Method for ATC Experiment E1

R. Boag

The Current Experiment

Our primary goal when designing this experiment was to reliably model both ongoing task and PM responses in a more complex and dynamic task than is typically used in either choice-RT modelling or PM research. To this end, our task deviates from traditional PM paradigms in several ways. First, most PM studies conducted using the traditional Einstein and McDaniel (1990) paradigm present only a small number of PM stimuli across an experiment, and as such they do not produce enough data to reliably constrain a model of PM processes. To reliably constrain our model, and increase the power of model fitting, we modified the paradigm by increasing the ratio of PM target trials to ongoing task trials to 1 in 5 (i.e., 20% of PM block stimuli were PM targets). This is a higher proportion of PM target trials than is typically used in PM literature, however, given that the present task is much more complex than traditional PM tasks (such as lexical decision), that decisions in this task unfolded over a much longer timescale (i.e., up to 10 seconds), and that overall PM accuracy was comparable to that of previous PM literature (~72%), the higher PM target ratio appears to be a reasonable modification to the task.

Second, we instructed participants to make their PM response instead of their ongoing task response, as we have done in our previous computational modelling efforts (Heathcote et al. 2015; Strickland et al., 2017). We used this type of PM because it is relevant to everyday situations in which cognitive errors are a result of failing to inhibit routine actions (Norman, 1981; Reason, 1990). In addition, this means we record only one RT and one response on every trial, allowing us to fit the standard LBA, whereas fitting data from a paradigm where both PM and ongoing task responses can be submitted to the same stimulus would require a more complex modelling approach.

Third, we modified the response key arrangements from the typical paradigm. Usually in the PM literature participants rest their fingers on the ongoing task keys and have to make a larger movement for the PM response. In terms of the LBA, this would cause uneven nondecision time between responses (via motor response production time), which adds complexity to the model that is not relevant to understanding PM processes, and neglecting unequal nondecision times between responses can lead to biased estimates of other model parameters (Voss, Voss, & Klauer, 2010). As such, we instructed participants to rest their fingers on both ongoing task response keys (with one hand), and on the PM response key (with the other hand), so we could assume an equal motor response time.

To assess the effects of time pressure and PM load on task performance, we used a repeated measures design in which participants completed a simulated ATC conflict detection task under two PM load conditions: control (no additional PM task) and PM (additional requirement to respond to PM targets), and four time pressure/trial load conditions: low load/low time pressure, low load/high time pressure, high load/low time pressure, and high load/high time pressure. Because we model all response types on both PM and non-PM trials, this design allows us to examine how ongoing task and PM response thresholds (i.e., proactive control) and accumulation rates (i.e., reactive control) might change under different time pressure and PM load conditions.

In terms of time pressure, our design is similar to research on the speed-accuracy trade-off in which

participants can choose to make faster responses at the expense of accuracy or make more accurate responses at the expense of speed. This research typically manipulates time pressure by imposing response deadlines of different durations or through task instructions that emphasise the importance of either fast or accurate responding. A benchmark finding in this literature is that higher time pressure (i.e., speed emphasis) leads people to lower their response thresholds to facilitate faster responding. Conversely, low time pressure (or accuracy emphasis) leads participants to raise their response thresholds in order to gain accuracy. We expect to replicate these findings for both ongoing task and PM thresholds.

In terms of PM load, our design is similar to Strickland (2017) and Heathcote, Loft and Remington (2015), which modelled responding to non-PM trials and found that PM costs were the result of strategic increases in response thresholds under PM load rather than decreased accumulation rates predicted by capacity-sharing theories of PM cost. Consistent with these results we expect thresholds for both ongoing task (conflict/nonconflict) responses to be higher during PM blocks than in control blocks. Concerning accumulation rates under PM load, Strickland (2017) found that ongoing task accumulation rates were lower on PM target trials compared to non-PM trials, suggesting the PM detector inhibits competing ongoing task responses via an automatic, stimulus-driven, 'reactive' control process. We expect to replicate these findings, although given the more complex and dynamic nature of our task and that we used a completely non-focal PM cue (non-focal PM cues generate less reactive inhibition), it is not immediately clear whether these effects will replicate as strongly.

This design also allows us to test the competing predictions of capacity-sharing and arousal/task engagement accounts of how ongoing task accumulation rates should be affected by time pressure and PM load on non-PM trials. Under high time pressure (i.e., shorter deadline/less time to respond) there is less time available to process task information before making a response. That is, more information per unit time must be processed under high time pressure compared to low time pressure conditions. Capacity-sharing accounts therefore predict slower accumulation rates under high time pressure, and faster accumulation rates under low time pressure. Similarly, the addition of a concurrent PM task means more information per unit time must be processed during PM blocks compared to control blocks. Here again capacity-sharing accounts predict slower accumulation rates when additional PM demand is present, and faster accumulation when PM demand is absent.

In contrast, from an arousal or task engagement perspective, increasing time pressure might be expected to create a more cognitively engaging task environment (particularly in a relatively monotonous, well-practiced task such as this) in which participants deploy more cognitive resources toward performaing the task during high pressure than low pressure blocks. This deployment of extra resources would be reflected in faster accumulation rates during high time pressure and slower accumulation rates during low time pressure, opposite to the predictions of the capacity sharing account. Likewise, the addition of a concurrent PM task may also make the ongoing task subjectively more engaging, leading more cognitive resources to be deployed under PM load. As before, an effort/arousal account would thus predict faster accumulation with additional PM demand, and slower accumulation without additional PM demand. Since these kinds of accumulation rate differences have not been looked at in a task of this nature, and given both sets of predictions seem plausible, we make no *a priori* predictions regarding how accumulation rate might change with time pressure. To summarise, we expect the following: higher ongoing task thresholds under PM load (proactive control), lower ongoing task and PM thresholds under higher time pressure, and lower ongoing task accumulation rates on PM-target trials than on non-PM trials (reactive inhibition). We make no predictions regarding whether non-PM trial accumulation rates will be increase or decrease with time pressure or PM load.

Method

Participants

Two participants were excluded (see Results), with 47 participants remaining (31 females). Ages ranged from 18 to 62 years ($M = 25.19$, $SD = 9.99$). Participants completed one two-hour testing session.

Materials

ATC Conflict Detection Task

The ATC conflict detection task was programmed using the ATC-Advanced Air-Traffic Control Simulator (Fothergill, Loft & Neal, 2009). The task involved classifying pairs of moving aircraft as either 'conflict' or 'nonconflict' depending on whether the aircraft would violate a 5 nautical mile (NM) minimum separation distance at some point during their flight. On some trials aircraft also contained a rare PM cue which required execution of an atypical PM response instead of the more frequent conflict/nonconflict ongoing task response.

During the simulation, the visual display included a countdown timer indicating seconds remaining in the trial and a 10 by 20NM (2 by 4cm approx.) scale marker to be used as a reference when judging distance. An information block showing callsign, airspeed, and flight level (altitude) information tracked alongside each aircraft, which appeared on screen as small circles. Each aircraft had a probe vector line which showed the aircraft's heading and predicted position one minute into the future. All aircraft appeared within a circular ATC sector boundary and followed converging straight-line flight paths (indicated by black lines) which always crossed over at a 90-degree angle in the center of the display. We now describe the spatial parameters of the aircraft stimuli used in the ATC simulation, the PM cue, and experimental design.

Aircraft Stimuli

We generated a set of unique conflict and nonconflict aircraft stimuli for each participant. To create the different conflict and nonconflict stimuli, each aircraft pair was given a distance of minimum separation (DOMS) either less than or greater than the 5NM separation standard. DOMS for conflict stimuli were drawn from the uniform distrubtion [0,3] NM. DOMS for nonconflict stimuli were drawn from the uniform distribution [7,10] NM.

In addition to DOMS, we allowed several other spatial properties of each aircraft to vary randomly during stimulus generation. Specifically, airspeed, direction of approach, time to minimum separation (TTMS), and order of passing (i.e., faster aircraft first vs. slower aircraft first) all varied randomly between stimuli. This was done to avoid instance learning, which can cause problems for modelling by introducing non-stationarity in model parameters, and may have resulted in participants performing the task using a different set of cognitive processes (e.g., memory for prior instances) than the ones we are interested in. The angle of approach between aircraft was fixed at 90 degrees to avoid interactions between angle and perceived conflict status (Loft, Bolland, Humphreys & Neal, 2009). Finally, the flight level for all aircraft was fixed at 37,000 feet. Table 1 specifies the range of values for each spatial variable and the distribution they were drawn from to generate the stimuli.

Table 1: Range and Distribution of Spatial Variables of Aircraft Stimuli

Spatial Variable	Distribution	Lower	Upper	Units
DOMS (Conflicts)	Uniform	0	3	NM
DOMS (Nonconflicts)	Uniform	7	10	NM
Airspeed	Uniform	400	700	Mph
Angle of approach	Constant	90	90	Degrees
Direction of approach	Uniform	0	360	Degrees
Flight level	Constant	37,000	37,000	Feet
TTMS	Uniform	120	210	Seconds
Order of passing	Bernoulli	0	1	0 = fastest first, 1 = slowest first

PM Cue

Aircraft with callsigns containing two of the same letter (e.g., APA169, RTR451) were designated as PM targets. This is an ecologically valid PM cue, since ATCos may be asked to look out for a specific flight number and perform an intended action (e.g., put the aircraft in a holding pattern) when that cue occurs. We note that this PM cue is completely non-focal, meaning the evidence used to make PM decisions (i.e., particular letters in an aircraft callsign) is independent of evidence used to make ongoing task conflict/nonconflict decisions (e.g., airspeed, relative distance, and motion information). In our experiment, 20% of PM block trials contained a PM target. On PM target trials only one of the aircraft on screen ever contained a PM cue, never both. Participants were instructed to respond to PM targets by pressing an alternate PM key instead of the typical ongoing task (i.e., conflict/nonconflict) keys.

NASA-TLX

Subjective task demand was assessed using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The NASA-TLX comprises six items: three that tap different aspects of task demand (i.e., mental, physical, and temporal demand) and three which tap the operator's subjective level of exertion and task performance (i.e., performance, effort, and frustration). Each item is rated on a 21-point Likert scale, with higher scores being indicative of a more demanding or effortful task. The NASA-TLX is a sensitive, reliable, and valid measure of workload (Hart & Staveland, 1988; Xiao, et al., 2005), and has been used in a wide variety of basic and applied tasks (Hart, 2006).

Experimental Design

Participants performed 8 blocks of trials consisting of four 80-trial control blocks paired with four 240-trial PM blocks, completed in alternating sequence. The order in which blocks were presented (i.e., Control or PM first) was counterbalanced across participants.

In control blocks, participants were presented with a randomised sequence of 40 conflict aircraft pairs and 40 nonconflict aircraft pairs, with no PM stimuli. In PM blocks, participants were presented

with a randomised sequence of 120 conflict aircraft pairs and 120 nonconflict aircraft pairs. In addition, a random 24 conflict pairs also contained a PM cue, and a random 24 nonconflict pairs also contained a PM cue, resulting in a total of 48 PM targets in each PM block. Thus 1/5 (48/240) PM block stimuli were PM targets.

Our time pressure and trial load factors were also blocked. Specifically, each Control-PM block pair had an associated trial load of either 2 or 5 decisions per trial (i.e., 2 or 5 aircraft pairs presented per trial) crossed with an associated level of time pressure (i.e., low or high - corresponding to either a long or short response deadline). This resulted in 4 unique trial load by time pressure combinations, with presentation order counterbalanced across participants. Table 2 shows the details of our time pressure by trial load manipulation.

We note that time pressure was not crossed orthogonally with trial load. That is, under low trial load (2 decisions per trial), low time pressure corresponded to a response deadline of 12 seconds (i.e., 6 seconds per decision on average) while high time pressure corresponded to a response deadline of 8 seconds (i.e., 4 seconds per decision on average). In contrast, under high trial load (5 decisions per trial), low time pressure corresponded to a response deadline of 20 seconds (i.e., 4 seconds per decision on average) while high time pressure corresponded to a response deadline of 10 seconds (i.e., 2 seconds per decision on average).

Table 2: Details of Trial Load and Time Pressure Manipulation

Block	Trial Load	Time Pressure	Decisions per Trial	Response Deadline (s)	Seconds per Decision
1	Low	Low	2	12	6
2	Low	High	2	8	4
3	High	Low	5	20	4
4	High	High	5	10	2

Procedure

Each testing session consisted of a training phase and an experimental phase which altogether took approximately 2 hours to complete. During the training phase participants received verbal instructions about the ATC task, watched an on-screen demonstration of the task environment, and completed a block of 40 training trials which included corrective feedback after each response. During the experimental phase participants completed eight blocks of experimental trials.

For the ATC task, participants' primary task was to judge whether pairs of aircraft would violate the 5 nautical mile (NM) minimum separation standard at any point during their flight. Participants were instructed that aircraft passing within the 5NM distance were defined as conflicts and required a *conflict* key-press response, whereas aircraft passing outside of that distance were defined as nonconflicts and required a *nonconflict* key-press response.

Participants were told that each aircraft pair would be presented sequentially (i.e., only two aircraft would appear on screen at a time), that all aircraft would be moving towards each other on converging flight paths which crossover in the center of the display, and that a number of spatial properties of the aircraft would vary from trial to trial, including their relative airspeeds, distance of minimum separation, and starting distance from the central crossing point. Participants were instructed to

consider these variables when forming their decisions and to refer to the on-screen distance scale and airspeed displays in order to make the best judgment possible.

Before each block of trials, participants saw visual instructions reminding them of the trial time limit (time pressure) and the number of aircraft pairs to be presented in each trial (trial load). Depending on the PM block, participants received either control or PM instructions. Before control blocks, participants were instructed that they only needed to make conflict/nonconflict responses for that block. Before PM blocks, participants were instructed to press a PM repsonse key instead of the conflict or nonconflict keys when they saw a PM target. Participants completed a short distractor task and saw a final reminder to respond as quickly and accurately as possible, after which the primary ATC task trials began.

Four response key assignments were counterbalanced across participants; 1) s = conflict, d = nonconflict, j = PM, 2) d = conflict, s = nonconflict, j = PM, 3) k = conflict, j = nonconflict, d = PM, and 4) j = conflict, k = nonconflict, d = PM. Participants were instructed to rest their fingers on their particular response key combination throughout the task. A screen with the text 'Press [Space] to continue' preceeded each trial, and each trial began once the spacebar was pressed. Within each trial, pairs of aircraft were presented sequentially; each pair disappearing from the screen once a response was made.

Trials ended when either all aircraft pairs had been responded to (2 pairs during low-load trials; 5 pairs during high-load trials) or when the response deadline expired (i.e., the timer counted down to zero). Aircraft pairs not responded to within the response deadline were recorded as nonresponses. Aside from the training trials, no further feedback was given concerning task performance. Participants completed a NASA-TLX workload questionnaire after each block. Participants took self-paced breaks between each block of trials and were also permitted short breaks at any point between-trials if required.