Evidence accumulation modelling in the wild: Understanding safety-critical decisions

Russell J. Boag[1,*]; Luke Strickland[2]; Andrew Heathcote[3,4]; Andrew Neal[5]; Hector Palada[5]; & Shayne Loft[1]

[1]School of Psychological Sciences, University of Western Australia, Crawley WA 6009, Australia

[2]Future of Work Institute, Curtin University, Perth WA 6000, Australia

[3]School of Psychology, University of Newcastle, Callaghan NSW 2308, Australia

[4]Department of Psychology, University of Amsterdam, 1018 WS Amsterdam, The Netherlands

[5]School of Psychology, University of Queensland, St Lucia QLD 4072, Australia

[*]Corresponding author: russell.boag@uwa.edu.au (R.J. Boag)

Manuscript accepted for publication at Trends in Cognitive Sciences, 16-Nov-2022

Citation:

Boag, R.J., Strickland, L., Heathcote, A., Neal, A., Palada, H., & Loft, S. (accepted 16-Nov-2022). Evidence accumulation modelling in the wild: Understanding safety-critical decisions. *Trends in Cognitive Sciences.*

Abstract

Evidence accumulation models are a class of computational cognitive model used to understand the latent cognitive processes that underlie human decisions and response times. They have seen widespread application in cognitive psychology and neuroscience. However, historically the application of these models was limited to simple decision tasks. Recently, researchers have applied these models to gain insight into the cognitive processes that underlie observed behaviour in applied domains such as air-traffic control, driving, forensic and medical image discrimination, and maritime surveillance. Here, we discuss how this modelling helps to understand how the cognitive system adapts to task demands and interventions such as task automation. We discuss future directions and argue for wider adoption of cognitive modelling in Human Factors research.

Key words: *evidence accumulation; computational cognitive model; decision making; human factors; performance and safety; applied cognition*

Bringing computational modelling out of the lab and into the wild

Computational cognitive models are powerful tools for understanding human cognition and behaviour. The models are *cognitive* because they explain how unobserved cognitive processes (e.g., attention, learning, working memory capacity) give rise to observed behaviour (e.g., choice, response time; RT). The models are *computational* because theorized relations between cognition and behaviour are defined unambiguously in terms of formal mathematics and instantiated in executable computer code. This enables the precise, quantitative measurement of latent cognitive processes and ultimately allows for stronger tests of competing cognitive theories than is possible through verbal (non-computational) reasoning or analysis of observed behaviour alone [1].

Evidence accumulation models (EAM) are among the most prominent and successful computational cognitive models in cognitive psychology and neuroscience [2-8]. EAMs explain the outcome and duration of decisions in terms of latent cognitive processes including the efficiency of information processing, the amount of evidence required to trigger a response, and the duration of encoding and motor response processes. In contrast to traditional analysis of mean RT and error rates, which can be ambiguous or difficult to interpret, EAMs account for all aspects of the data (e.g., skew and variability of RT distributions) and can identify differences in underlying decision processes that cannot be inferred from traditional descriptive analyses [9].

In the cognitive (neuro)sciences, EAMs have been most widely applied to simple, highly controlled decision-making tasks (e.g., brightness discrimination, random dot motion, lexical decision, stop-signal, go/no-go tasks). The simplicity of highly controlled tasks enables precise, targeted measurement of cognitive processes and facilitates interpretation of neurophysiological measures (e.g., EEG, fMRI) [2,10-12]. However, such tasks are seldom representative of the more complex and cognitively demanding decision-making contexts that humans face in the modern workplace [13,14]. Consequently, the practical implications of such work for how humans make decisions ‘in the wild’ are often unclear for those seeking to understand the cognitive underpinnings of human performance and errors in safety-critical work domains. Bringing EAMs ‘into the wild’ holds reciprocal benefits for applied and basic research: Applied research benefits from greatly enhanced measurement

of the latent cognitive mechanisms underlying performance. Basic research benefits from understanding how cognitive theories generalise to representative complex work tasks [15].

In this article, we review recent work pioneering the use of EAMs to study representative simulations of real-world decisions in such diverse domains as air-traffic control (ATC), driving, forensic and medical image discrimination, and maritime surveillance. We first outline the theory and key computational features of EAMs. Next, we review several recent novel insights into human decision making in safety-critical work tasks made possible by EAMs. In doing so, we show that EAMs provide a common theoretical framework that explains human performance across a diverse set of modern work tasks. Finally, we discuss future directions and argue for wider adoption of computational cognitive modelling approaches in applied (Human Factors) research.

The architecture of evidence accumulation

Two of the most successful EAMs, the diffusion decision model (DDM) [16] and linear ballistic accumulator (LBA) [17] are illustrated in Fig. 1. In these models, decision making involves sampling evidence from the task environment until a threshold amount of evidence is reached. There is typically one threshold for each possible choice option in the experimental task, and the first threshold reached triggers the corresponding overt response. Across repeated decisions, the distribution of threshold crossing times (plus the time for non-decision processes like stimulus encoding and response production) describes a decision maker's distribution of empirical RTs, and the proportion of times evidence terminates at each threshold describes the empirical response proportions. Predicting both choices and RTs is critical because the slowest or fastest responses can pose unique risks (e.g., rash decisions or slow detection of an unsafe event can both be hazardous). Explaining the entire shape of RT distributions in terms of latent cognitive processes is a critical advantage of EAMs (over analyses of behavioural summaries such as mean RT and error rate) that allows EAMs to provide a coherent account of complex or ambiguous observed effects [9].

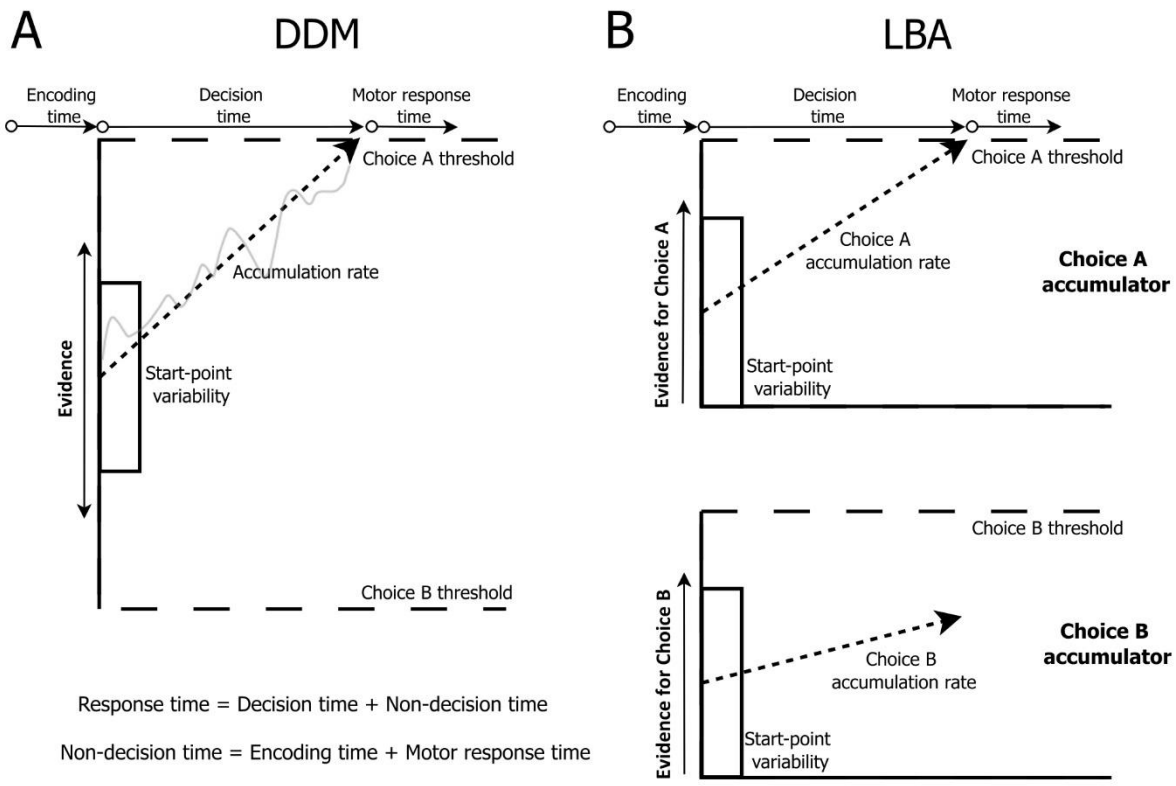

Figure 1. Diffusion decision model (DDM) and linear ballistic accumulator (LBA) evidence accumulation architectures.

(A) In the DDM, noisy evidence starts accumulating at a point between two response boundaries (thresholds) and terminates when either threshold is reached. Accumulation rate measures the difference in evidence strength for the two response options (a step toward one boundary is a step away from the other). The distance between the response boundaries represents response caution, which controls trade-offs between speed and accuracy. Moving the evidence start point closer to one boundary (relative to the other) creates a bias towards that response. (B) In the LBA, instead of relative evidence, evidence for each response accumulates in independent 'racing' accumulators, each with its own response threshold. The first accumulator to reach threshold triggers the overt response. Threshold height controls response caution settings. Biases can be induced by setting a low threshold for the target response and high threshold(s) for the other response(s). Owing to its modular architecture, the LBA is easily applied to decisions involving an arbitrary number of response options. Other common architectures include the single-boundary diffusion (noisy evidence accumulates toward a single threshold) [54,68,69] and racing diffusion models (noisy evidence accumulates independently in two or more racing accumulators) [70,71].

Importantly, EAM parameters have psychologically meaningful interpretations in terms of latent cognitive processes [18,19]. The mean rate of evidence accumulation represents the efficiency of information processing. Accumulation rates are jointly determined by stimulus

characteristics (e.g., salience and discriminability from other choice options) and the amount of attention/cognitive resources devoted to the task. When stimulus characteristics are held constant, accumulation rates measure the level of attention devoted to the task and hence are a powerful tool for quantifying cognitive demands when attentional capacity is exceeded. For example, increased task demands can impair the rate at which an air-traffic controller processes potential conflicts (i.e., violations of minimum aircraft separation standards), leading to slower and more error-prone observed performance. Accumulation rates converge with other rigorous measures of cognitive capacity such as Systems Factorial Technology [20-22], the gold-standard nonparametric method for determining whether a human processing architecture has limited, unlimited, or super capacity. Several studies reviewed below use accumulation rates to identify conditions of unmanageable workload in which task demands exceed the operator's capacity to manage them, leading to significant performance degradations [23-33].

A threshold's height relative to where evidence starts accumulating (i.e., the amount of evidence required to trigger a response) measures response caution, with higher thresholds producing more cautious decisions since more evidence is required to reach a decision. The relative position of thresholds to each other measures bias towards responding one way or another. Continuing our running example, an air-traffic controller may respond to perceived heightened time pressure both by adopting lower response thresholds overall (producing a global speed-up) and by shifting bias towards classifying aircraft as in-conflict versus not-in-conflict (increasing false alarms but ensuring no conflicts are missed) [34]. Since thresholds are set in advance of stimulus presentation, they are considered a locus of proactive cognitive control strategies [35,36]. Several studies reviewed below use thresholds in this manner to identify when and how individuals proactively adapt decision-making strategies to deal with anticipated task demands (e.g., when facing heightened time pressure and/or additional task complexity), and to identify potential drawbacks of certain strategies [23,24,26-28].

Non-decision time measures the duration of perceptual encoding and motor response processes. Several studies reviewed below use non-decision time to identify situations where individuals fail to encode stimuli with sufficient detail to make reliable decisions [25,26]. For example, an air-traffic controller under extreme time pressure may

inadequately encode information about potential conflicts, leading to a shortened non-decision time and high miss rate.

Finally, parameters controlling between-trial variability in accumulation rate and starting point account for commonly observed differences in the relative speed of correct and incorrect responses [37,38]. Although less commonly interpreted than accumulation rate, threshold, and non-decision time, some studies use variability parameters to identify task factors that lead to increased uncertainty (greater variability) in decision making [26,27,39].

Decomposing performance into these underlying cognitive processes has clear implications for the modern workplace. Identifying whether errors are due to the quality of information provided by the task display has implications for interface design. In contrast, an operator using a suboptimal strategy has implications for work training, whereas an operator having reduced cognitive resources due to excessive workload has implications for work design.

Box 1 contains considerations regarding the application of EAMs to decisions that unfold over longer timescales than those typical of highly controlled lab settings.

Box 1. Modelling long timescale decisions with EAMs

An important question concerns whether standard EAMs represent an appropriate model of naturalistic tasks in which decisions unfold over longer timescales than are typically seen in highly controlled lab settings (e.g., mean RT < 1.5 seconds). Most EAMs assume that decisions are the result of a single continuous evidence accumulation process. However, violations of this assumption become increasingly plausible at longer timescales, where decisions may be the result of multiple, potentially sequential, unobserved processing stages.

In every study throughout this review, standard (single accumulation process) EAMs provided close fits to relatively long decisions (e.g., 2-10 seconds mean RT) and generated inferences consistent with those in the short-RT literature (i.e., accumulation rates as the locus of capacity sharing, discriminability, and reactive control effects; thresholds as the locus of proactive control and response bias effects). This suggests that the standard EAM framework is robust to potential violations of the single accumulation process assumption and can be a valid measurement model of longer timescale naturalistic decisions. This is supported by simulation studies showing standard EAMs provide close fits and theoretically sensible parameter effects for tasks with mean RT up to 7.4 seconds in which the single accumulation process assumption is explicitly violated [72].

In some settings, it is also possible to test empirically for the appropriate processing architecture. For example, when examining performance of a task involving asynchronous stimuli with different onsets

within each trial, [27] compared model fits to RTs computed assuming either parallel or serial processing of stimuli (i.e., RT from stimulus onset for parallel; RT from termination of the previous response for serial). Given that [27] found that assuming the incorrect architecture resulted in severe miss-fit to RT distributions, this approach suggested the appropriate processing architecture whilst demonstrating the falsifiability of the EAM framework.

For situations in which the standard models fail, one can construct EAMs that explicitly account for multiple, potentially sequential, processing stages [73,74]. Such models have shown promise in highly controlled lab settings and could in principle be applied to longer timescale tasks. However, the additional complexity of these models renders some of them very computationally expensive to fit and the mechanisms describing unobserved within-trial dynamics may suffer from poor identifiability, particularly in less controlled applied settings. Generally, researchers should seek converging evidence about whether a single accumulation process can be assumed, especially when RTs are long.

We now turn to reviewing recent work that has used EAMs to understand human performance in representative simulations of complex dynamic work tasks. We demonstrate that EAMs provide a unified theoretical framework for explaining human performance across a diverse set of decision-making contexts and offer unique insights that practitioners can use to improve operator training and work design, and to inform the development of automated decision-support tools. The first section discusses findings surrounding limitations on operator attention and processing capacity, including when only limited or impoverished information is available from the task environment. The second section discusses findings regarding cognitive control strategies individuals use to adapt to task demands.

Attention, processing, and performance in the red zone

A central goal of Human Factors research is to identify limits on operators' ability to process task information while maintaining acceptable performance. When task demands exceed operator capacity, or when multiple channels of task demand compete for the same cognitive resource (e.g., attention, memory) [40], performance can suffer and potentially catastrophic errors may result (e.g., a pilot responding to multiple instrument warnings forgets to set flaps for landing; a driver attending to a passenger's conversation fails to

brake for an unexpected hazard). These situations are referred to as red zones/lines of workload [41], and designers of work systems must be aware of them to predict when task demands may degrade performance. However, red zones/lines are difficult to identify because humans employ counter measures (e.g., getting assistance from another operator or relying more on task automation) and/or adjust task processing strategies to avoid them [42,43]. In this manner, task demands are not simply imposed upon an operator, but rather, actively managed through resource allocation and strategy change [44,45]. EAMs provide a means of disentangling these effects, which are difficult to identify in traditional analyses of mean RT and/or error rates.

Recently, researchers have turned to EAMs to study the limits of attention and performance using representative simulations of ATC conflict detection [23,24], distracted driving [29,31-33,46], and maritime surveillance [25-28]. Two studies [23,24] investigated how prospective memory (PM) demands (i.e., the need to remember to perform a deferred action in the future) and time pressure affect the allocation of attention and cognitive capacity in individuals tasked with detecting potential conflicts between aircraft in simulated ATC. Understanding the resource requirements of PM is a critical applied question because PM tasks can impair controllers' performance on critical routine tasks (e.g., slower acceptance/hand-off of aircraft, slowed or failed conflict detection) [47,48]. Moreover, the experimental PM literature at the time was largely uninformed about PM capacity demands because the simple tasks (e.g., lexical decision) typically used did not place sufficient demands on cognitive capacity to necessitate resource sharing [35,49-52]. Using LBA accumulation rates to measure capacity, it was found in the more complex ATC task that PM demands did in fact drain resources from the conflict detection task, causing lower accumulation rates and resulting in slower and more error-prone conflict detection (Fig. 2A) [23,24]. In addition, the slowing induced by PM demands was especially detrimental under high time pressure (tighter response deadlines), with participants significantly more likely to fail to respond to potential conflicts on-time. Additionally, participants flexibly allocated capacity according to task priority, such that prioritized tasks received proportionally more resources at the expense of lower priority tasks: Prioritizing conflict detection reduced the severity of time pressure- and PM-induced costs to conflict detection performance whereas prioritizing the PM task increased the severity of those costs [24].

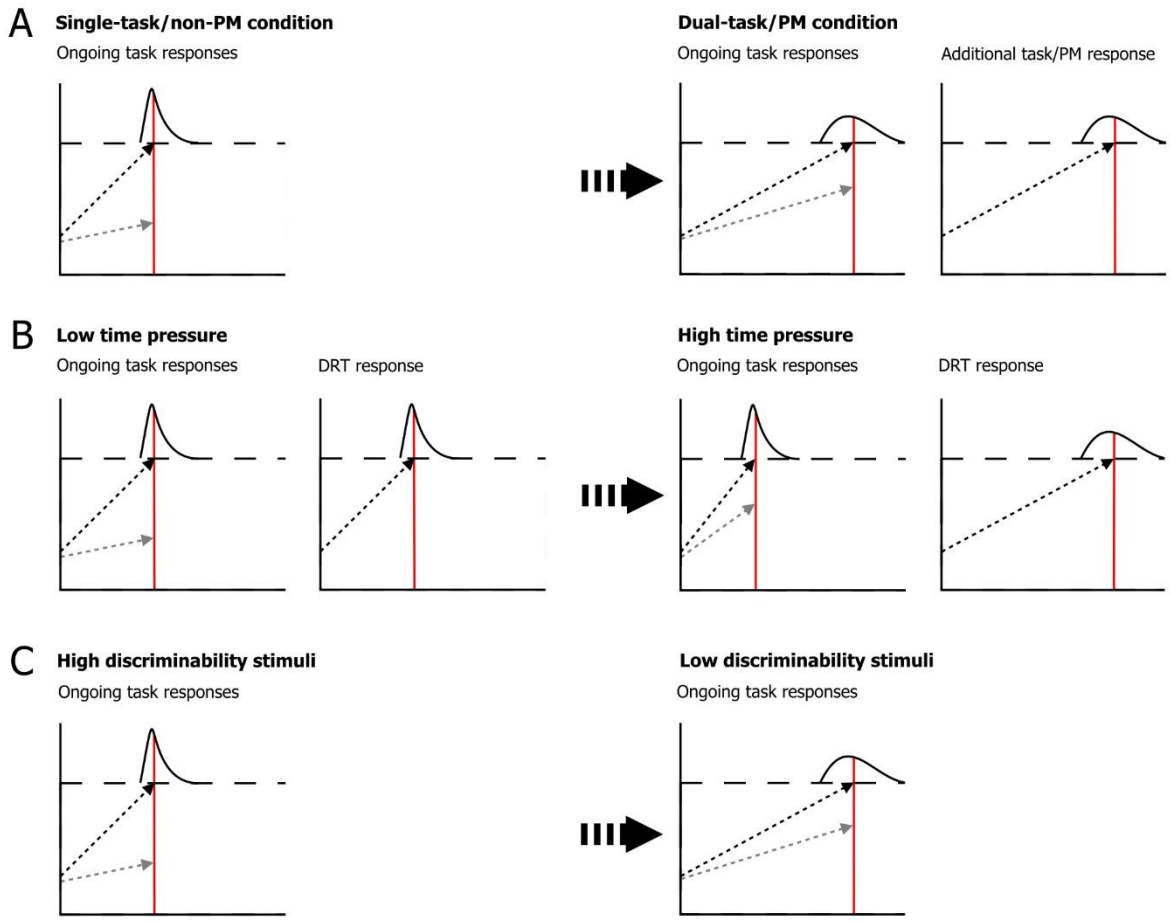

Figure 2. EAM accounts of capacity sharing and stimulus discriminability effects discussed in text.

In each accumulator diagram, black (grey) arrows represent mean accumulation rate for correct (incorrect) responses. Atop each threshold (dashed horizontal lines) is an RT distribution whose shape reflects the illustrated task/parameter effect. Red vertical lines indicate mean RT. Note that one ongoing task accumulator is used to represent potentially multiple ongoing task responses. (A) In single-task conditions (e.g., without a concurrent PM task) accumulation is fast and high quality (large rate difference for correct and incorrect responses), producing fast and accurate responses. In dual-task conditions (e.g., with a concurrent PM task) ongoing task accumulation is slow and poor quality (small rate difference for correct and incorrect responses), producing slower and less accurate responses. (B) At low time pressure, ongoing task and DRT accumulation is high quality and responses accurate. At high time pressure, some DRT accumulation is diverted to the ongoing task, increasing the speed of responses. However, the quality of ongoing task accumulation is lower, reducing accuracy. (C) With highly discriminable stimuli, ongoing task accumulation is fast and high quality and responses fast and accurate. With poorly discriminable stimuli, ongoing task accumulation is slow and poor quality and responses slow and inaccurate.

Similar PM-induced resource sharing effects occurred in a cognitively demanding maritime surveillance task, in which participants monitored an aerial view from an uninhabited aerial vehicle of five shipping lanes (partially obscured by cloud cover) and were tasked with classifying passing ships according to equipment visible on deck (e.g., cranes, masts, lifeboats) [28]. Reduced LBA accumulation rates under PM load indicated that adding a concurrent PM task drained resources that would have otherwise gone to the ongoing ship classification task, causing slower and less accurate classification performance (Fig. 2A).

Several studies have assessed operator attention and cognitive capacity by combining a primary task (e.g., maritime surveillance, driving) with a concurrent detection response task (DRT) [25,29,32,53]. The DRT is designed to measure spare (off-task) capacity and to infer cognitive workload on the primary task. In one study [25], participants performed the maritime surveillance task described above under varying degrees of time pressure (longer and shorter response deadlines) while simultaneously monitoring for DRT stimuli. Accumulation rates from LBA and racing diffusion models [54] measured how individuals allocated resources between the two tasks as demands increased. With greater time pressure, accumulation rates increased for ship classifications (speeding up responses) but decreased for DRT responses, indicating that individuals diverted resources away from the DRT and reallocated them to the ship classification task to compensate for greater task demands (Fig. 2B). Convergent results have been obtained without a concurrent DRT: Tighter deadlines led individuals to devote a greater quantity of resources (higher average accumulation rates) to classifying ships. However, the quality of processing was poorer (smaller difference in accumulation rates for correct and incorrect responses), which reduced accuracy. Expending additional resources and effort thus only partially compensated for increased task demands [26].

Similar work has studied distracted driving by pairing a simulated driving task with the DRT [29,31-33] or by directly modelling aspects of driver behaviour (e.g., braking and steering-wheel turning RTs) [30,46]. Understanding how drivers handle distraction is important for road safety because distractions (e.g., mobile phone use) increase driver RTs and reduce hazard detection rates [55,56]. Drivers show lower DRT accumulation rates and slower DRT responses when multi-tasking (steering while counting backwards by threes) compared to a single-task condition (steering only) [29]. Impaired DRT processing suggests that resources

were reallocated from the DRT to maintain performance on the other tasks (Fig. 2B; see also [33]).

Other work [30] used a single-boundary DDM to directly model braking RTs when a simulated driving task was performed either alone or while holding a distracting conversation (on a mobile phone and with a passenger) [57]. Drivers had lower accumulation rates for braking responses when distracted compared to when driving with no distraction. Distractions may thus impair drivers' ability to respond quickly and effectively to safety-critical events (e.g., a vehicle braking suddenly, a pedestrian stepping into traffic), since low accumulation rates produce slow, inaccurate responses. Follow-up work [32] found that conversation impaired DRT accumulation rates for both drivers and passengers, and that speaking drained more resources than listening. Importantly, the resources drivers allocated to driving and conversing traded-off according to the natural ebb and flow of the conversation, demonstrating that individuals allocate resources adaptively to meet dynamically evolving task demands.

Naturalistic stimuli vary widely in complexity and perceptual discriminability, and thus it is crucial to understand how stimulus characteristics affect operators' ability to process information and meet task demands. Studies using the maritime surveillance task described above varied the complexity of the decision rule (the number of features that defined a target) [26] and stimulus discriminability (the degree to which ships were obscured by passing cloud cover) [27]. Both factors (greater complexity, lower discriminability) impaired information processing (reduced accumulation rates) and caused slower and more error-prone ship classifications (Fig. 2C).

Recent work has applied EAMs to forensic and medical image discrimination using highly complex naturalistic stimuli encountered in the field (e.g., forensic fingerprint images [39], histological cell images [58,59]). One study [39] varied the amount of visual noise that was added to naturalistic fingerprint images in an image discrimination task (deciding whether a crime scene print matches a suspect). Matching prints were processed less efficiently (with lower accumulation rates) when degraded by visual noise (Fig. 2C), which produced more frequent errors, and this deficit was ameliorated by a brief training intervention. These findings have implications for the trustworthiness of crime scene-suspect fingerprint

matches produced by human decision makers and may inform training programs aimed at improving identification accuracy.

Seeking to understand diagnostic decision making in medicine, recent work [59] obtained naturalistic cell histology images that varied in perceptual difficulty (as judged by subject-matter experts) and were then judged by novices and experts as either positive or negative for pathology. Reduced DDM accumulation rates indicated that hard-difficulty images were processed less efficiently than easy-difficulty images, causing less accurate diagnoses (Fig. 2C). Additionally, experts processed information more efficiently than novices, and both novices and experts accumulated evidence more slowly for negative than positive diagnoses. These results could be used to reduce the frequency of misdiagnoses by improving diagnostician training.

Overall, these findings indicate that, in tasks that embody the complexity in which most decision-making occurs, additional task demands (e.g., PM, time pressure) require operators to redistribute limited cognitive resources. When demands divert resources away from safety-critical primary tasks, performance may be impaired—with potentially catastrophic outcomes (e.g., failing to identify potential aircraft conflicts or brake for a traffic hazard). Additionally, stimulus characteristics, such as complexity and discriminability, affect how efficiently operators process task-relevant information. Performance on tasks involving complex or poorly discriminable stimuli is especially likely to be impaired when demands venture into the 'red zone'. Importantly, this work demonstrates that EAMs offer a unified theoretical account of a complex set of behavioural effects across a wide range of naturalistic tasks and experimental manipulations.

Box 2 discusses potential connections between EAMs and other models used to understand safety-critical decisions.

Box 2. Towards an integrated theory of safety-critical decisions

As highlighted throughout this review, EAMs are advantageous over traditional analyses of mean RT and error rates because they allow researchers to disentangle effects that may otherwise be ambiguous or masked. For example, analyses of mean RT or error rates alone cannot establish why one participant is fast-but-inaccurate and another is slow-but-accurate, because shifts in these variables can arise from (combinations of) different latent processes. One participant might respond more slowly and accurately

than another due to relatively higher response thresholds (a difference of strategy not ability). An alternative explanation for accurate but slow responding is high-quality cognitive processing (high accumulation rate) but slow motor responding (long non-decision time). In fully accounting for the entire shape of RT distributions, including variability and skew, EAMs make it possible to differentiate such cases [9].

A further advantage is that EAMs allow researchers to establish formal connections with other theories of cognition (e.g., reinforcement learning) and broader cognitive and task network architectures commonly used to understand safety-critical decisions. For example, recent work has integrated EAMs with reinforcement learning (RL) models [75-78], which allow decision mechanisms (e.g., accumulation rate, threshold) to be parameterised in terms of trial-by-trial learning dynamics. RL-EAMs have successfully explained learning effects in several simple lab tasks (e.g., value-based choice, category learning), and have been shown to provide a close account of data that standard EAMs fail to fit [76]. By incorporating RT distributions, this approach substantially extends the explanatory scope of, and constraint upon, models of learning. Practically speaking, these models can improve our understanding of how operators perform in dynamic work settings that require adapting decision making from moment-to-moment (e.g., as new information is learned, as critical events unfold).

Similar opportunities exist for integrating EAMs with more general cognitive architectures (e.g., ACT-R [79], SOAR [80]), task network models (e.g., IMPRINT [81]), and multiple-goal pursuit models (e.g., MGPM [82]) that explain how operators prioritise the allocation of time and effort as they pursue a set of competing goals with different deadlines. These models explain task scheduling and goal prioritization but are largely silent in modelling the dynamics of individuals choices. Considered as a 'front end' model that explains choices and RTs, EAMs could bring these models in closer contact with empirical performance data, allowing for more detailed predictions of safety-critical decisions and stronger tests of competing theories.

Taking control of cognition: Proactive and reactive decision control

In the preceding section, we reviewed effects arising due to either limitations on the human operator's processing capacity or limitations on the information available from the task environment. In this section, we discuss the ways in which operators exert cognitive control to meet specific task demands and prioritize different goals. Flexible adaptation depends upon this "ability to regulate, coordinate, and sequence thoughts and actions in accordance with internally maintained behavioral goals" [60, p. 1]. According to the dual-mechanisms framework, cognitive control comes in two forms: proactive and reactive [60]. Proactive control is volitional control engaged before a cognitively demanding event or change in task demands to bias the cognitive system in a goal-driven manner. Key to proactive control is

that it is deployed to be already active when the target event/context occurs. Reactive control, in contrast, is automatic, event-driven control engaged after the onset of a target event/context to influence responding "only as needed, in a just-in-time manner" [60, p. 2]. These two control modes allow operators to flexibly adapt to changes in task demands and task priority that often occur in dynamic decision-making contexts (e.g., a maritime surveillance operator adopting a more conservative threshold for classifying enemy targets when friendly forces are nearby; an air-traffic controller strategically shifting bias toward making conflict responses when under time pressure to ensure aircraft remain separated [34]). As we will demonstrate, EAMs provide a coherent framework for measuring and interpreting numerous cognitive control effects.

In EAMs, key loci of proactive control strategies are threshold and bias settings, since these are set by the operator in advance of stimulus onset (i.e., it is circular for the amount of evidence used to identify a stimulus to depend upon the identity of that stimulus). Several studies have used threshold and bias to quantify how decision makers use proactive control to adapt to changes in demands. For example, individuals detecting aircraft conflicts use proactive control to adapt to PM demands and time pressure [23,24], albeit in different ways. When facing tighter deadlines, participants set lower thresholds to ensure that responses were executed before the deadline (Fig. 3A). In contrast, when given a concurrent PM task, participants set higher conflict detection thresholds, which delayed conflict detection responses relative to PM responses (Fig. 3B). Model simulations indicated this strategy allowed individuals to avoid pre-empting PM responses (if appropriate) and thus achieve higher PM accuracy. However, with tighter deadlines, this slowing strategy led to a substantial increase in non-responses (responses not executed before the deadline) that would be unacceptable for controllers in the field. Similar modelling of PM in the maritime surveillance task converged with these results: Individuals adapted to PM demands by setting higher ship classification thresholds to avoid pre-empting the atypical PM responses [28].

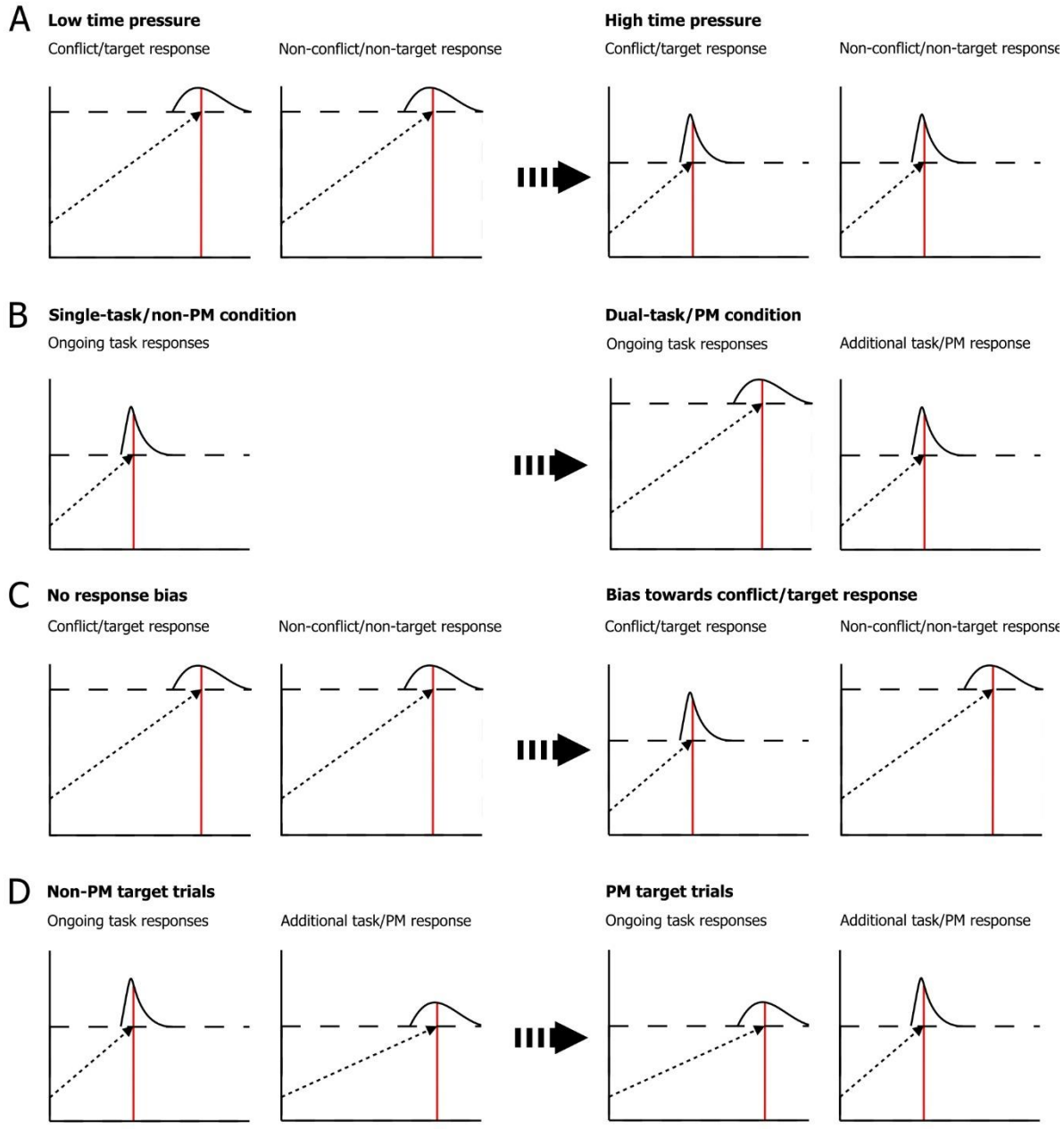

Figure 3. EAM accounts of proactive and reactive control effects discussed in text.

In each accumulator diagram, black arrows represent mean accumulation rate. Atop each threshold (dashed horizontal lines) is an RT distribution whose shape reflects the illustrated task/parameter effect. Red vertical lines indicate mean RT. Note that, in panels B and D, one ongoing task accumulator is used to represent potentially multiple ongoing task responses. (A) With low time pressure, thresholds are high and responses slow and accurate. With high time pressure, thresholds are low and responses fast and inaccurate. (B) In single-task conditions (e.g., without a concurrent PM task), ongoing task thresholds are low and responses fast. In dual-task conditions (e.g., with a concurrent PM task), ongoing task thresholds are set higher. This slows down ongoing task responses, allowing more time for the PM target accumulator to reach threshold (if appropriate). (C) When responding is unbiased, similar threshold settings are used for all responses. When responding is biased, the threshold for the prioritized response is lowered, making it easier to trigger

relative to less-prioritized responses. (D) On non-PM target trials, ongoing task accumulation is fast and likely to win against the PM accumulator. On PM target trials, ongoing task accumulation is inhibited (lowered) and less likely to win against the PM accumulator.

Drivers use similar proactive control strategies to manage the demands of distracted driving. For example, drivers set higher thresholds for pressing the brake pedal when experiencing distraction (mobile phone or passenger conversation) to compensate for the poorer quality of information uptake caused by dividing attention between the conversation and events on the road [30]. DRT studies [29,31-33] show that drivers set higher DRT thresholds and respond more slowly while engaged in conversation [31,32] and when multi-tasking (counting backwards by threes) [29,33] compared to driving undistracted. Such proactive control strategies (e.g., delaying certain responses) can help compensate for additional task demands but are undesirable if operators become too slow to react to critical events.

In terms of managing time pressure, every study in this review that increased time pressure found that individuals proactively set lower thresholds to decrease RT and ensure responses were executed on-time [23-27,39,59,61,62]. Individuals adopt this strategy regardless of whether increased time pressure takes the form of tighter deadlines [23-26,62], being required to process more stimuli per unit time [25-27], or being instructed to prioritise speed over accuracy [24,39,59,61], consistent with thresholds as a general mechanism for controlling speed-accuracy trade-offs [7,63]. However, setting thresholds too low can cause unacceptably high error rates and may thus be an undesirable strategy in operational settings in which errors are extremely costly (e.g., an air-traffic controller quickly misclassifying aircraft heading for a conflict as safely separated).

Response bias is another proactive control mechanism used to adapt to task demands. In maritime surveillance, participants compensated for greater stimulus complexity (ships with more features) by shifting bias to favour classifying ships as targets (setting a lower threshold for responding ‘target’ than ‘non-target’; Fig. 3C) [26]. Similarly, in ATC conflict detection, controllers adapt to increased uncertainty (e.g., longer time to minimum separation, greater angle of convergence) and time pressure by becoming biased towards classifying aircraft as in-conflict [34,61,62,64]. This strategy ensures targets are not missed but increases false alarms. Operators are likely to employ such biases in operational settings

that prioritize safety over accuracy, particularly when misses are significantly more costly than false alarms, as they are in ATC conflict detection [34,47].

Individuals also use biases to incorporate prior knowledge and expectations about the task environment, such as the expected frequency with which stimuli occur (i.e., base rate/prevalence) [58] and the expected reliability of predictive cues [59]. In a medical image classification task, participants were given prior information in the form of a predictive cue that indicated the correct diagnosis with 65% reliability [59]. As expected, individuals shifted bias towards the cued response. Later work [58] varied the relative prevalence (presentation frequency) of healthy versus diseased cell images. Both novices and experts shifted bias towards the more prevalent diagnostic category, responding more often with the high base-rate diagnosis regardless of the identity of a given image. Although less pronounced in experts than novices, this strategy can lead to increased false positives (for high-prevalence categories) and increased misses (for low-prevalence categories) [65], either of which may be undesirable in certain contexts (e.g., medical and airport screening).

Finally, we discuss reactive control, which is engaged only as needed to deal with critical events as they occur (not engaged in advance like proactive control). Recent studies of PM in ATC conflict detection [23,24] found that, in addition to the proactive control and capacity effects outlined earlier, participants deployed reactive control upon encountering PM targets (the onset of which could not be predicted). Specifically, accumulation rates for conflict detection responses were lower when a PM target was present versus absent (Fig. 3D). That is, when the cognitive system detects features consistent with PM targets, inhibitory input slows down accumulation for the competing conflict detection responses. This increases the likelihood of the PM accumulator winning against the more habitual ongoing responses (see also [35,66]). These effects, which were not otherwise obvious, have been replicated in the maritime surveillance paradigm, where accumulation rates for classifying ships were inhibited in the presence of PM targets [28]. Additionally, it has been shown that individuals can vary the strength of reactive inhibition according to task priority: Prioritizing PM led to greater inhibition of conflict detection responses compared to when conflict detection was prioritized [24].

Interesting practical consequences of reactive inhibitory control were identified in a study in which ATC conflict detection participants were provided with an imperfectly reliable

automated decision aid [67]. In automation blocks, the decision aid advised whether the aircraft displayed were in conflict. Behaviourally, incongruent responses (responses disagreeing with the decision aid) were slower and less frequent relative to responses made on matched trials from manual blocks. Crucially, EAM analysis showed that this pattern could be explained in terms of a single inhibition mechanism, whereby decision aid advice inhibited accumulation rates for incongruent responses. Practically, this strategy allowed operators to integrate automated advice whilst still requiring them to process task-relevant information to trigger a response, ensuring actions were not initiated solely on potentially erroneous automated advice. In situations without time pressure, this strategy increases accuracy, making decision aids desirable. However, with high time pressure, inhibited responses may become unacceptably slow.

In sum, individuals use a variety of cognitive control strategies to flexibly adapt to changes in operational demands. They use proactive control to adjust threshold and bias settings to manage anticipated demands (e.g., time pressure, PM demands, stimulus complexity and prevalence). They use reactive control to influence processing (e.g., inhibiting accumulation rates for competing/incongruous responses) only as needed (e.g., when encountering PM targets, when automation gives incongruous advice). These findings highlight that, although it is undoubtably important to understand how cognitive resource limits constrain operators' ability to meet operational demands, Human Factors practitioners must also understand the broader array of proactive and reactive strategies operators use to adapt to task demands. EAMs provide a unified framework for disentangling all these processes, which holds enormous potential for Human Factors/Ergonomics research. Box 3 outlines several best practices for getting the most out of EAM analyses.

Box 3. Getting the most out of EAMs with good modelling practices

When developing a new model or applying an existing model to a novel or naturalistic task, researchers must ensure that their model produces valid and generalisable inferences. To this end, we outline several best practices for modelling that should form part of any thorough model-based analysis (for more detailed discussion, see [1,15,83]).

One common question concerns model complexity. A model should not fail to capture important trends in the data (under-fitting) but also not be so complex as to capture spurious or idiosyncratic variation (over-fitting). To avoid both under- and over-fitting, researchers should 'bookend' selected models with more and

less flexible model variants to establish upper and lower bounds on model complexity. This can help find the model that most parsimoniously describes the data.

Another issue closely related to complexity and over-fitting is that of generalizability, or how well a model predicts new data (data not used in model fitting). To encourage generalizability, researchers can incorporate cross-validation techniques into their model fitting procedures (e.g., fitting the model to a subset of data and predicting the withheld portion) or conduct simulation studies that test the predictive validity of their models.

Once an appropriate model has been selected, researchers should conduct parameter recovery studies to establish whether the model produces reliable inferences and to diagnose weakly identified (unreliable) parameters or model mechanisms. This is done by fitting the model to simulated data and assessing whether the model recovers the known data-generating parameters. Good recovery indicates a reliable model. Poor recovery points to potentially unreliable model mechanisms and may suggest ways of improving future experimental designs (e.g., increased trial numbers, stronger experimental manipulations).

Further confidence in model-based inferences can be obtained by comparing several different models that instantiate the same cognitive theory, and by replicating results across multiple studies. For example, at least one study featured in this review compared the same theories instantiated in both the LBA and DDM frameworks [27] and several included replication studies [24-28]. Across models and studies, points of agreement provide convergent validity and increase confidence in inferences. Points of disagreement indicate where more caution should be exercised in interpreting a theory and may suggest avenues for further research and theoretical development.

Concluding remarks

In this article, we reviewed a recent body of work that has brought EAMs 'into the wild' by applying them to tasks that embody the complexity and demands of modern workplaces. Across many complex work tasks, EAMs provide a coherent account of the latent cognitive mechanisms that drive human performance. Accumulation rates can be used to measure attention and cognitive capacity, and to identify points at which demands compromise performance and safety. Threshold, bias, and rate parameters can identify proactive and reactive cognitive control strategies that operators use to meet both expected and unexpected changes in demands. Human Factors practitioners can use such insights to improve operator training and task design, develop automated support tools, and identify when operators risk entering the 'red zone'. As discussed, many exciting possibilities exist for using EAMs to inform important current applied topics, such as automation reliance, and

for integrating EAMs with learning models and broader cognitive architectures to further understand operator performance (see Outstanding Questions). In conclusion, the application of computational cognitive models like EAMs to representative tasks carries reciprocal benefits for applied and basic research. Human Factors research benefits from more detailed analyses of latent cognitive processes provided by formal modelling. Equally, theoreticians benefit from understanding how their cognitive theories generalise to the complex and demanding environments in which most decision making occurs. We therefore encourage more researchers to bring computational cognitive models out of the lab and into the wild.

Outstanding questions

- Can EAMs be used to track operators' cognitive state in real time to predict when support is needed (e.g., during periods of high workload or operator fatigue)? Can EAM's detailed measurement of latent cognitive processes open the door to providing operators with individualised support (e.g., different interventions when slow responses are due to impaired accumulation versus an overly cautious response strategy)?
- Can integrating EAM's account of choice and RT into more general cognitive or task network architectures give insight into performance in even higher-fidelity work tasks involving more continuous, evolving stimuli and events with less predictable onsets and durations? Could such integrated models enable researchers to understand and predict system-level work performance in complex work tasks? Can EAMs that incorporate learning mechanisms account for moment-to-moment changes in operators' cognitive state due to task experience and adaptation?
- What can EAMs tell us about how operators handle multiple concurrent goals with different deadlines, that additionally may vary along several dimensions relevant to operator motivation (e.g., the value of achieving the goal, the work required to reach the goal, how quickly progress can be made towards the goal, and whether the goal entails approaching a desired state or avoiding an undesired state)? Can such knowledge inform the development of automated scheduling algorithms that help operators allocate their time efficiently?
- How can EAMs be used to inform personnel selection and training for cognitively demanding safety-critical work tasks? Can the detailed picture of latent cognitive abilities provided by EAMs be used to select (exclude) candidates who possess (lack) certain cognitive abilities, such as efficient information processing, or to train candidates found to be using suboptimal decision-making strategies, such as setting thresholds too low in situations that require high accuracy?

Acknowledgments

This work was supported by an ARC Discovery Project grant (DP200101842) awarded to S.L. and L.S and DP2101003130 awarded to A.H.

References

1. Farrell, S. and Lewandowsky, S. (2018) *Computational modeling of cognition and behavior.* Cambridge University Press
2. Forstmann, B.U. *et al.* (2016) Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology* 67, 641-666
3. Ratcliff, R. *et al.* (2016) Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences* 20, 260-281
4. Evans, N.J. and Wagenmakers, E.-J. (2020) Evidence accumulation models: Current limitations and future directions. *Quantitative Methods for Psychology* 16, 73-90
5. Gold, J.I. and Shadlen, M.N. (2007) The neural basis of decision making. *Annual Review of Neuroscience* 30, 535-574
6. Schall, J.D. (2019) Accumulators, neurons, and response time. *Trends in Neurosciences* 42, 848-860
7. Mulder, M. *et al.* (2014) Perceptual decision neurosciences–a model-based review. *Neuroscience* 277, 872-884
8. Donkin, C. and Brown, S.D. (2018) Response times and decision-making. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, pp. 349-377, John Wiley & Sons
9. Lerche, V. and Voss, A. (2020) When accuracy rates and mean response times lead to false conclusions: A simulation study based on the diffusion model. *The Quantitative Methods for Psychology* 16, 107-119
10. Forstmann, B.U. *et al.* (2011) Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in Cognitive Sciences* 15, 272-279
11. Forstmann, B.U. and Wagenmakers, E.-J. (2015) Model-based cognitive neuroscience: A conceptual introduction. In *An Introduction to Model-based Cognitive Neuroscience*, pp. 139-156, Springer
12. Turner, B.M. *et al.* (2019) Advances in techniques for imposing reciprocity in brain-behavior relations. *Neuroscience & Biobehavioral Reviews* 102, 327-336
13. Tsang, P.S. and Vidulich, M.A. (2003) Introduction to aviation psychology. In *Principles and Practice of Aviation Psychology,* pp. 1-19, Lawrence Erlbaum Associates
14. Dismukes, R.K. (2012) Prospective memory in workplace and everyday situations. *Current Directions in Psychological Science* 21, 215-220
15. Lee, M.D. *et al.* (2019) Robust modeling in cognitive science. *Computational Brain & Behavior* 2, 141-153
16. Ratcliff, R. (1978) A theory of memory retrieval. *Psychological Review* 85, 59-108
17. Brown, S.D. and Heathcote, A. (2008) The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology* 57, 153-178
18. Donkin, C. *et al.* (2011) Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology* 55, 140-151
19. Voss, A. *et al.* (2004) Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition* 32, 1206-1220
20. Eidels, A. *et al.* (2010) Converging measures of workload capacity. *Psychonomic Bulletin & Review* 17, 763-771
21. Donkin, C. *et al.* (2014) Assessing the speed-accuracy trade-off effect on the capacity of information processing. *Journal of Experimental Psychology: Human Perception and Performance* 40, 1183-1202
22. Little, D.R. *et al.* (2019) Systems factorial technology analysis of mixtures of processing architectures. *Journal of Mathematical Psychology* 92, 102229
23. Boag, R.J. *et al.* (2019) Cognitive control and capacity for prospective memory in complex dynamic environments. *Journal of Experimental Psychology: General* 148, 2181-2206
24. Boag, R.J. *et al.* (2019) Strategic attention and decision control support prospective memory in a complex dual-task environment. *Cognition* 191, 103974
25. Palada, H. *et al.* (2019) Using response time modeling to understand the sources of dual-task interference in a dynamic environment. *Journal of Experimental Psychology: Human Perception and Performance* 45, 1331-1345
26. Palada, H. *et al.* (2018) Understanding the causes of adapting, and failing to adapt, to time pressure in a complex multistimulus environment. *Journal of Experimental Psychology: Applied* 24, 380-399
27. Palada, H. *et al.* (2016) Evidence accumulation in a complex task: Making choices about concurrent multiattribute stimuli under time pressure. *Journal of Experimental Psychology: Applied* 22, 1-23

28. Strickland, L. *et al.* (2019) Prospective memory in the red zone: Cognitive control and capacity sharing in a complex, multi-stimulus task. *Journal of Experimental Psychology: Applied* 25, 695-715
29. Castro, S.C. *et al.* (2019) Cognitive workload measurement and modeling under divided attention. *Journal of Experimental Psychology: Human Perception and Performance* 45, 826-839
30. Ratcliff, R. and Strayer, D. (2014) Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic Bulletin & Review* 21, 577-589
31. Tillman, G. *et al.* (2017) Modeling cognitive load effects of conversation between a passenger and driver. *Attention, Perception, & Psychophysics* 79, 1795-1803
32. Castro, S.C. *et al.* (2022) Dynamic workload measurement and modeling: Driving and conversing. *Journal of Experimental Psychology: Applied*. Advance online publication, https://doi.org/10.1037/xap0000431
33. Damaso, K.A. *et al.* (2022) A cognitive model of response omissions in distraction paradigms. *Memory & Cognition 50*(5), 962-978
34. Loft, S. *et al.* (2009) A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. *Journal of Experimental Psychology: Applied* 15, 106-124
35. Strickland, L. *et al.* (2018) Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review* 125, 851-887
36. Verbruggen, F. and Logan, G.D. (2009) Proactive adjustments of response strategies in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance* 35, 835-854
37. Ratcliff, R. and Rouder, J.N. (1998) Modeling response times for two-choice decisions. *Psychological Science* 9, 347-356
38. Ratcliff, R. and McKoon, G. (2008) The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation* 20, 873-922
39. Palada, H. *et al.* (2020) An evidence accumulation model of perceptual discrimination with naturalistic stimuli. *Journal of Experimental Psychology: Applied* 26, 671-691
40. Wickens, C.D. (2002) Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3, 159-177
41. Wickens, C.D. *et al.* (2015) Mental workload, stress, and individual differences: Cognitive and neuroergonomic perspectives. *Engineering Psychology and Human Performance (International Edition)*, 346-376
42. Sperandio, J. (1971) Variation of operator's strategies and regulating effects on workload. *Ergonomics* 14, 571-577
43. Loft, S. *et al.* (2007) Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human factors* 49, 376-399
44. Carver, C. and Scheier, M. (1998) *On the self-regulation of behavior.* Cambridge University Press
45. Hockey, G.R.J. (1997) Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology* 45, 73-93
46. Ratcliff, R. (2015) Modeling one-choice and two-choice driving tasks. *Attention, Perception, & Psychophysics* 77, 2134-2144
47. Loft, S. (2014) Applying psychological science to examine prospective memory in simulated air traffic control. *Current Directions in Psychological Science* 23, 326-331
48. Loft, S. *et al.* (2021) Prospective memory in safety-critical work contexts. In *Current Issues in Memory*, pp. 192-207, Routledge
49. Ball, B.H. and Aschenbrenner, A.J. (2018) The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review* 25, 1114-1122
50. Strickland, L. *et al.* (2017) Accumulating evidence about what prospective memory costs actually reveal. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, 1616-1629
51. Horn, S.S. and Bayen, U.J. (2015) Modeling criterion shifts and target checking in prospective memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41, 95-117
52. Heathcote, A. *et al.* (2015) Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological review* 122, 376-410
53. International Organization for Standardization (2016) Road vehicles–Transport information and control systems–Detection Response Task (DRT) for assessing attentional effects of cognitive load in driving (Rep. ISO 17488). Geneva, Switzerland
54. Heathcote, A. (2004) Fitting Wald and ex-Wald distributions to response time data: An example using functions for the S-PLUS package. *Behavior Research Methods, Instruments, & Computers* 36, 678-694

55. Strayer, D.L. *et al.* (2006) A comparison of the cell phone driver and the drunk driver. *Human Factors* 48, 381-391
56. Strayer, D.L. and Drews, F.A. (2007) Cell-phone–induced driver distraction. *Current Directions in Psychological Science* 16, 128-131
57. Cooper, J.M. and Strayer, D.L. (2008) Effects of simulator practice and real-world experience on cell-phone—related driver distraction. *Human Factors* 50, 893-902
58. Trueblood, J.S. *et al.* (2021) Disentangling prevalence induced biases in medical image decision-making. *Cognition* 212, 104713
59. Trueblood, J.S. *et al.* (2018) The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cognitive Research: Principles and Implications* 3, 1-14
60. Braver, T.S. (2012) The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences* 16, 106-113
61. Vuckovic, A. *et al.* (2014) A sequential sampling account of response bias and speed–accuracy tradeoffs in a conflict detection task. *Journal of Experimental Psychology: Applied* 20, 55-68
62. Vuckovic, A. *et al.* (2013) Adaptive decision making in a dynamic environment: A test of a sequential sampling model of relative judgment. *Journal of Experimental Psychology: Applied* 19, 266-284
63. Bogacz, R. *et al.* (2010) The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences* 33, 10-16
64. Neal, A. and Kwantes, P.J. (2009) An evidence accumulation model for conflict detection performance in a simulated air traffic control task. *Human Factors* 51, 164-180
65. Wolfe, J.M. and Van Wert, M.J. (2010) Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology* 20, 121-124
66. Loft, S. and Remington, R.W. (2013) Wait a second: Brief delays in responding reduce focality effects in event-based prospective memory. *Quarterly Journal of Experimental Psychology* 66, 1432-1447
67. Strickland, L. *et al.* (2021) Inhibitory cognitive control allows automated advice to improve accuracy while minimizing misuse. *Psychological Science* 32, 1768-1781
68. Logan, G.D. *et al.* (2014) On the ability to inhibit thought and action: general and special theories of an act of control. *Psychological Review* 121, 66-95
69. Ratcliff, R. and 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, 11285-11290
70. Tillman, G. *et al.* (2020) Sequential sampling models without random between-trial variability: The racing diffusion model of speeded decision making. *Psychonomic Bulletin & Review* 27, 911-936
71. Hawkins, G.E. and Heathcote, A. (2021) Racing against the clock: Evidence-based versus time-based decisions. *Psychological Review* 128, 222-263
72. Lerche, V. and Voss, A. (2019) Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research* 83, 1194-1209
73. Little, D.R. (2012) Numerical predictions for serial, parallel, and coactive logical rule-based models of categorization response time. *Behavior Research Methods* 44, 1148-1156
74. Provost, A. and Heathcote, A. (2015) Titrating decision processes in the mental rotation task. *Psychological Review* 122, 735-754
75. Miletić, S. *et al.* (2020) Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia* 136, 107261
76. Miletić, S. *et al.* (2021) A new model of decision processing in instrumental learning tasks. *Elife* 10, e63055
77. Sewell, D.K. *et al.* (2019) Combining error-driven models of associative learning with evidence accumulation models of decision-making. *Psychonomic Bulletin & Review* 26, 868-893
78. Pedersen, M.L. *et al.* (2017) The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review* 24, 1234-1251
79. Anderson, J.R. and Lebiere, C.J. (2014) *The atomic components of thought.* Psychology Press
80. Newell, A. (1992) Unified theories of cognition and the role of Soar. In *SOAR: A Cognitive Architecture in Perspective*, pp. 25-79, Springer
81. Samms, C. (2010). Improved performance research integration tool (IMPRINT): Human performance modeling for improved system design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54, 624-625
82. Ballard, T. *et al.* (2018) On the pursuit of multiple goals with different deadlines. *Journal of Applied Psychology 103*(11), 1242-1264

83. Heathcote, A. *et al.* (2015) An introduction to good practices in cognitive modeling. In *An Introduction to Model-based Cognitive Neuroscience*, pp. 25-48, Springer