linear with several assumptions for each decision being made. The sequential sampling process has a rate, or speed of processing, which varies randomly from trial to trial according to a normal distribution, since factors such as arousal and attention are assumed to fluctuate across trials (Donkin & Brown, 2018). Similarly, each accumulator starts with an amount of evidence (a starting point), which, on each trial, is a random value drawn from a uniform distribution.

Despite the participants being exposed to the stimulus in the tracking phase, they are not required to make a decision until the interrogation phase. Therefore we view the interrogation phase as the decision phase of the experiment, with five decisions per trial of the MOT. In modelling the decision strategies and processes of the decision phase of the MOT task, we assume that as soon as a single dot is cued for interrogation, the evidence accumulation process commences for the decision of whether to respond "target" or "non-target", with participants sampling evidence from an internal representation. This is the same principle used to model decisions in recognition memory research, where participants are first exposed to the stimuli and at a later point asked to make a decision regarding these stimuli (e.g., Ratcliff, 1978; Ratcliff & Starns, 2009; Ratcliff & Van Dongen, 2011; Dube et al., 2012; Rae et al., 2014; Aschenbrenner et al., 2016; Osth et al., 2017; Osth et al., 2018).

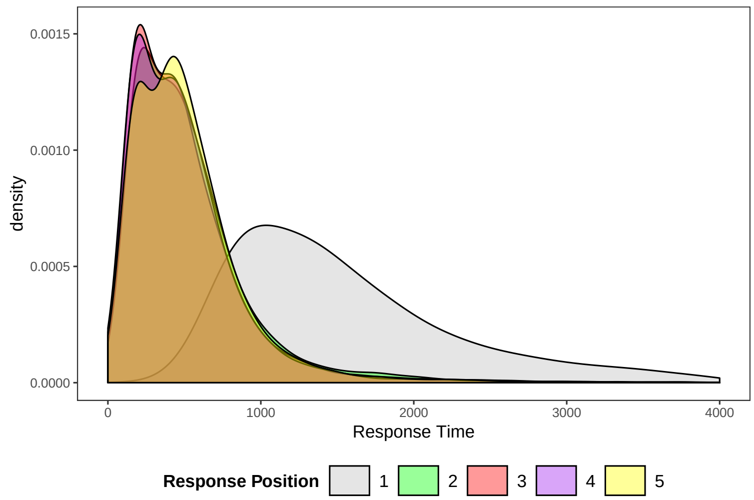
The parameters of the model were as follows; the threshold (or amount of evidence required for a response to be made), *b*; the drift rate or average speed of accumulation, *v*; the range of start points, *A*; the standard deviation of drift rates across trials, *s*; and the non-decision time, or time accounted for by processes unrelated to the decision process such as encoding and responding time, T_0 (Brown & Heathcote, 2008). In any one condition, there is an accumulator for the choice of responding with "target" and an accumulator for responding with "non-target", and the parameters for each accumulator were not the same. For example, with regards to drift rate, the accumulator that matches the correct response (correct drift v_c) will have a higher average drift rate across trials compared to the accumulator that does not match the correct response (error drift v_e). Drift rates for each accumulator were allowed to vary across difficulty conditions, as responding "non-target" in the 0 dot condition should take little processing, compared to the 4 dot condition where evidence may accumulate at a slower rate given the increased difficulty. This decision is supported by literature suggesting that increased workload affects drift estimates (Tillman, Strayer, Eidels, & Heathcote, 2017; Castro, Strayer, Matzke, & Heathcote, 2019). The threshold (*b*) parameter factored in biases in responding and was allowed to vary as a function of both response ("target" or "non-target") and difficulty (since bias toward the "non-target" response should reduce when there are more targets to track). For example, in the 1 dot condition, it would be sensible to show a bias toward responding "non-target", since 9 of the 10 dots shown to participants were not targets to track. This bias also meant that in the 0 dot condition, it was always the case that participants should respond "non-target", hence why parameter estimates for responding "target" show a large amount of error. Since the encoding and motor response demands should be identical in the 0, 1, and 4 dot load conditions, we did not allow T_0 to vary as a function of difficulty (for examples, see Voss, Rothermund, & Voss, 2004; Dutilh et al., 2019), however, we conducted exploratory analysis regarding T_0 differences across difficulty, which are discussed in our results.

The above outlines our complex model where both *v* and *b* were able to vary across difficulties for each accumulator. However we also considered two simpler models; one which only enabled *v* to vary across difficulties for each accumulator and one which only enabled *b* to vary across difficulties for each accumulator. All models included a threshold (*b*) parameter which factored in the biases in responding discussed above.

It should be noted that in the 0 dot condition, there is no choice to be made. However, fitting the LBA to the 0 dot condition provides information about motor-processing. This is not a novel approach to fitting data without a choice (e.g., see LATER model in Noorani & Carpenter, 2016). Across all conditions, we fit a two-accumulator model. An alternative approach to this would be to fit a one-accumulator model in the 0 dot condition. In our two-accumulator model, the model is free to capture the data trend (i.e. not responding "target") by freely estimating a large "target" threshold. Furthermore, the posterior distribution for some parameters under the 0 dot condition will be unconstrained by data (as the "target" response in the 0 dot condition had few data, and consequently, the posterior is not estimated). As a result, in the 0 dot condition, our two-accumulator model is the numerical equivalent to fitting a one-accumulator model. Analysis not reported here confirmed this. Whilst choosing to fit a one-accumulator model would be a simpler model with respect to number of parameters, the two accumulator model provides a more simple explanation of the decision process, with model consistency across all task conditions. Although fitting the 0 dot condition tells us nothing about decision strategies, we are able to use these estimates to compare simple reaction time between the student and RAAF samples.

The three models were fit separately to the student and RAAF data. We estimated parameters for each group from the model using a Bayesian hierarchical approach. Each subject was allowed individual parameters which were constrained to follow truncated (positive only) normal distributions. At the group level, each of these parameters

Figure 4 ■ Response time distributions for the position of each sequential decision during the interrogation phase. In the interrogation phase, decision number one is the first decision made in each MOT trial and five is the final decision made relating to the previous MOT trial. As can be seen, the distribution of response times for the first decision is very different to the overlapping distributions for decisions two through to five.



varied across the mean and standard deviation. Uninformative and diffuse priors were specified. Group level mean parameter priors were truncated normal distributions and priors for standard deviation of the group level distributions were gamma. For a full overview of the prior distributions, see Appendices.

Samples were drawn from the posterior distribution using a differential evolution Markov chain Monte-Carlo method (de MCMC; Turner, Sederberg, Brown, & Steyvers, 2013). We ran 25 identical chains in parallel, with 500 iterations for burn-in and 1000 iterations for convergence for each group. The initial group level samples were drawn from the prior distributions and the subject level samples were randomly drawn from broad distributions of parameter values, covering a range wider than general LBA fits in simple decision making experiments.

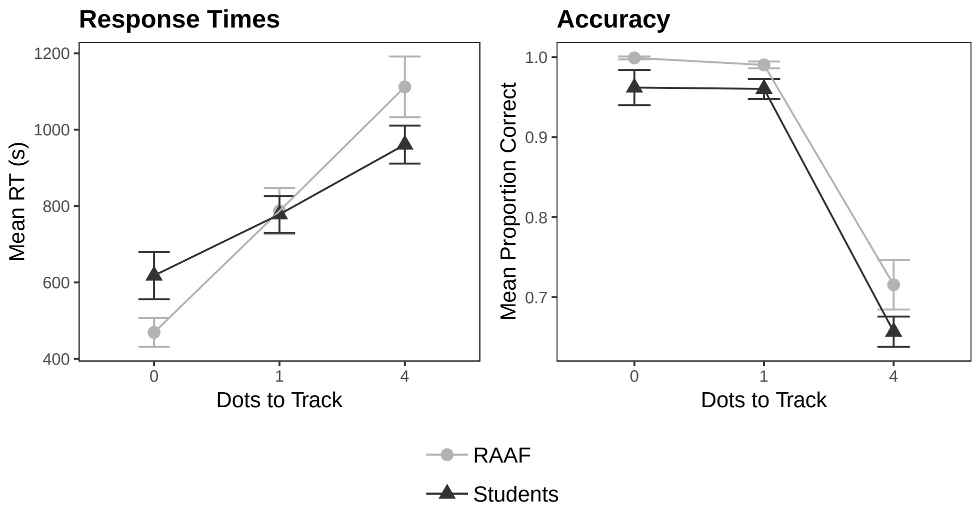
A point estimate of caution threshold for each participant was calculated from the posterior distributions of the parameters of the winning LBA model. This is calculated as the mean value of $b - A/2$ across the posterior samples. The drift rate was similarly calculated for each group across conditions and response types. For the sake of simplicity, we compare v_c (correct drift) to v_e (error drift) and b_T ("Target" response threshold) to b_{NT} ("Non-Target" response threshold) across difficulty and between groups.

Results

MOT response times greater than 2000ms were removed, however, there was no lower bound for response times. The reason for keeping response times under 200ms was that in some trials short response times were expected. For example in the 1 dot to track condition, if the target was identified early in the interrogation phase, only 'no' responses should remain, meaning participants could respond before perceiving or processing the remaining distractors. The first response in each trial was also removed. These responses were removed as the mean response times were affected by task switching costs (from the DRT to the MOT and from tracking to responding), meaning the distribution of response times was neither shifted nor a linear increase, but rather a distribution which was unrepresentative of participant responding. Figure 4 shows the distribution of response times for decision positions of the interrogation phase, where the first choice (shown in grey) is vastly different from the other four response positions.

For the Bayesian inferential statistics, we treated our study as a two-way mixed design, with the within-subject variable of difficulty (number of dots to track in the MOT – 0, 1 or 4) and the between-subject variable of group (RAAF or student). We assessed the response time and the proportion correct for responses to the MOT. We used two-way Bayesian repeated measures ANOVAs for each of the above measures of interest.

Figure 5 ■ Response times (left) and accuracy (right) across conditions of dots to track for both groups in the MOT task. Error bars are 95% confidence intervals.



The average proportion of correct responses in the MOT was 83.9% and average RT was .608 seconds. The change in RT across the different levels of difficulty and between groups can be observed in the left panel of Figure 5, with mean proportion correct shown in the right panel. Mean RT increased and mean proportion correct decreased as the level of difficulty increased. For mean RT there appears to be a crossover interaction effect of group and difficulty, where the RAAF group were faster in the 0 dot condition, but slower in the 4 dot condition. For mean proportion correct, it is clear that the RAAF group showed higher mean proportion correct across all levels of difficulty.

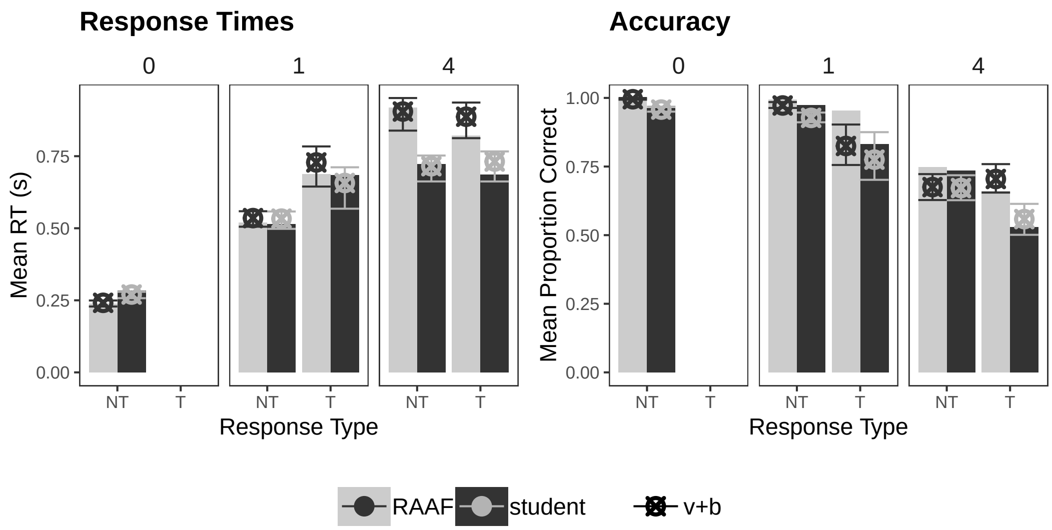
Bayesian ANOVAs showed a strong preference for the model that included the main effects of difficulty, group & interaction between these two factors for mean RT ($BF_{10} > 1000$) and a preference for the model that only included the main effects of difficulty and group for mean proportion correct ($BF_{10} > 1000$).

Bayesian paired samples t-tests revealed that these differences were reliable between all levels of difficulty for mean RT (0 dots vs. 1 dot, $BF_{10} > 1000$; 1 dot vs. 4 dot, $BF_{10} > 1000$) and for some levels of proportion correct (0 dots vs. 1 dot, $BF_{10} = 0.109$; 1 dot vs. 4 dots, $BF_{10} > 1000$). Bayesian independent samples t-tests also revealed strong evidence for differences in mean RT between the RAAF and student groups in the 0 and 4 dot conditions (0 dots, $BF_{10} = 11.86$; 4 dots, $BF_{10} = 23.45$), but not in the 1 dot condition ($BF_{10} = 0.208$). There was strong evidence for a difference in proportion correct between groups in the 1 and 4 dot difficulty conditions (1 dot, $BF_{10} = 14.41$; 4 dots, $BF_{10} = 26.41$), but no reliable difference in the 0 dot condition (0 dots, $BF_{10} = 1.79$). It is important to note the crossover interaction of groups and difficulty on mean RT, clearly illustrated on the left panel of Figure 5. This interaction may be the result of a strategic shift and/or processing difference between the groups, processes which could be investigated via sequential sampling modelling. These overall patterns are consistent with findings that increasing task difficulty leads to decreased response time and accuracy. The results show trends in line with our first hypothesis, with the RAAF group outperforming the student group for accuracy measures, but unexpectedly showing a crossover interaction effect for response times (where the RAAF group were faster in the 0 dot condition, but slower in the 4 dot condition).

Model Fit

To asses the goodness-of-fit of our models we chose to base our judgements on graphical evidence instead of any numerical index. The graphical approach and the statistical approach to model selection both show varying strengths and weaknesses (Shiffrin, Lee, Kim, & Wagenmakers, 2008; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008; Evans & Brown, 2018), with current goodness of fit methods unable to solve the dilemma of model inference (Dutilh et al., 2019). As shown in Figure **??** in Appendices, the predictions of the v only and b only models were clearly inconsistent with the data. The v+b model, displayed in Figure 6, showed a much better fit.

Figure 6 ■ Model fits for both response time (left) and accuracy (right). The bars indicate the data and the dots indicate the model predictions. Error bars for the dots are the inter-quartile range across different samples from the posterior. Three panels are included for each statistic representing the three levels of dots to track. Correct response types ("target" or "non-target") are shown on the x-axis. In the 0 dot condition, only "non-target" responses are shown as there were no targets and so there was no probability of having a correct "target" response.

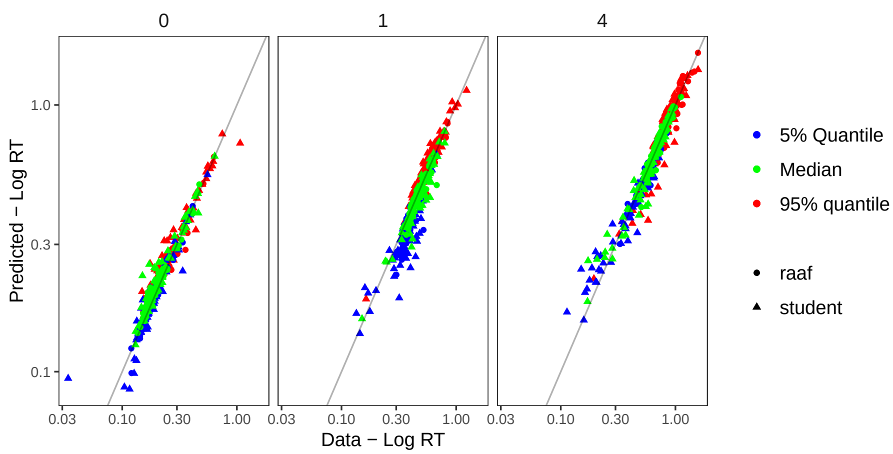


With a tightly constrained model, varying across v and b, the winning model fits the data well, as shown in Figure 6. Figure 6 shows the model predictions (dots) against the observed data (bars), with the inter-quartile range for the model included. We show both the mean response times (left panel) and proportion of correct responses (right panel) across response types ("target" and "non-target") for all three conditions of dots to track. For the model, the mean and inter-quartile range was calculated from 100 posterior predicted samples per subject, for both groups.

On inspection, the winning model appears to capture the key trends in the data and follows an expected pattern, with no gross misfit. Figure 7 and Figure 8 show response time and accuracy data (respectively) for the posterior predictive data plotted against the observed data. Figure 7 shows that the model appears to capture the trends in response time across the quantile range for the three conditions of difficulty. Figure 8 shows that the model accurately captures accuracy trends in the 4 dot condition, but appears to underestimate accuracy in the 1 dot condition for a subset of the student group. This may be due to students using cognitive "shortcuts" in this condition, where fast "non-target" guesses will produce higher accuracy than predicted. Both Figure 7 and Figure 8 highlight that as difficulty increases, response time increases and proportion correct decreases across both "target" and "non-target" responses. In the 0 dot condition, participants have fast response times and close to perfect accuracy, which is captured by the model (shown in Figure 6). This is similar in the 1 dot condition, where we see a high proportion correct for "non-target" responses and a slightly lower proportion correct for "target" responses (as is expected based on the low frequency of correct "target" responses). An associated increase in response time is also captured by the model with participants responding slower in more difficult conditions than in the 0 dot condition, and slower to "target" responses than to "non-target" responses. The 4 dot condition shows a change in responding, as "target" responses have a faster response time, but lower accuracy. This may be due to the difficulty of the condition rather than purely a bias in responding, but is captured by the model. Although the model estimates are a close fit across the data there is a slight misfit for the RAAF group in the highest difficulty. It is evident that the model slightly overestimates the RAAF group proportion correct on "target" responses and underestimates "non-target" responses. Furthermore, there is a slight misfit for the RAAF group response times, with the model overestimating response times for "target" conditions.

With a clear threshold effect across difficulty conditions, driven by inbuilt task biases we considered the pos-

Figure 7 ■ Model fit for the posterior predictive response times (y-axis) against the data (x-axis). The two groups are represented by different shapes (circles for RAAF and triangles for students). Each participant has three coloured dots in each panel representing the lower quartile (blue), upper quartile (red) and median values. A $y = x$ diagonal line is shown. Each panel shows the fit under each of the difficulty conditions.



sibility that T_0 may also change across difficulty conditions (Dutilh et al., 2019). We therefore conducted exploratory post-hoc analyses of the data, where we looked at the fastest 1% of responses in the 0, 1, and 4 dot conditions for both groups, since changes in the leading edge of the response time distribution have been taken to indicate changes in non-decision time (Ratcliff & Tuerlinckx, 2002). Figure 9 shows the fastest 1% of responses for both the RAAF group and student group. The figure indicates that these response times increased as difficulty increased, indicating a possible T_0 effect. More interestingly however, the model generated posterior predictive data shows trends consistent with this pattern. This analysis indicates that, although the difficulty condition did shift the leading edge of the response time distribution, the winning *v*+*b* model captured this effect, and these changes are consistent with changes in threshold.

Model Results

Given the winning LBA model accounts for changes in response time and proportion correct by varying *v* and *b*, we can observe the change to these parameters in Figure 10. Figure 10 shows the *v* parameter (left panel) for correct (v_c) and error (v_e) responses and *b* (right panel) for responding "non-target" (b_{NT}) and "target" (b_T) across levels of difficulty for both groups. We show that as difficulty increases, v_c decreases and b_{NT} increases. The decrease in v_c indicates that the difficulty manipulation slows the rate of processing at the decision level. The increase in b_{NT} as difficulty increases is in line with the expected proportion of responding "non-target", with b_{NT} low for both groups at the 0 dot condition, but gradually increasing with added targets. We also show that b_T decreases from being extremely high at the 0 dot condition, to almost the same as b_{NT} at the 4 dot condition.

The differences between groups is also evident in Figure 10 and 11. The *b* parameter shows a large difference in response thresholds between RAAF and students. RAAF personnel display higher thresholds at all levels of difficulty and response types, possibly indicating a greater level of caution. For *v*, there is little difference between the RAAF and student groups across difficulty levels, however the v_c parameter for 1 dot to track does show some difference with RAAF personnel showing higher drift for correct responses. Combining both *v* and *b* provides a clear explanation for the effects observed at response time and accuracy levels. The RAAF personnel display higher thresholds in the more difficult conditions leading to higher accuracy than the students and slower response times. This is not true of the 1 dot condition, where despite the RAAF thresholds being higher, response times show no difference to students. This may be accounted for by a faster drift rate observed at v_c. Finally, Figure 11 shows a clear difference in the *b* parameter between RAAF and student groups. This parameter was used to calculate the bias in b_{NT} and b_T, and shows a more efficient strategy used by the RAAF group to set their threshold bias.

Figure 8 ■ Model fit for the posterior predictive accuracy (y-axis) against the data (x-axis). The two groups are represented by different shapes (circles for RAAF and triangles for students). A $y = x$ diagonal line is shown. Each panel shows the fit under each of the difficulty conditions.

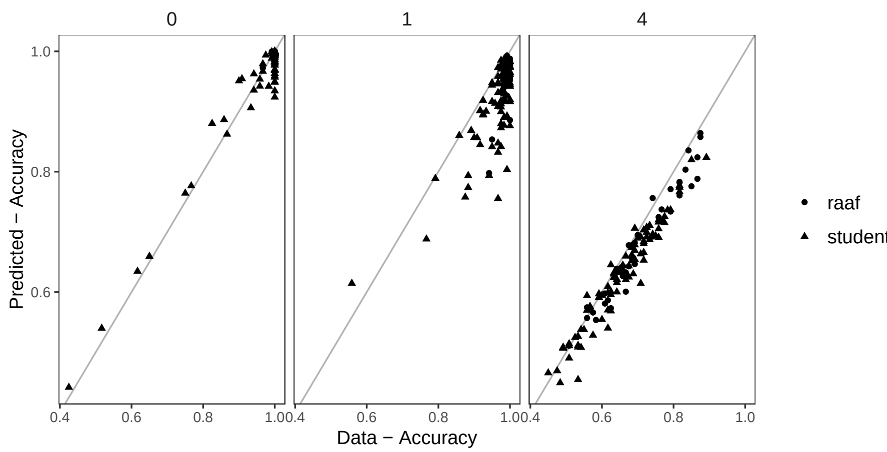
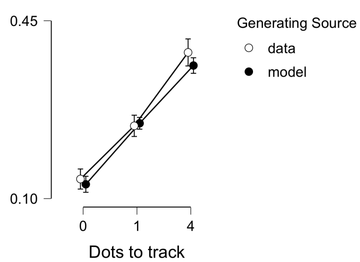
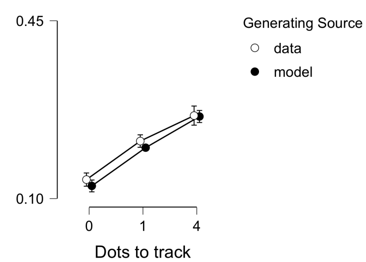


Figure 9 ■ Fastest 1% of response times across difficulty conditions, shown for both groups (RAAF left and Students right). Colour indicates the data generating source, where “data” is taken from the behavioural data and “model” is taken from the posterior predictive samples. Error bars indicate 95% confidence intervals.

(a) RAAF



(b) Students



Discussion

Evidence accumulation models are emerging as a dominant method of analysis for a wide variety of decision making tasks because they allow us to discriminate between the effects of latent cognitive processes. These decision making theories propose that we require set amounts of evidence for each response option before we make a decision. It is this “threshold”, as well as the rate of processing, non-decision time and certain cognitive biases that play a role in not only the speed of a decision, but also the accuracy. How these elements are traded off against each other has been shown to vary across tasks and populations.

Data from decisions in the MOT task is typically analysed only through a traditional hypothesis testing lens, looking at differences in accuracy and response time data, but not accounting for any speed-accuracy trade off or joint distribution. Here we show that sequential sampling models can be fit to decision data from the MOT to reveal underlying decision strategies used by different groups. Further, since prior research has tended to focus on investigating cognitive deficits within certain populations, rather than proficiencies (e.g., Dillon et al., 2015; Forstmann et al., 2011; Heathcote et al., 2015; van Ravenzwaaij et al., 2012), it was opportunistic to model the decision strategies and processes from a tightly selected group of RAAF personnel.

Initial basic analysis revealed trends consistent with hypotheses; that increasing the difficulty of the MOT task led to a deterioration of performance (slower response times and lower accuracy), and that the RAAF group outperformed the students in their proportion of correct responses. We did not expect to find an interaction between

Figure 10 ■ Parameter estimates for *v* and *b* across both groups and conditions. For the *v* parameter plot, each condition (0,1 and 4 dots to track) has two associated parameters - v_c for correct responses; and v_e for error responses. Similarly, in the *b* plot, each condition has two parameters - b_{NT} the threshold for responding "non-target" and b_T the threshold for responding "target".

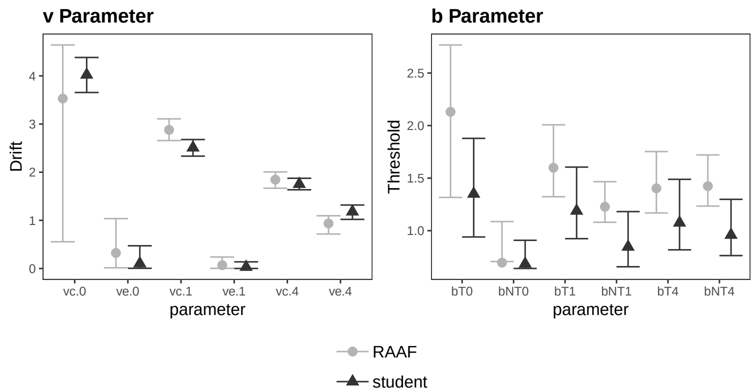
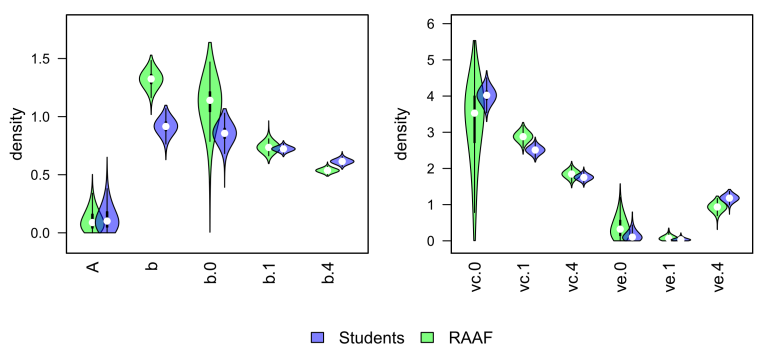


Figure 11 ■ Violin plots of posterior distributions of parameter estimates for the two groups. T_0 is not shown here as there was no group differences and relatively small estimates.



groups and difficulties for response time data, however, this provided a strong exemplar of why further modelling should be undertaken. In modelling this data, we found that response threshold and rate of processing may provide an account for the large difference in both speed and accuracy between a highly trained RAAF group and a control student group. The RAAF group appear to set higher thresholds for the 1 dot condition, but this is moderated by their superior drift rates, which may explain why the RAAF group had higher accuracy, but no difference in response times in this condition. In the 4 dot condition, the RAAF group appear to set higher thresholds, but have a similar drift rate to the student group, which is consistent with the finding that the RAAF had slower response times, but retained their higher level of accuracy. It is not clear as to why the RAAF group had a higher drift rate for the correct accumulator in the 1 dot condition but not the 4 dot condition. A possible explanation for this phenomenon is a ceiling effect in the most difficult condition. Other factors such as attention, incentive and cognitive ability may have also contributed to this effect, however these are difficult to quantify from the data. As only 0, 1 and 4 dot conditions were observed, it would be of interest to also test participants at 2 and 3 dot conditions to see if differences in processing speed followed a linear decay under these intermediate conditions.

In the MOT task, it was clear that increasing the number of dots to track leads to a marked decrease in accuracy, as has been shown in previous studies (Pylyshyn & Storm, 1988; Pylyshyn, 2004). This is especially prevalent between the 1 dot and 4 dot conditions. This performance decrease indicates the high difficulty of the task at the higher level of dots to track. Furthermore, when observing response times, it is clear that increased difficulty contributed to an increase in decision time. When accounting for groups, it is evident that the RAAF group outperforms the student group at all levels of MOT when observing accuracy scores. However, there is a clear crossover interaction effect observed in mean response times between groups and levels of difficulty. Whilst the RAAF group had a faster mean response time than the students in the 0 dot condition, they have a slower mean response time in the 4 dot condition. A more informative model-based approach enables insight into underlying cognitive processes, in order to see what causes differences in descriptive data.

The winning model appeared to fit the data well, however a slight misfit was observed in the posterior predictive data for the accuracy of the student group in the 1 dot condition (as shown in Figure 8). For the student group, the model under-predicted the level of accuracy for a subset of students, however, the nature of biased responding may mean there are "shortcuts" that contribute to this effect. That is, in the 1 dot condition, a participant could choose to guess in some trials by only responding "non-target" and still exhibit high accuracy and fast response times. Another possibility for the under-fit relates to the high sequential dependence in the 1 dot condition (participants are unlikely to make multiple "target" responses in one MOT trial). The model also showed a slight misfit (as seen in Figure 6) for the RAAF when responding "target", slightly overestimating their response time and underestimating proportion correct. This misfit should be considered when modelling such data in future and could potentially be addressed with more interrogations per trial in the MOT or through building a joint model which factors in sequential effects in responding or the between subjects effects of groups.

The current study is limited in that we do not account for DRT data, which is part of a broader project, nor do we account for the sequential effects of the response position. DRT data could be highly informative to the model with a between task trade off evaluated alongside the speed-accuracy trade off of both tasks. However, this modelling approach would require a more sophisticated analysis in combining two different types of data from two divergent, semi-parallel tasks. We decided to exclude data from the first response given after each MOT tracking period due to task switching demands causing an increase in the first response time, however, these responses could provide informative data if they were modelled effectively. Models that account for this may then provide information about the task switching costs as well as the underlying decision components. The current model treats all sequential decisions as independent, thus not accounting for any secondary sequential effects of prior responses, which may be particularly prevalent in the 1 dot condition. More advanced modelling strategies might attempt to include these secondary effects and see how this affects the interpretation of the key latent psychological processes and strategies we were interested in. Future experiments should look to extend the number of decisions made in the interrogation phase to cover all of the objects in the display. Further advanced modelling exercises should also be undertaken to make use of the well researched tracking models in conjunction with the decision processes.

Despite these limitations, our study provides the first use of a sequential sampling models for decisions in the MOT task to our knowledge, but more-so we highlight the importance of deeper investigation into participant responding. Our results clearly show a difference in responding between the two groups, with the RAAF group showing higher accuracy. However, it is only through a model framework that we are able to tease apart group differences and show evidence of underlying processes that may be responsible for the overall differences.

Authors' note

We would like to thank No. 4 Squadron, Williamtown for the acquisition of data from combat controller candidates.

References

Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (2016). A diffusion model analysis of episodic recognition in preclinical individuals with a family history for alzheimer's disease: The adult children study. *Neuropsychology*, *30*(2), 225.

Batchelder, W. H. (2010). Cognitive psychometrics: Using multinomial processing tree models as measurement tools. doi:10.1037/12074-004

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

Brown, S. D., Marley, A., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological review*, *115*(2), 396. doi:10.1037/0033-295X.115.2.396

Castro, S. C., Strayer, D. L., Matzke, D., & Heathcote, A. (2019). Cognitive workload measurement and modeling under divided attention. *Journal of experimental psychology: human perception and performance*, *45*(6), 826.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in cognitive sciences*, *9*(7), 349–354.

Dillon, D. G., Wiecki, T., Pechtel, P., Webb, C., Goer, F., Murray, L., ... Weissman, M., et al. (2015). A computational analysis of flanker interference in depression. *Psychological medicine*, *45*(11), 2333–2344.

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, *41*(4), 1095–1110.

Donkin, C., & Brown, S. D. (2018). Response times and decision-making. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, *5*, 1–33.

Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychonomic Bulletin & Review*, *16*(2), 411–417.

Dube, C., Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Beyond roc curvature: Strength effects and response time data support continuous-evidence models of recognition memory. *Journal of Memory and Language*, *67*(3), 389–406.

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., ... Krypotos, A.-M., et al. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*(6), 1026–1036.

Evans, N. J., & Brown, S. D. (2018). Bayes factors for the linear ballistic accumulator model of decision-making. *Behavior research methods*, *50*(2), 589–603.

Evans, N. J., Rae, B., Bushmakin, M., Rubin, M., & Brown, S. D. (2017). Need for closure is associated with urgency in perceptual decision-making. *Memory & cognition*, *45*(7), 1193–1205.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience*, *31*(47), 17242–17249.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389.

Haritaoglu, I., Harwood, D., & Davis, L. S. (1998). W 4 s: A real-time system for detecting and tracking people in 2 1/2d. In *European conference on computer vision* (pp. 877–892). Springer.

Hawkins, G. E., Marley, A., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive science*, *38*(4), 701–735.

Heathcote, A., Suraev, A., Curley, S., Gong, Q., Love, J., & Michie, P. T. (2015). Decision processes and the slowing of simple choices in schizophrenia. *Journal of Abnormal Psychology*, *124*(4), 961.

Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed-accuracy tradeoff. *Neuron*, *76*(3), 616–628.

Ho, T. C., Yang, G., Wu, J., Cassey, P., Brown, S. D., Hoang, N., ... Duncan, L. G., et al. (2014). Functional connectivity of negative emotional processing in adolescent depression. *Journal of affective disorders*, *155*, 65–74.

JASP Team. (2019). JASP (Version 0.11.0)[Computer software]. Retrieved from https://jasp-stats.org/

Koller, D., Weber, J., & Malik, J. (1994). Robust multiple car tracking with occlusion reasoning. In *European conference on computer vision* (pp. 189–196). Springer.

Noorani, I., & Carpenter, R. (2016). The later model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251.

Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial eeg correlates of drift-diffusion model parameters. *Journal of mathematical psychology*, *76*, 117–130.

Osth, A. F., Bora, B., Dennis, S., & Heathcote, A. (2017). Diffusion vs. linear ballistic accumulation: Different models, different conclusions about the slope of the zroc in recognition memory. *Journal of Memory and Language*, *96*, 36–61.

Osth, A. F., Jansson, A., Dennis, S., & Heathcote, A. (2018). Modeling the dynamics of recognition memory testing with an integrated model of retrieval and decision making. *Cognitive psychology*, *104*, 106–142.

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. *Journal of Experimental Psychology: Applied*, *22*(1), 1.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. tracking without keeping track of object identities. *Visual cognition*, *11*(7), 801–822.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, *3*(3), 179–197.

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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226.

Rasmussen, C., & Hager, G. D. (1998). Joint probabilistic techniques for tracking multi-part objects. In *Proceedings. 1998 ieee computer society conference on computer vision and pattern recognition (cat. no. 98cb36231)* (pp. 16–21). IEEE.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological review*, *111*(1), 159.

Ratcliff, R., & McKoon, G. (2015). Aging effects in item and associative recognition memory for pictures and words. *Psychology and aging*, *30*(3), 669.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological review*, *116*(1), 59.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and iq in two-choice tasks. *Cognitive psychology*, *60*(3), 127–157.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and iq on item and associative memory. *Journal of Experimental Psychology: General*, *140*(3), 464.

Ratcliff, R., & Van Dongen, H. P. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*(27), 11285–11290.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive psychology*, *38*(2), 259–290.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-.-.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*, 283–317.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and aging*, *25*(2), 377.

Tillman, G., Strayer, D., Eidels, A., & Heathcote, A. (2017). Modeling cognitive load effects of conversation between a passenger and driver. *Attention, Perception, & Psychophysics*, 1–9.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological methods*, *18*(3), 368.

van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2012). A diffusion model decomposition of the effects of alcohol on perceptual decision making. *Psychopharmacology*, *219*(4), 1017–1025.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & cognition*, *32*(7), 1206–1220.

Wagenmakers, E.-J., Lee, M. D., Lodewyckx, T., & Iverson, G. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). New York: Springer Verlag.

2020 ■ Vol. 16 ■ no. 2

Appendices

The *v*+*b* LBA Model

The winning LBA model that allowed drift rate and threshold to vary across conditions. This model also builds the *b* parameter bias:

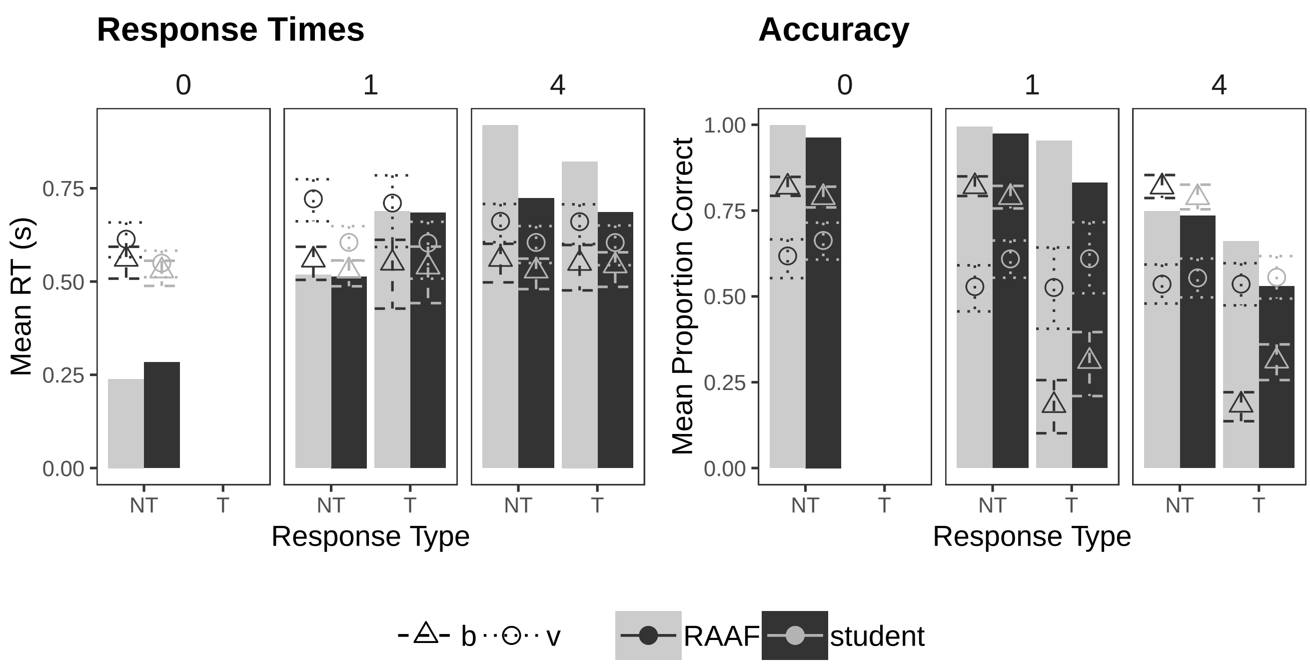
$$
\begin{aligned}
Data\ \ level: & & \\
(RT, resp) &\sim LBA(A_i, b'_T, b'_{NT}, v_{j.T,i}, v_{j.NT,i}, s_T = 1, s_{NT} = 1, t_{0,i}) \\
b'_T &= A_i + b_i(1 + (2 \times tanh(B_j) - 1)) \\
b'_{NT} &= A_i - b_i(1 + (2 \times tanh(B_j) - 1)) \\
& \quad Where\ \ i = person\ \ and\ \ j = condition \\
Group\ \ level: & & \\
A_i &\sim N_+(\mu_A, \sigma_A) \\
b_{\{0,1,4\}.T,i} &\sim N_+(\mu_{b,\{0,1,4\}.T}, \sigma_{b,\{0,1,4\}.T}) \\
b_{\{0,1,4\}.NT,i} &\sim N_+(\mu_{b,\{0,1,4\}.NT}, \sigma_{b,\{0,1,4\}.NT}) \\
v_{\{0,1,4\}.T,i} &\sim N_+(\mu_{v,\{0,1,4\}.T}, \sigma_{v,\{0,1,4\}.T}) \\
v_{\{0,1,4\}.NT,i} &\sim N_+(\mu_{v,\{0,1,4\}.NT}, \sigma_{v,\{0,1,4\}.NT}) \\
t_{0,i} &\sim N_+(\mu_{t0}, \sigma_{t0}) \\
Prior\ \ distributions: & & \\
\mu_A, \mu_{b,0}, \mu_{b,1}, \mu_{b,4} &\sim N_+(2, 2) \\
\mu_{v,T,0}, \mu_{v,T,1}, \mu_{v,T,4}, \mu_{v,NT,0}, \mu_{v,NT,1}, \mu_{v,NT,4} &\sim N_+(3, 3) \\
\mu_{t0} &\sim N_+(.5, .5) \\
\sigma_A, \sigma b, 0, \sigma_{b,1}, \sigma_{b,4} &\sim \Gamma(1, 1) \\
\sigma_{v,T,0}, \sigma_{v,T,1}, \sigma_{v,T,4}, \sigma_{v,NT,0}, \sigma_{v,NT,1}, \sigma_{v,NT,4} &\sim \Gamma(1, .5) \\
\sigma_{t0} &\sim \Gamma(1, 3)
\end{aligned}
$$

For the model, *i* indexes participants and *j* indexed the level of dots to track. The data level distribution was an LBA model with eight parameters, and the data to which this was applied consisted of six response time distributions (correct and incorrect responses at each level of dots to track) and two response probabilities (the proportions of correct responses in each condition).

2020 ■ Vol. 16 ■ no. 2

Model Fits for *b*–only and *v*–only

Model fits for both response time (left) and accuracy (right). The bars indicate the data and the dots indicate the model predictions. Error bars for the dots are the inter-quartile range across different samples from the posterior. Three panels are included for each statistic representing the three levels of dots to track. Correct response types ("target" or "non-target") are shown on the x-axis. In the 0 dot condition, only "non-target" responses are shown as there were no targets and so there was no probability of having a correct "target" response. The shape of the dot indicates the generating model.



Citation

Innes, R. J., & Kuhne, C. L. (2020). An LBA account of decisions in the multiple object tracking task. *The Quantitative Methods for Psychology*, *16*(2), 175–191. doi:10.20982/tqmp.16.2.p175

Copyright © 2020, *Innes and Kuhne*. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 29/01/2019 ∼ Accepted: 21/01/2020