

Impact of a Multiobject-directed Spatial Orientation Ability on Air Traffic Conflict Detection Performance in Free Route Airspace

Jimmy Y. Zhong

Nanyang Technological University

Sim Kuan Goh

Nanyang Technological University

Chuan Jie Woo

Nanyang Technological University

Sameer Alam (✉ sameeralam@ntu.edu.sg)

Nanyang Technological University

Research Article

Keywords: multiobject-directed spatial orientation, spatial orientation ability (SOA), free route airspace (FRA)

Posted Date: March 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-258016/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Impact of a multiobject-directed spatial orientation ability on air traffic conflict detection performance in free route airspace

Jimmy Y. Zhong¹, Sim Kuan Goh¹, Chuan Jie Woo¹, and Sameer Alam¹

¹Air Traffic Management Research Institute (ATMRI), School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore.

ABSTRACT

With a focus on psychometric assessment, the current study investigated the extent to which spatial orientation ability (SOA), as conceptualized in the spatial cognition and navigation literature, predicted air traffic conflict detection performance in a simulated free route airspace (FRA). Within a FRA, airspace users have the flexibility to plan flights by selecting preferred routes between predefined waypoints. Despite such benefits, FRA implementation can introduce conflicts that are geometrically complex, and of which would require a high level of SOA engagement. Based on a sample of 20 young adults who have the prospect to become air traffic controllers (ATCOs), we found that response time-based performance on a newly developed computerized spatial orientation test (SOT) predicted time to loss of minimum separation (tLMS)-based performance on a conflict detection task to a moderately large extent under scenarios with high air traffic density. We explained these findings in light of similar or overlapping mental processes that were most likely activated optimally under task conditions featuring approximately equal numbers of outcome-relevant stimuli. We also discussed the potential use of the new SOT in relation to the selection of prospective ATCOs who can demonstrate high levels of conflict detection performance in FRA during training simulations.

Introduction

Over the past few years, there has been a significant amount of human factors studies in the Air Traffic Management (ATM) domain that focused on (i) situation awareness^{1,2}, (ii) mental workload^{3,4}, and (iii) spatial attention^{5,6}. This is perhaps unsurprising considering that ATM-related human factors research is largely concerned with understanding the psychological processes that support interactions between humans and ATM systems, and how theories, principles, data and methods can be used to optimise human well-being and overall system performance⁷. By contrast, psychometrics is a field of study in psychology concerned with the theory and methods of psychological measurement – a specialized field that holds the design and deployment of tests for an objective evaluation of psychological attributes (e.g., attitudes, emotions, cognitive abilities, personality) as part of its top priorities^{8,9}. In the ATM domain, psychometric testing of candidates applying for Air Traffic Controller (ATCO) positions is a commonplace phenomenon that has been done by various air navigation service providers (ANSPs) such as European Organization for the Safety of Air Navigation (EUROCONTROL)^{10–12} and Federal Aviation Administration (FAA), USA^{13,14}. ATCO selection by these aviation organizations involve psychometric test batteries evaluating one's mental capabilities spanning across spatial attention and reasoning, visual perception, and working memory^{10,11,14}, alongside tests of aptitude for air traffic control (ATC) and occupational knowledge of the aviation field^{13,14}. The administration of these tests is important because it increases the probability of selecting candidates who are likely to pass formal ATC training and move on to become professional ATCOs^{10,14}.

Compared with the aforementioned developments in ATM-related human factors research, not many ATM-related psychometric studies had been done over the past few years¹⁰, and the current study was conducted with an intention of contributing to the latter field of research by investigating the extent to which spatial orientation ability (SOA) could predict traffic conflict detection performance under simulated ATC scenarios. By "air traffic conflict detection," we refer to an ATCO's professional task of detecting potential safety infringement events (i.e., "conflicts") in which a breach of separation minima occurs between in-flight aircraft in a controlled airspace¹⁵ (see Methods, for specific details). Professional competence in detecting such safety infringement or conflict events is extremely important because any failure to do so can lead to near miss events that trigger short term conflict alerts (STCAs) [for ATCOs] and the traffic collision avoidance system (TCAS) [for aircrews], or in worse cases, aviation accidents resulting from late aircraft maneuvers¹⁶. Before explaining the purpose for investigating the impact of SOA on conflict detection, we shall first state the definition and primary features of SOA.

In this paper, we described SOA with respect to the commonly accepted definition used in the spatial cognition and navigation literature. Specifically, SOA is defined as a specific spatial ability that allows one to: (i) imagine how an object appears relative to surrounding objects from a particular orientation or perspective and (ii) make directional judgments from the standpoint inherent to that imagined perspective^{17–24}. As the imagination of how one object appears relative to other objects requires the visualization of a particular perspective or heading direction, SOA has been widely referred to as *perspective-taking ability* (PTA) in the spatial cognition and navigation literature^{17–24}. Notably, through the use of confirmatory factor analysis on performance data collected from a large sample of participants, SOA has been shown as a unique type of spatial ability that is separable from two other types of component spatial abilities: (i) *speeded mental rotation* and (ii) *spatial visualization*²⁰. The former requires performing mental rotations of a schematically drawn object [either two-dimensional (2D) or three-dimensional (3D)] while the latter involves executing a series of spatial operations or transformations on the same type of object²⁰. Principally, speeded mental rotation and spatial visualization focused on the execution of mental operations on single, discrete objects, and have thus been conceptualized as object-centric spatial abilities^{20,22}.

In contrast to these two types of spatial abilities, SOA involves mental simulation of positioning oneself among multiple objects^{17,18,20,23,24} and making quick judgments of inter-object spatial relations that vary with imagined changes in orientation or perspective^{17–24}. In view of these spatial reasoning processes, we characterized SOA, within the context of this study, as a *multiobject-directed* spatial ability – an ability that incorporates the spatial perception and cognition of multiple objects^{20–22}.

Research Gap

In the ATM domain, SOA has been regarded by both ATCOs^{25,26} and aircraft pilots^{25,27} as one of the most important cognitive abilities for operational competence. Performance-wise, experimental studies involving spatial ability assessment have shown that higher levels of SOA were associated with (i) higher situation awareness of downlinked messages among professional ATCOs²⁸, (ii) improved management of ground level traffic among ATCO applicants²⁹, and (iii) better aircraft handling performance during different flight phases (i.e., departure, en route, arrival) among university students and employees who can become prospective ATCOs.^{30–32}

Previous ATM studies that involved psychometric testing mainly involved spatial orientation tests (SOTs) that required: (i) judgments of lateral heading directions (i.e., differentiating between left and right turning directions when following a schematically drawn route)²⁸, and (ii) perspective-taking focused on a schematically drawn 3D object (e.g., visualizing how a 3D cube-like object would appear from different viewpoints^{30–32}). Notably, the SOT currently administered by EUROCONTROL ("spot the side") focuses on lateral directional judgments, requiring test takers to take the perspective of an on-screen avatar and pinpoint the location of a geometric object (left or right) that is held by one of his/her outstretched hands^{33,34}. This test is conceptually similar to the aforementioned test requiring lateral heading judgments²⁸ and can be seen theoretically as an elementary test of egocentric PTA.²⁰

These aforementioned SOTs showed that alternative types of SOTs involving a wider array of objects and presenting higher orders of task complexity were absent in the ATM domain. This could be explained by the fact it has taken almost two decades of intensive psychometric research, involving large samples of participants, to provide strong support for the reliability and validity of SOTs featuring multiple objects (or landmarks)^{17–24,35–39}. In general, these SOTs require participants to execute pointing responses toward multiple objects (or landmarks) arranged in an array with reference to an initial heading direction^{19–24,39}. Some of these SOTs were programmed to provide sensitive recordings of pointing performance on computers and such computerized SOTs have been shown to correlate significantly with real-world landmark pointing performance that reflected stored environmental knowledge acquired from route learning^{23,24}.

On the other hand, studies from the ATM domain showed that performance on a classical paper-based version of these computerized SOTs²⁰ did *not* predict how successful ATCOs were in handling aircraft for safe landings⁴⁰. Likewise, performance on a paper-based SOT involving the judgments of left and right heading changes along a 2D route did not predict ATCOs' conflict detection performance²⁸. These negative findings could be explained by the fact that aircraft handling (i.e., monitoring and sequencing aircraft for safe departures, hand-overs, and arrivals) relies more on other cognitive abilities like spatial attention and mental rotation than on SOA⁴⁰ and that the cognitive processes involved in making lateral directional judgments²⁸ did not match well with higher-order attentional and spatial reasoning processes required for air traffic conflict detection^{41,42}. Conceptually, these negative findings were crucial for highlighting the notion that there needs to be a convergence or overlap of similar mental processes (perceptual and/or cognitive) between a spatial ability test and an ATC task in order for the former to predict performance on the latter⁴⁰.

Study Motivation and Aims

In view of the aforementioned SOTs in the ATM psychometric literature that presented small numbers of objects requiring judgments of changes in orientation and that there are currently no multiobject-based SOT that can predict air traffic conflict detection performance, we aimed to assess conflict detection through the use of a computerized SOT that was recently developed

by Friedman et al. (2020) and made freely accessible to researchers from all disciplines¹⁹. This new SOT was designed with respect to the notion of SOA as a multiobject-directed and navigationally relevant construct that involves judging orientation or perspective changes^{21,23,24,35,39}. Compared with its paper-based predecessor²⁰, this computerized SOT offers greater ease of stimuli presentation and higher precision in data recording (i.e., recording of pointing responses to the accuracy of two decimal places, see Methods, for more details).

To our knowledge, the current study is the first that applied this new SOT to air traffic conflict detection. Prior to this study, several studies had directed this new test successfully to investigations of individual differences in SOA and environmental learning^{18,43,44}. In this study, we chose conflict detection as the criterion variable because conflict detection requires spatial attention to and simultaneous information processing of multiple aircraft^{41,42,45}, mental processes which we expected one to engage equally well in when he/she observes an array of 2D objects on the SOT and visualize how they relate spatially to one another.

A more important and practical reason for relating SOA to conflict detection concerns recent technological advances in Space-Based Automatic Dependent Surveillance-Broadcast (SB ADS-B)⁴⁶, which paved the way for the deployment of free route airspace (FRA)^{47,48}. Compared with current fixed route networks, FRA offers greater flexibility for airspace users (i.e., airlines) to plan flights through the selection of preferred routes between predefined entry and exit points in the face of few constraints (e.g., avoidance of dangerous areas and compliance to fixed entry and exit points)^{47,49}. Notably, FRA has been implemented successfully in Europe, providing greater capacity for sustaining air traffic flow while reducing CO₂ emission, flight time, and fuel waste^{48–51}. However, the benefits of FRA carries with them the potential of introducing conflicts that are geometrically more complex than those in a fixed route network due to increased number and distribution of conflict points, as well as elevated uncertainty in aircraft maneuvering (e.g., in the event of changes in aircraft heading or orientation)^{50,52}. These issues will make conflict detection more challenging for ATCOs, and consequently, it will be prudent for ANSPs to select ATCOs with suitably high levels of SOA in order to ensure flight safety under FRA implementation. Putatively, such ATCOs would be better positioned to detect unexpected aircraft heading or orientation changes that can lead to conflicts in a FRA.

Specifically, we investigated the extent to which SOA, as conceptualized and measured in the recent spatial navigation literature, could predict conflict detection performance among a sample of university students and employees who are eligible for application as ATCOs. For a detailed examination of this predictive relationship, we introduced a mental workload component of *air traffic density*⁴⁵ to explore how this predictive relationship could change with varying numbers of aircraft in a simulated, controlled airspace (see Methods, for details). Like what Friedman et al. (2020) did, we also examined the concurrent validity of the new SOT¹⁹ in relation to a classical paper-based SOT, the Money Road Map Test⁵³, that requires quick judgments of lateral heading changes. By investigating all these relationships, we aimed to provide novel findings to argue for the relevance of this new SOT for predicting conflict detection performance in a simulated FRA environment among potential ATCOs. By directing a SOT whose design has been well-conceptualized by spatial navigation researchers to ATM research, we also aimed to promote more communication and collaboration between researchers from the spatial navigation and ATM domains.

Methods

Participants

20 young adults (11 females, nine males), ranging from 19 to 31 years of age ($M = 23.65$, $SD = 2.96$), participated in the study. This participating group comprised university students (undergraduates and postgraduates) and employees. They were recruited through email invitations and advertisements posted on notice boards in the university campus. All participants possessed normal or corrected vision (20/40 vision or better) and were found to be right-handed. The majority of participants ($n = 17$) majored in the hard sciences (i.e., Aerospace Engineering, Electrical Engineering, Computer Science). The remaining participants ($n = 3$) majored in the social sciences (i.e., Business Administration, Psychology, Sociology). Age and educational qualifications of all participants met the entry requirements for ATCO application specified by the Civil Aviation Authority of Singapore (CAAS)⁵⁴. Informed consent for study participation was approved by the Institutional Review Board (IRB) of Nanyang Technological University (NTU) [Reg. no.: 200604393R]. Signed consent was obtained from each participant prior to his/her participation. All tasks and procedures mentioned below were carried out in accordance with relevant ethical and institutional guidelines.

Spatial Orientation Test (SOT)

The SOT used in this study was a computerized test developed recently by Friedman et al. (2020)¹⁹. It was packaged as a Java applet made freely available for download and research use under the auspices of the international Creative Commons 4.0 (CC BY 4.0) license provided by the Open Science Framework (OSF), Center for Open Science, Charlottesville, VA, USA (<https://osf.io/t58ab/>).

A 15.6-inch LCD monitor with an aspect ratio of 16:9 (13.60 inches x 7.65 inches, equivalent to 34.51 cm x 19.42 cm) was used for presenting the SOT. The task was shown in full screen at a resolution of 1,920 x 1080 pixels and a refresh rate of 60

Hz. Each participant sat 20 inches (50.8 cm) away from the center of the screen while performing the task. The task comprised three practice trials and 12 test trials. On each trial, participants viewed an array of seven inanimate objects whose locations were kept fixed across all trials. A drum was depicted at the center of the array surrounded by six other objects, rendering a spatial layout that appeared quasi-hexagonal (see Fig. 1). To eliminate pointing biases in any direction, these inanimate objects were intentionally designed by Friedman et al. (2020) to have no inherent directionality (i.e., a salient initial heading direction).

The object array was shown on the left half of the computer screen, subtending a visual angle of 16.2°, while a circle for making pointing responses was shown on the right half of the screen, subtending a visual angle of 11.2°. Following on-screen instructions located at the bottom of the object array, participants were told to imagine themselves standing at the location of an initial object, facing a second object, and then point to a third object based on that imagined perspective. On each trial, the imagined initial heading direction was shown on the adjacent response circle and each participant responded by positioning a straight line that rotated about the center of the response circle in a desired pointing direction. The straight line represented a pointing arrow and showed up when a participant made a left-button mouse click at any position within the response circle. To confirm one's decision on a pointing direction, a participant must press the "Enter" key on the keyboard, upon which data recording occurred. Participants' responses were recorded in the form of angular pointing errors to the accuracy of two decimal places. This error is unsigned and refers to the deviation in degrees from a point-to-target angle (i.e., target angle) that was pre-computed as correct. For instance, if the target angle was pre-computed as 40°, and a participant's response angle was 30°, he/she would have registered a pointing error of 10°.

Each participant performed three practice trials first before performing the test trials. To each participant, a total of 12 test trials, arranged in a randomized sequence that differed between participants, were presented. The target angles ranged from a minimum of 25° to a maximum of 333° ($M = 191.83^\circ$, $SD = 103.98$). The number of trials completed, mean and standard deviation of pointing errors were computed automatically by the program at the end of the task. For any trial that was not completed by a participant, the program automatically registered a pointing error of 90°. To ensure accuracy in our data analysis, such data from uncompleted trials were removed in the computation of mean and standard deviation errors for each participant.

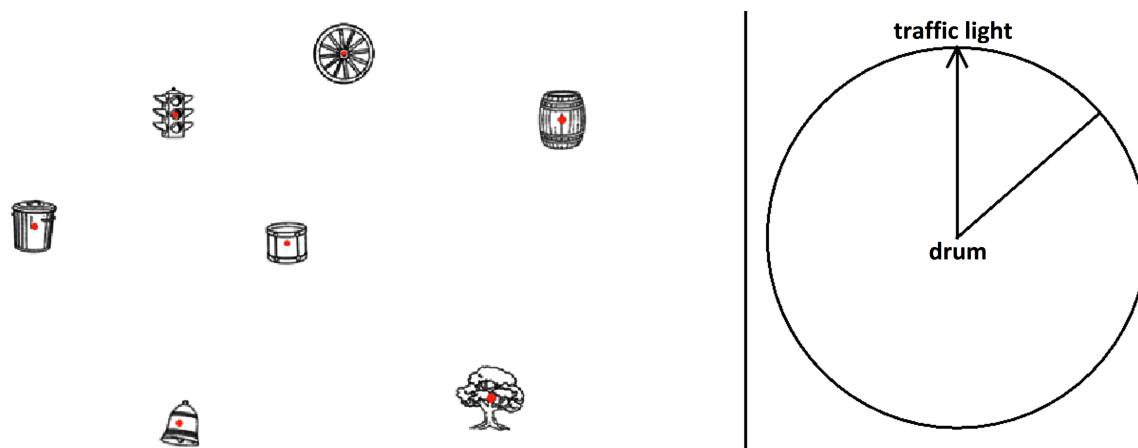


Figure 1. Screenshot of a practice trial in Friedman et al.'s (2020) SOT. For clarity of presentation, each 2D object are shown in higher contrast, verbal instructions and object labels are also enlarged. The target direction in the upper right quadrant of the response circle (derived from the computer mouse actions of the lead author) shows the estimated position of the wheel relative to an initial heading that is to be imagined. [Reproduced with permission under the Creative Common Attribution 4.0 license.]

Money Road Map Test (MRM-T)

The MRM-T, designed originally by Money, Alexander, and Walker (1965)⁵³, was the second test used for assessing participants' SOA. It is a paper-based test and has been used previously by Friedman et al. (2020). We used it to assess the concurrent validity of Friedman et al.'s (2020) SOT, that is, we wanted to examine how well pointing errors from the new SOT correlated with accurate performance on the MRM-T. In this study, we presented an open-source MRM-T used previously by Elman et al. (2013)^{55,56}. It shows a paper map of an imaginary town, with a route running through its streets in the form of a continuous dashed line. The map shows 32 turning points in total and participants were told to imagine travelling along the continuous

172 route and to indicate at each turning point whether the route turned left or right. At each turning point, each participant was
173 told to write a capital letter “L” besides it if he/she perceived the route as turning left and to write a capital letter “R” if he/she
174 perceived the route as turning right. A short route for pre-test practice was located at the bottom-right corner of the map.

175 Air Traffic Conflict Detection Task (AT-CDT)

176 Apparatus and Stimuli

177 The AT-CDT was designed using a real-time ATC simulation software that was produced by the National Aerospace Laboratory
178 [Dutch: Nationaal Lucht- en Ruimtevaartlaboratorium (NLR)] of the Netherlands: NLR Air Traffic Management Research
179 Simulator (NARSIM, version 8.0). A square-shaped 2K radar monitor with an aspect ratio of 1:1 (20 inches by 20 inches,
180 equivalent to 50.8 cm x 50.8 cm) was used to display air traffic scenarios in a controlled airspace (see Fig. 2, central image).
181 Within this airspace, simulated ATC services were provided to aircraft in accordance to instrument and visual flight rules,
182 international regulations that govern flight operations relying on instrument and visual references, respectively. The surveillance
183 range of the radar simulator covered a radius of 44 nm over the controlled airspace based on NARSIM’s default view settings,
184 allowing all aircraft to be shown within the dimensions of the radar monitor. Air traffic was shown in full screen at a resolution
185 of 2048 x 2048 pixels and a refresh rate of 54.6 Hz. Each participant sat 25 inches (66.04 cm) away from the center of the
186 screen when performing the task.

187 The entire radar display showing the full range of the controlled airspace and the air traffic embedded within it subtended a
188 visual angle of 36.3°. The radar display of each aircraft was symbolized by a radar blip (white dot) showing its current location
189 and a short trail (dotted line) representing the extent of displacement from the aircraft’s previous radar-detected location (Fig. 2,
190 right panel). With each aircraft’s true airspeed standardized at a constant of 450 knots (see subsection below, for details),
191 each radar blip and its trail spanned 10 mm in length and subtended a visual angle of 0.9°. In addition, the data block placed
192 next to each blip, showing the aircraft’s callsign, flight level, true airspeed, and landing runway, spanned 20 mm in height
193 and subtended a visual angle of 1.8° (Fig. 2, right panel). Information about each aircraft’s location was updated every four
194 seconds at a rate of 0.25 frames per second (equivalent to 15 frames per minute). This related to an antenna rotation speed of 15
195 rounds-per-minute (rpm) of a surveillance radar that transmits and receives in return (from a transponder on each aircraft) radio
196 frequency signals about the identity, location, and motion-related information of the aircraft under its detection range.

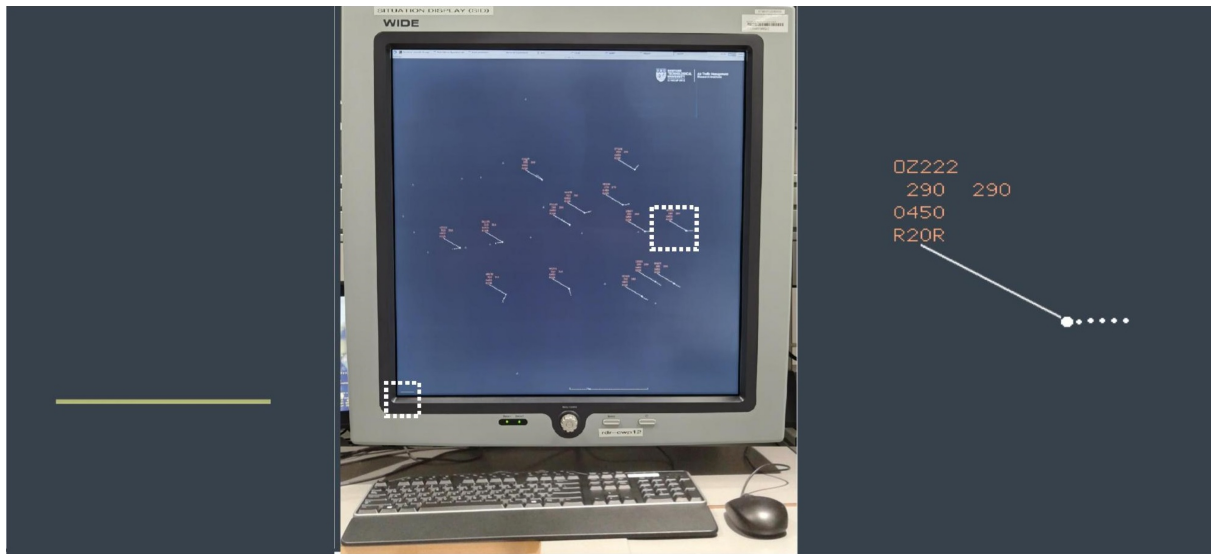


Figure 2. 2K radar monitor (2048 pixels x 2048 pixels) used for showing simulated air traffic data. The air traffic scenario shown was taken from the third AT-CDT test trial – the first high high air traffic density (ATD) condition with 14 aircraft. The left panel shows the 20 mm ruler (boxed on bottom left corner of the radar display) that indicates 5.0 nm separation between a pair of aircraft while the right panel shows details from an enlarged data block of an aircraft (boxed on the right side of the radar display) appearing in the eastern sector of the controlled airspace. The university logo and the full name of the research institute were displayed at the top right corner of the radar display.

197 In addition to these aircraft-based stimuli, a fixed number of small triangles, symbolizing the waypoints along air routes,
198 were displayed repeatedly across all simulation trials. Each triangular symbol subtended a visual angle of 0.5°. These waypoint
199 symbols were introduced for the purpose of easing the perception of changes in aircraft locations and headings. Fixed air
200 routes – conventional routes that were pre-designated within the controlled airspace for aircraft flying between a specific set of

waypoints – and non-fixed air routes – straight-line flight paths that pass through waypoints different from those of the fixed air routes – were used during the task design phase for conflict generation and traffic flow arrangement (see subsections below). Lines delineating these air routes were kept invisible during task performance to deter participants from using them as visual cues to simplify the task of perceiving aircraft heading changes and conflict occurrence. By hiding air route lines, our task did not provide participants with obvious visual cues about where an aircraft would turn, compelling them to make use of their spatial perception and orientation skills. This means that participants must pay attention to aircraft heading changes and judge whether such changes could lead to any conflicts.

Conflict Event Generation

To simulate an ATC experimental trial, we scripted an initial flight plan (IFP) in eXtensible Markup Language (XML), specifying the air routes, initial aircraft coordinates, heading directions, flight levels (i.e., altitude of aircraft operations), and relevant air speeds (i.e., calibrated airspeed, true airspeed). One demonstration trial, one practice trial, and nine test trials constituted the task in full and conflict events were specified in the IFP of each trial (see Procedures, for specific details). In general terms, an air traffic conflict refers to an event in which two or more aircraft experience a loss of minimum separation. Following conventional norms, this minimum separation refers to a minimum lateral separation of 5.0 nautical miles (nm) on the same flight level and a minimum vertical separation of 1,000 feet¹⁵. Specifically, in our task, we focused on lateral conflicts occurring on the same flight level and restricted the number of aircraft involved in each conflict to two. The former criterion ensured that SOA was assessed unambiguously for stimuli presented on the same spatial plane, akin to the stimuli presented in the SOT, while the latter criterion ensured the ease of generating conflict events and data recording.

We created two types of conflict events (i.e., conflict event models) using certain geometric or topological rules that were kept constant across the trials. We introduced two distinct conflict events instead of one in order to collect more behavioral data on each simulated test trial, as well as to have more dynamic situations for evaluating the effect of SOA. Importantly, we specified aircraft heading changes in both conflict event models based on the notion that mental processes inherent to SOA are engaged in the perception of such changes. Moreover, to keep participants focused on detecting aircraft heading changes and to minimize individual differences in the perception of dynamic stimuli, the true airspeed of all aircraft, irrespective of their conflict status, was fixed at 450 knots (833.4 km/h) throughout the entire duration of every trial. Consequently, all conflicts in our task occurred during the en route flight phase. Aircraft that had the potential of coming into conflict with each other were set to cruise at the same flight level while aircraft with no chances for conflict were set to cruise at flight levels separated from the former clusters of aircraft by a minimum of 1,000 feet.

In the first conflict event model (Model A), two aircraft had an initial angle of approach that was obtuse (approximately 130° on average) and were set to arrive at a head-on conflict after one of these aircraft made an acute angular turn (approximately 45° on average) at a waypoint (see Fig. 3A). In the second conflict event model (Model B), two aircraft were set to approach each other from opposite directions along two air routes whose initial segments ran parallel to each other (see Fig. 3B). The initial segment of both air routes was set to differ at a minimum orthogonal lateral separation of 7.0 nm. One of these aircraft followed an air route that involved an acute angular turn (approximately 45° on average) at a waypoint that would bring it into conflict with the other aircraft flying in the opposite direction.

In each model, the length of the air routes of the conflicting aircraft merging at the center of the conflict zone (i.e., an airspace with 5.0 nm radius) was kept the same. In both models, non-conflicting aircraft cruising at the same flight level as the conflicting aircraft were separated from the conflict zone's center at an initial distance that was at least 1.5 times that of the initial air route distance separating either conflicting aircraft from the conflict zone's center (Fig. 3). For all non-conflicting aircraft, they were initially placed at distances that were sufficiently far away from a conflict zone to offset any chances of them entering into the conflict zone at the same time as the conflicting aircraft.

Conflict Event Occurrence, Location, and Air Traffic Presentation

All participants were actively involved in conflict detection on the practice and test trials. On each trial, the two types of conflict events were shown, in sequence, at distinct locations in the controlled airspace. The conflict event shown in model A always occurred before that shown in model B due to the longer air routes travelled by the conflicting aircraft pair in the latter scenario. For both models, we measured the time it took for a conflict to occur in terms of *time at the closest point of approach* (*tCPA*), which refers to the time it took for a pair of conflicting aircraft to reach a lateral separation of 5.0 nm from their initial coordinates since the onset of simulated air traffic on a particular trial (Fig. 3). Across the practice and test trials, the *tCPAs* for the conflicting aircraft pairs in conflict events A and B were measured at means of 122 seconds ($SD = 6.2$ seconds) and 143 seconds ($SD = 5.0$ seconds), respectively, from trial onset. The initiation of heading changes of the aircraft involved in conflict events A and B began at means of 85 seconds ($SD = 5.1$ seconds) and 100 seconds ($SD = 3.1$ seconds), respectively, from trial onset. This means that on average, heading change of the conflicting aircraft (colored in red in Fig. 3) *first* occurred at means of 37 seconds before conflict event A and 43 seconds before conflict event B.

On each test trial, the coordinates at which airborne collisions were specified to occur, marking the center of a conflict

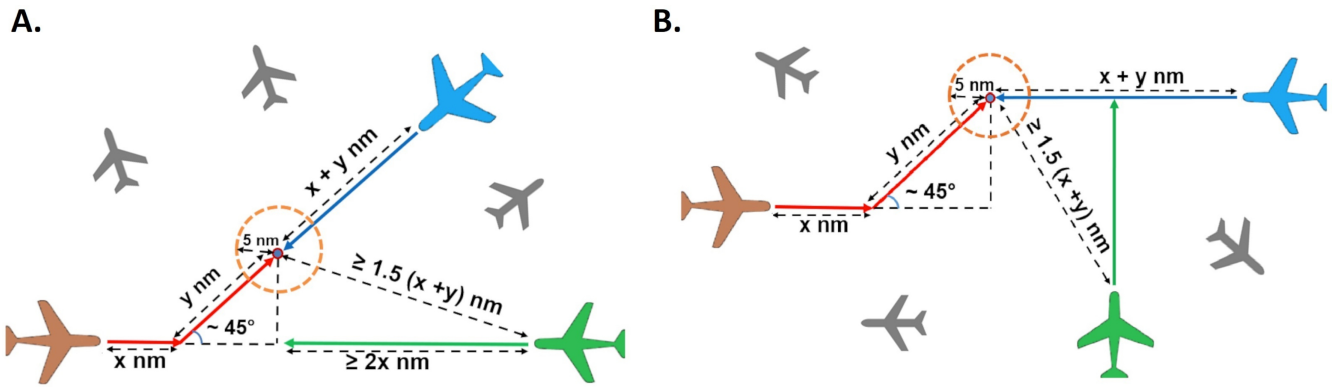


Figure 3. Conflict event models A (left panel) and B (right panel). Algebraic symbols and equations specify the length of the path segments of the air routes (not drawn to scale) and their topological relationships. These routes were kept hidden from participants' view. Colored aircraft represent aircraft cruising on the same flight level while grey aircraft in the background represent aircraft cruising at flight levels different from that of the former cluster. On the same flight level, a non-conflicting aircraft in each model (colored in green) was placed initially at a safe distance away from the conflict zone, which is represented by the dotted circle with a radius of 5.0 nm. The total number of non-conflicting aircraft located in the vicinity of two conflicting aircraft that constituted each conflict event increased from one in the low air traffic density (ATD) condition to three and five in the medium and high ATD conditions, respectively (see text, for more details).

zone, coincided with the coordinates of waypoints along an air route or the mid-section of an air route between two waypoints. These predefined waypoints and air routes were kept invisible during task performance to prevent participants from using them as visual cues for judging the locations of the conflicts. We programmed the movements of aircraft about these waypoints with reference to FRA scenarios with high levels of aircraft manoeuvring uncertainty and distribution of conflict points across time^{49,50}. Separation between the airborne collision sites were kept apart at a minimum separation of 10.0 nm. Two collision sites were introduced on each trial, one for each conflict event model. The location of these sites on the practice trial differed from that on the test trials. On each test trial, each pair of conflicts occurred, in succession, at two discrete collision sites (see Fig. 5). There were three pairwise combinations of collision sites in total and they were shown in different orders over three trial blocks (see Procedures, for more details).

Across the test trials, we manipulated mental workload by increasing the number of aircraft operating within the boundaries of the specified common airspace over three levels of air traffic density (ATD): (i) low (six aircraft), (ii) medium (10 aircraft), and (iii) high (14 aircraft). This manipulation was based on the premise that demands on information processing would increase with higher numbers of aircraft and had been attempted by previous studies^{45,57}. Critically, the numbers of aircraft were determined empirically with respect to real-world air traffic data collected using SB ADS-B technology over flight levels ranging from 21,000 to 32,000 feet in the Singaporean airspace. This range of flight levels corresponded with the range used in the current AT-CDT (see Table 1). Figure 4 shows the area covered by an ATC sector located to the south of Singapore island and the frequency distribution of aircraft over a 24 hour time span on a randomly chosen day in 2019 before COVID-19 outbreak. With reference to the statistical information shown in Figure 4, the aircraft number in the low ($n = 6$) and medium ($n = 10$) ATD conditions matched exactly with the 25th and 50th percentile values, respectively, from the entire distribution of real-world aircraft numbers. The aircraft number in the high ATD condition was one more than the real-world aircraft number shown at the 75th percentile level ($n = 13$, see Figure 4). This was done because we presented a fixed even number of conflicting aircraft in each ATD condition ($n = 4$) and assigned even numbers of non-conflicting aircraft that increased systematically from the low to high ATD conditions (see paragraphs below).

In addition, we applied the current set of aircraft numbers based upon professional feedback given by an ATCO who has expert knowledge about air traffic flow and volume under operational ATC settings. Furthermore, with respect to pre-existing ATM-related human factors studies, the aircraft numbers we presented were approximately equal to the numbers used in a neurophysiological study on ATCO mental workload⁵⁷ and were slightly lower than the numbers used in another human factors study that manipulated ATD⁴⁵. Considering that our participants are not professional ATCOs and that each test trial was capped at a maximum time-limit of three minutes (see Procedures), we did not present larger numbers of aircraft in each ATD condition (e.g., ≥ 10 under low ATD, ≥ 20 under medium and high ATD) to prevent our task from getting overly complicated.

We set up three blocks of test trials with each block composing three trials. Within each trial block, low ATD was always shown first, followed by medium and high ATD. On each trial, the initial lateral separation between any pair of aircraft was set at a minimum of 7.0 nm. The number of non-conflicting aircraft cruising at the same flight level as the two pairs of conflicting

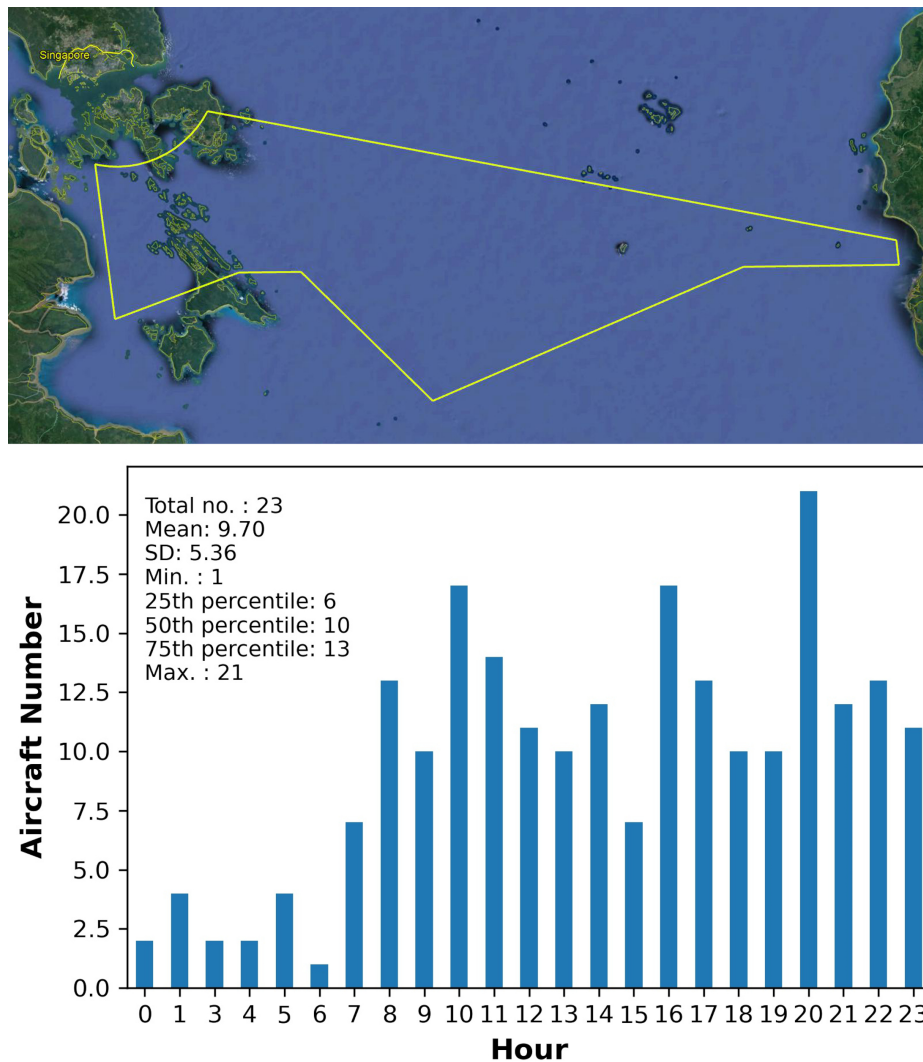


Figure 4. Figure showing the boundaries of an air traffic control (ATC) sector (colored in yellow) covering the airspace in the southeastern region of Singapore island (top panel, generated using Google Earth) and the aircraft number recorded on a randomly selected 24-hour working day in 2019 before COVID-19 outbreak (bottom panel, descriptive statistics shown on the top left corner; SD = Standard Deviation. Min. = Minimum. Max. = Maximum).

aircraft on each trial was increased from one per conflict in the low ATD condition, to two and three per conflict in the medium and high ATD conditions. These procedures ensured a gradual increment in the difficulty of conflict detection.

In addition, we presented non-conflicting aircraft cruising at flight levels that were *different* from the two flight levels on which the two types of conflict events occurred on the medium and high ATD trials. The numbers of these special group of non-conflicting aircraft were set at two and four, respectively, on the medium and high ATD trials. Regardless of ATD condition, all non-conflicting aircraft were arranged to travel along straight-line air routes. Importantly, we assigned a unique callsign, standardized at five characters (two letters and three digits), to each aircraft within and across trials. This was done to eliminate any responses that relied on verbal memory or attentional biases to aircraft showing the same callsigns within and across trials.

Procedure

All participants were tested individually over single experimental sessions that lasted approximately one hour per participant. Each participant completed the SOT and MRM-T before performing the AT-CDT. The experimenter administered the first two tests in a counter-balanced order. Half of the participants performed the SOT first while the other half performed the MRM-T first. Before formal testing, participants were given some practice trials. In the MRM-T, participants practiced indicating three turning directions correctly on the short route located on the bottom-right corner of the paper maze. Importantly,

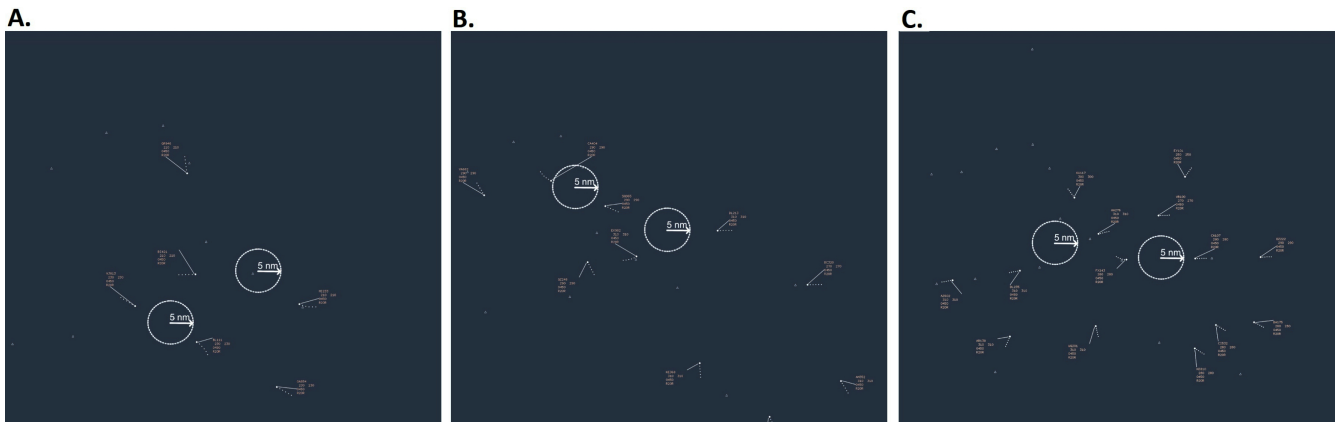


Figure 5. First block of three simulation trials showing three levels of air traffic density (ATD): (A) low (six aircraft), (B) medium (10 aircraft), and (C) high (14 aircraft). White circles, each with a radius of 5.0 nm, delineate the boundaries of the conflict zones. Such circles were not shown during simulation. Through a Latin square design, the pairwise combinations of conflict sites recurred in two different, non-overlapping sequences across the three ATD conditions over the next two blocks of trials (three trials each). The initial arrangement of aircraft in each subsequent ATD condition differed from the arrangement of aircraft shown within the first trial block and varied from each other.

the experimenter fixed the paper map in a portrait orientation in front of each participant during this practice phase and the subsequent one-minute test phase. A participant was only allowed to proceed to the test phase upon correct indication of all three turns on the practice route. In the presence of any errors, the particular participant was told to repeat the practice until no errors were made. Likewise, on the SOT, each participant performed three practice trials prior to the test phase. Before performing these trials, to ensure that all participants knew the identities of each object comprising the object array, the experimenter pointed to each object and mentioned their names to each participant. On each practice trial, automatic feedback showing the correct target direction was shown immediately after the registration of a response. In the event of large angular error ($\geq 90^\circ$) on any practice trial, the particular participant was told to repeat that trial until there was a noticeable reduction in error. After passing the practice phase of each test, each participant was given maximum time-limits of one minute and five minutes, respectively, to complete the MRM-T and SOT in the formal testing phases. Before the start of each testing session, the experimenter told each participant to respond as fast and as accurately as he/she could and recorded each participant's test completion time with a stopwatch.

After completing the two spatial tests, all participants were introduced to the AT-CDT. They were given detailed descriptions of the main characteristics of a conflict event (mentioned above in "Conflict Event Generation" subsection) and told explicitly that the conflicts under examination in the current task referred only to lateral conflicts that involved a loss of minimum separation between a pair of aircraft at a distance of 5.0 nm or less. Before starting a practice trial, the experimenter presented a demonstration trial explaining the visual stimuli presented by NARSIM. This trial featured conflict event model B and involved three aircraft. With the simulation in progress on this trial, the experimenter introduced the radar blips representing each moving aircraft, the waypoints, and the data blocks containing information about each aircraft's callsign (first row), flight level (second row), true airspeed (third row), and landing runway (last row). First, they were told that all aircraft cruised at the same true airspeed of 450 knots and would land at the same runway, and that they did not need to pay attention to the information indicated on the third and last rows of the data block (Fig. 2, right panel). Second, they were told to attend to the flight level data and informed that any pair of aircraft that were separated by 10 units (representing 1,000 feet) would not be in conflict. Third, they were told to pay close attention to aircraft heading changes and to all aircraft cruising at the same flight level.

In addition, all participants were shown a scale located at the bottom-right of the screen that indicated the lateral conflict distance threshold of 5.0 nm on the radar screen (Fig. 2, left panel). This scale spanned 20 mm in length and subtended a visual angle of 1.8° . Together with this scale, the experimenter also brought up a measuring line by clicking on the scroll wheel of the mouse that enabled an automatic measurement of the lateral distance separating any pair of aircraft. Specifically, when the conflicting aircraft pair reached a distance approximating the length of the scale, the experimenter activated the measuring line, showing the magnitude of lateral distance separation. This procedure was done to give each participant an initial percept of what a loss of minimum lateral separation entailed and was not repeated in the practice and test trials. More importantly, all participants were told to indicate their detection of the conflicting aircraft pair by double-clicking their callsigns in quick succession. To demonstrate this, the experimenter double-clicked on the correct conflicting pair. With each series of double-clicks, time elapsed from trial onset was recorded to millisecond accuracy. This form of temporal data recording

Table 1. Flight Parameters used in AT-CDT Practice and Test Simulation Trials

Trial Type	ATD	Total Aircraft No.	Aircraft on same FL	FL (CE Model A)	FL (CE Model B)	Flight Levels
Practice	Low	6	6 [3 (Model A), 3 (Model B)]	30,000	25,000	25,000, 30,000
Test 1	Low	6	6 [3 (Model A), 3 (Model B)]	21,000	23,000	21,000, 23,000
Test 2	Medium	10	8 [4 (Model A), 4 (Model B)]	24,000	29,000	24,000 – 29,000
Test 3	High	14	10 [5 (Model A), 5 (Model B)]	30,000	32,000	23,000 – 32,000
Test 4	Low	6	6 [3 (Model A), 3 (Model B)]	25,000	30,000	25,000, 30,000
Test 5	Medium	10	8 [4 (Model A), 4 (Model B)]	29,000	31,000	27,000 – 32,000
Test 6	High	14	10 [5 (Model A), 5 (Model B)]	27,000	32,000	27,000 – 32,000
Test 7	Low	6	6 [3 (Model A), 3 (Model B)]	28,000	26,000	26,000, 28,000
Test 8	Medium	10	8 [4 (Model A), 4 (Model B)]	30,000	27,000	27,000 – 32,000
Test 9	High	14	10 [5 (Model A), 5 (Model B)]	28,000	31,000	25,000 – 31,000

Note. ATD = Air Traffic Density. CE = Conflict Event. FL = Flight Level (in feet).

applied to all subsequent trials.

After the demonstration, the experimenter started a practice trial. This trial featured six aircraft and two conflict events based on models A and B. Each conflict event occurred in a discrete sector of the controlled airspace and involved three aircraft cruising together on a flight level (25,000 feet) that was distinct from the flight level of the other trio (30,000 feet). Each participant was given control of the mouse and told to indicate two pairs of conflicting aircraft in the same way demonstrated by the experimenter. A participant was only allowed to commence on the test trials after the correct detection of both pairs of conflicting aircraft. In the event of any incorrect detection, participants were re-briefed about the task and re-attempted the practice trial from the start. After completing the practice trial, all participants proceeded to the test trials. There was a total of nine test trials arranged in three trial blocks. Table 1 shows the flight parameters (i.e., ATD, aircraft numbers, flight levels) associated with these practice and test trials.

In each trial block, to ensure that there was a gradual exposure to trials of increasing difficulty, the trial with low ATD was always shown first, followed by medium and high ATD trials (column 2, Table 1). As shown in Fig. 5, three pairwise combinations of discrete collision sites were presented in each block of three trials. We configured them using a Latin square design, meaning that they appeared in a distinct and non-overlapping sequence within each trial block and ATD condition. This ensured that no two collision sites were in the same air sectors between trials and reduced the chances for any participant to predict the whereabouts of a conflict on any trial based on knowledge acquired from the trial that immediately preceded it. Across all trials, inclusive of the demonstration and practice trials, the flight levels of all aircraft ranged between 21,000 and 32,000 feet ($M = 28,111.11$, $SD = 2895.77$) [last column, Table 1]. On the low, medium, and high ATD test trials, each conflict event occurred within clusters of three, four, and five aircraft cruising at the same flight level, respectively (column 4, Table 1). Between trials, the flight levels assigned to each conflict event involving these clusters of aircraft differed by a margin ranging from 2,000 to 5,000 feet. Additional non-conflicting aircraft in the medium and high ATD trials cruised at flight levels that are safely separated from the flight levels of the quartets and quintets of aircraft designed for conflict detection by a minimum margin of 2,000 feet. Crucially, there was no repetition of the same flight level information across the test trials. This means that that no aircraft cruised at the same flight level on two consecutive trials. This deterred participants from making responses based on attentional biases or memory of aircraft with repeated flight level information.

On each test trial, like on the practice trial, participants responded by double-clicking on the callsigns of the aircraft that were perceived as coming into conflict. They were told that their responses were non-amendable and that they must think carefully before making each response. A time-limit of three minutes was given for the detection of both conflict events and each participant was told to respond as fast and as accurately as he/she could.

Results

Test Measures for Data Analysis

From each of the three tests, we selected relevant accuracy and response time (RT) measures for use in the correlation- and regression-based data analyses. We detailed their recording and computation in the subsections below.

MRM-T

From the MRM-T, we used the total accuracy score recorded from each participant as the sole performance indicator. As all participants used up one minute to perform the MRM-T, we did not involve RT on this test in our data analysis. Accuracy was scored in a binary fashion, with "1" given to a correctly labeled turn and "0" given to an incorrectly labeled turn. The total accuracy score achievable by any participant was 32.

SOT

From the SOT, we used the mean (M) and standard deviation (SD) of pointing errors (in degrees), and mean RT (in seconds, recorded to the accuracy of two decimal places) exhibited by each participant as the key performance indicators. For each participant, mean RT was computed by dividing the total time taken to complete the SOT (maximum five minutes) by the total number of test trials completed (out of 12). Due to the fact that the SOT did not feature an automatic recording of the time taken to complete each trial, we could not compute the exact total and mean RT taken to complete the test trials correctly based on a preset accuracy criterion (e.g., a maximum error limit of 90° , under and above which a trial could be scored as correct and incorrect, respectively). Nonetheless, we circumvented this problem by computing a corrected SOT mean RT for each participant that factored his/her mean pointing error into consideration. The formula is given as follows:

$$t_{\mu_i^n}^c = (1 - \frac{\epsilon_{\mu_i^n}}{180^\circ}) * t_{\mu_i^n} \quad (1)$$

in which,

$t_{\mu_i^n}^c$ = Corrected SOT mean RT computed for the i^{th} participant who completed n no. of test trials

$\epsilon_{\mu_i^n}$ = Mean pointing error shown by the i^{th} participant who completed n no. of test trials

$t_{\mu_i^n}$ = Raw (uncorrected) SOT mean RT shown by the i^{th} participant who completed n no. of test trials

Here, the raw SOT mean RT is multiplied by a coefficient ($1 - \frac{\epsilon_{\mu_i^n}}{180^\circ}$) that we conceived as representing an *estimate* of the mean proportion of time dedicated to performing the SOT in an accurate or optimal fashion. Note that the mean pointing error is divided by 180° , which is the maximum pointing error that can be committed by any participant. A simple way to understand this formula lies in the example in which a participant committed 0° error on average. In this case, the coefficient would equate to one, and his/her corrected mean RT would equate to his/her uncorrected mean RT. Conceptually, it can be assumed that this participant devoted all his thinking time to making perfectly accurate pointing responses. In practice, however, such a feat showing perfect performance is highly unlikely, and equation (1) gives an truncated estimate of the time taken on average by a participant to decide on and conduct an *accurate* pointing response. In other words, we partialled out an estimate of the mean proportion of time spent on committing pointing errors in the computation of this corrected RT.

AT-CDT

From the AT-CDT, we computed a *time to loss of minimum separation* ($tLMS$) measure based on equation (2) below:

$$tLMS_{ij} = t_{2C_{ij}} - tCPA_{C_j} \quad (2)$$

in which,

$tLMS_{ij}$ = tLMS shown by the i^{th} participant on the j^{th} trial

$t_{2C_{ij}}$ = Response latency shown by the i^{th} participant on the j^{th} trial in the detection of a second aircraft in a perceived conflict event (C)

$tCPA_{C_j}$ = Pre-recorded tCPA attached to a particular conflict event (C) on the j^{th} trial (constant across all participants, see "Conflict Event Occurrence" subsection above)

Here, it must be clarified that $t_{2C_{ij}}$ marks the time that elapsed from trial onset to the moment when a participant clicked on the *second* aircraft (denoted by the subscript "2") – immediately after clicking on a first aircraft – in a pair that he/she perceived as coming into conflict. By applying the difference score measure shown in equation (2), we obtained a standardized RT measure that corrected for the different times at which two conflict events (Models A and B, Fig. 3) occurred on each test trial. A negative tLMS value indicates that a pair of conflicting aircraft was identified *before* conflict onset whereas a positive tLMS value indicates that the conflicting pair was identified *after* conflict onset. The tLMS value of zero indicates conflict

onset, that is, the moment when two aircraft reach a minimum lateral separation of 5.0 nm. By having a signed standardized RT measure, we were able to perform time averaging across trials of interest and apply the mean tLMS measure for correlational analysis involving other variables.

In computing the mean tLMS (τ_{μ_i}) for each participant [(3)], we only factored in correctly detected conflicts over a particular number of completed trials (n). Incorrect detections were ignored because the baseline tCPA pertained to a pair of aircraft that were in actual conflict and hence tLMS can only be interpreted meaningfully with respect to correct detections. In the formula below [(3)], $x_{\delta_{ij}}$ represents the accuracy score associated with the detection of one conflict event. The maximum accuracy score achievable on the AT-CDT is 18, which represents the total number of conflict events across nine test trials.

$$\tau_{\mu_i} = \frac{\sum_{j=1}^n (tLMS_{ij} * x_{\delta_{ij}=1})}{\sum_{j=1}^n x_{\delta_{ij}=1}} \quad (3)$$

$$x_{\delta_{ij}} = \begin{cases} 1, & \text{if correct detection} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Here, it must be emphasized the computation of mean tLMS is conceptually similar to the computation of the corrected SOT mean RT shown in equation (1). In both cases, the mean test accuracy displayed by each participant is factored into RT computation. On each trial, mean tLMS is computed by averaging the tLMS tied to the *correct* detections of both conflict events A and B. In the presence of only one correct detection, mean tLMS on that trial referred to that detection only.

We found that the pair of mean tLMS values from both conflict event types correlated highly with each other ($r_{20} = .85$, $p < .001$). Critically, a paired sample t-test showed non-significant difference between the mean tLMS values from both conflict event types, $t_{19} = 1.05$, $p = .348$, $M(SE)_{Difference} = 1.23(1.28)$, suggesting that there were no features inherent to either conflict event type that bias participants toward faster (or slower) responses. Consequently, we did not analyze the tLMS values obtained from either conflict event type separately from the other in the analyses below. By doing so, we did not endorse an opinion that qualitatively different mental processes were involved in the detection of these two conflict event types.

Descriptive Statistics

Table 2 shows the descriptive statistics of all relevant test measures obtained from the three tests performed by the total sample of 20 participants. Total accuracy scores from the MRM-T and AT-CDT were converted into percentages for ease of comparison. For each participant, SOT mean RTs and and AT-CDT mean tLMS were computed to the accuracy of two decimal places. Consequently, the means of these temporal measures shown in Table 2 reflect the averages of means taken from each participant.

An examination of the data distribution of Table 2 test measures showed that that most participants achieved perfect or close to perfect accuracy scores on the AT-CDT (Mean % accuracy = 95). Specifically, a scrutiny of AT-CDT accuracy data using histograms, box-plots, and Q-Q plots showed a negatively skewed distribution ($skewness = -1.28$, $SE = 0.51$), with most participants scoring full marks ($n = 11$) or close to full marks ($n = 8$, accuracy score ≥ 15). Only one participant achieved a "low" score of "14." A test of the normality assumption using the Shapiro Wilk's test further showed that the distribution of these accuracy scores deviated prominently from normality (Shapiro-Wilk's $W = .750$, $p < .001$). Due to this observed ceiling effect and the fact that accuracy had already been factored into the computation of mean tLMS, we removed AT-CDT accuracy in the correlational analysis presented below.

With respect to AT-CDT mean tLMS computed based on correct detections, we applied a general measure of mean tLMS, averaged across both models, in the analyses below. To examine whether participants had a general tendency to respond before or after the start of a conflicting aircraft's heading change in each model, we computed two sets of mean tLMS values derived from conflict events A and B, respectively (see last two rows in Table 2). One-sample t-tests showed that the mean tLMS values derived from Models A and B, respectively, were higher than the mean reference tLMS values of -37 seconds that denoted heading change onset in Model A [$t_{19} = 4.31$, $p < .001$, $M(SE)_{Difference} = 6.77(1.57)$] and -43 seconds that denoted heading change onset in Model B [$t_{19} = 4.38$, $p < .001$, $M(SE)_{Difference} = 9.78(2.23)$] (see "Conflict Event Occurrence" subsection above, for details of computing these reference tLMS values). This showed that participants generally responded *after* the start of the conflicting aircraft's heading change in both models. Importantly, this finding suggested that observing an aircraft's turning motion helped participants to detect the correct pair of conflicting aircraft.

Correlational Analysis

Table 3 shows results from multiple bivariate correlations between MRM-T percent accuracy, SOT pointing errors (mean and SD), SOT mean RTs (raw and corrected), and AT-CDT mean tLMS. In particular, we computed four sets of mean tLMS

Table 2. Descriptive Statistics of Test Measures collected from Participants ($N = 20$)

Test	<i>M</i>	<i>SD</i>	Min.	Max.	Range
MRM-T					
% Accuracy	78.91	20.58	34.38	100.00	65.63
SOT					
Pointing Error Mean (°)	17.42	12.58	4.00	59.83	55.83
Pointing Error <i>SD</i> (°)	16.50	15.18	2.00	55.00	53.00
Mean RT (raw) [s]	22.36	8.01	10.00	42.86	32.86
Mean RT (corrected) [s]	20.10	7.27	9.78	40.55	30.77
AT-CDT					
% Accuracy	95.00	6.96	77.78	100.00	22.22
Mean tLMS (s) - Both Models	-28.69	8.15	-42.17	-14.4	27.77
Mean tLMS (s) - Model A ^a	-29.23	6.65	-41.94	-17.71	24.23
Mean tLMS (s) - Model B	-28.00	10.12	-44.54	-8.35	36.19

Note. *M* = Mean. *SD* = Standard Deviation. Min. = Minimum. Max. = Maximum. RT = Response Time.

tLMS = time to loss of minimum separation. ^aNon-significant difference from Model B mean tLMS ($p > .05$).

Table 3. Bivariate Correlations of Key Test Variables of Interest ($N = 20$)

Measure	1	2	3	4	5	6	7	8	9
1. MRM-T % Accuracy	-								
2. SOT Pointing Error Mean (°)	-.50*	-							
3. SOT Pointing Error <i>SD</i> (°)	-.51*	.89**	-						
4. SOT Mean RT (raw) [s]	-.28	.18	.19	-					
5. SOT Mean RT (corrected) [s]	-.17	-.05	-.03	.97**	-				
6. AT-CDT Mean tLMS (s) - All Trials	.04	.31	.31	.35	.21	-			
7. AT-CDT Mean tLMS (s) - Low ATD	.14	.31	.33	.14	.05	.95**	-		
8. AT-CDT Mean tLMS (s) - Medium ATD	.01	.29	.28	.27	.19	.91**	.88**	-	
9. AT-CDT Mean tLMS (s) - High ATD	-.08	.21	.19	.61**	.57**	.77**	.57**	.53**	-

Note. * $p < .05$ (two-tailed). ** $p < .01$ (two-tailed). ATD = Air Traffic Density. *SD* = Standard Deviation. RT = Response Time. tLMS = time to loss of minimum separation.

values: an overall mean that encompassed all nine AT-CDT test trials and three categorical means associated with the three ATD conditions.

We found three sets of noteworthy findings. First, there were significant and moderately high negative correlations between MRM-T percent accuracy and the SOT pointing errors [$r_{20} = -.50$, $p = .023$ (*M* error); $r_{20} = -.51$, $p = .021$ (*SD* error)]. These findings were important for showing that the SOT possessed concurrent validity with respect to the MRM-T. Second, there were significant and moderately high positive correlations between AT-CDT mean tLMS obtained from the high ATD condition and the two types of SOT mean RTs [$r_{20} = .61$, $p = .005$ (raw); $r_{20} = .57$, $p = .008$ (corrected)]. These findings were crucial to the central aim of this study and we further examined them these using linear regression analysis in the section below. Third, there were significant and positive correlations between the four sets of mean tLMS measures that ranged from moderate to high ($.53 \leq r_{20} \leq .95$, $p_s \leq .015$). Mean tLMS from the high ATD condition did not correlate highly with those from low ($r_{20} = .57$, $p = .008$) and medium ($r_{20} = .53$, $p = .015$) ATD conditions, suggesting that performance on the high ATD trials were not highly consistent with performance on the low and medium ATD trials (see ANOVA findings below, for details on the distribution of data points).

Moderated Linear Regression Analysis

Continuing from the correlational analysis above, we regressed each of the four sets of AT-CDT mean tLMS values on SOT mean RT (raw and corrected, respectively) to address the main aim of this study, that is, the extent to which SOT performance

could predict AT-CDT performance. By doing so, ATD was applied as a categorical moderator of the relationship between SOA and conflict detection. Table 4 shows the results of these eight linear regression models. Figs. 6 and 7 show the scatter-plots tied to these regression models. They show four scatter-plots apiece with SOT mean raw RT and corrected RT, respectively, as the predictor.

With alpha set at .017 (Bonferroni-corrected), both sets of SOT mean RTs emerged as significant predictors of AT-CDT mean tLMS in the high ATD condition [SOT raw RT: $\beta = 0.61$, $t_{18} = 3.24$, $p = .005$ (see Fig. 6A); SOT corrected RT: $\beta = 0.57$, $t_{18} = 2.96$, $p = .008$ (see Fig. 7A)], consistent with our correlational findings above. Notably, around a third of the variance in AT-CDT mean tLMS was explained by variation in SOT mean RTs [$R^2 = .37$ (raw); $R^2 = .33$ (corrected)]. The remaining relationships between SOT mean RTs and AT-CDT mean tLMS were found to be low and non-significant [SOT raw RT: $0.14 \leq \beta_s \leq 0.35$, $.130 \leq p_s \leq .551$ (see Figs. 6B, C, D); SOT corrected RT: $0.05 \leq \beta_s \leq 0.27$, $.248 \leq p_s \leq .824$ (see Figs. 7B, C, D)]. With alpha set at .017 (Bonferroni-corrected), comparisons of the beta values from the three ATD conditions (identical to their r -values) using Fisher's Z test of dependent correlations⁵⁸ showed that the correlation between SOT mean RTs and AT-CDT mean tLMS was significantly stronger in the high ATD condition than in the low ATD condition [$Z = 2.38$, $p = .009$ (SOT raw RT); $Z = 2.56$, $p = .005$ (SOT corrected RT)].

To ascertain that bivariate normality was present and that our regression findings were not skewed by the presence of outliers, we further performed one-sample tests of normality (Shapiro-Wilk and Kolmogorov-Smirnov) on the standardized residuals emanating from the eight linear regression models. Joint assessments of normality test statistics and Q-Q plots showed that none of these distributions of standardized residuals deviated significantly from normality with alpha set at .05 ($.932 \leq$ Shapiro-Wilk's $W \leq .974$, $.165 \leq p_s \leq .827$; $.095 \leq$ Kolmogorov-Smirnov's $D \leq .153$, $.680 \leq p_s \leq .986$).

Statistical power analysis

In addition, to confirm that the sample size of 20 participants was sufficient for confirming the significant effects derived from the regression analysis at an acceptable statistical power of 80%, we performed a post-hoc power analysis using G*Power (version 3.1.9.4)⁵⁹. With power and two-tailed alpha set at .80 and .05 respectively, we found that our sample size was adequate for confirming significant findings with an linear regression effect size (f^2) of .45, which corresponds to a lower-bound R^2 value of 0.21 (or a r -value of .46). As all significant R^2 and r values exceeded these lower-bound values, the significant correlations derived from our sample size should not be seen as under-powered.

Sex differences

Furthermore, to resolve a concern that individual variations in SOT mean RTs and AT-CDT mean tLMS could be explained simply as reflecting sex differences in general spatial ability or cognitive styles⁶⁰, we correlated the binary sex variable (with females coded as "0" and males coded as "1") with the two types of SOT mean RTs and AT-CDT mean tLMS (across all trials and within each ATD condition) using point-biserial correlations. With alpha set at the default value of .05, no significant correlations were found ($p_s \geq .218$). The correlations between sex and the two types of SOT mean RTs were virtually zero [$r_{pb} = -.001$, $p = .997$ (SOT raw RT); $r_{pb} = .04$, $p = .882$ (SOT corrected RT)] while the correlations between sex and the respective sets of AT-CDT mean tLMS were low [$r_{pb} = .18$, $p = .438$ (all trials); $r_{pb} = .13$, $p = .576$ (low ATD); $r_{pb} = .07$, $p = .775$ (medium ATD); $r_{pb} = .29$, $p = .218$ (high ATD)]. Altogether, these findings showed that the sex variable played a negligible role in mediating the observed correlational relationships between SOT mean RTs and AT-CDT mean tLMS.

Repeated Measures ANOVA of AT-CDT Performance

Owing to the fact that the AT-CDT is not a psychometric test but a controlled experimental task that we designed to assess SOA, we performed a 3 X 3 Repeated Measures Analysis of Variance (RM ANOVA) on the mean tLMS values to examine whether or not our experimental manipulations exerted significant effects. With mean tLMS specified as the dependent variable (DV), we examined the main effects of ATD (low, medium, high) and trial block (TB; triplets of trials presented over three cycles), as well as the ATD $\times$ TB interaction effect. We applied ATD as the first independent variable (IV) to check for the presence of mental workload effect, that is, we wanted to know if an increment in ATD elicited more positive mean tLMS values that signified longer delays before responses. We applied TB as the second IV to check for the presence of spatial learning effect, that is, we wanted to know if participants acquired some conflict geometry-related knowledge as the trials progressed, becoming faster in response on trials presented later in the sequence than on trials presented earlier.

Through RM ANOVA, we found significant main effects of both ATD ($F_{2,38} = 21.32$, $p < .001$, partial $\eta^2 = .529$, observed power = 1.00) and TB ($F_{2,38} = 9.91$, $p < .001$, partial $\eta^2 = .343$, observed power = .976), plus a significant ATD $\times$ TB interaction effect ($F_{2,76} = 5.07$, $p = .001$, partial $\eta^2 = .211$, observed power = .955). Fig. 8 shows this significant interaction effect by plotting out participants' mean tLMS distribution and averages on each test trial while Fig. 9 shows the two significant main effects by plotting out participants' mean tLMS distribution and averages in each ATD condition and TB.

To examine the simple effects emanating from these omnibus test effects, two-tailed dependent t-tests were performed with alpha set at Bonferroni-corrected thresholds of .0014 for between-trial comparisons (.05 / 36 possible comparisons) and

Table 4. Regression of AT-CDT Mean tLMS values on SOT Mean RTs (Raw and Uncorrected)

Predictor	Criterion Variable	R^2	Adj. R^2	$F_{1,18}$	I/C	B	SE	β	T_{18}	p
SOT Mean RT (raw) [s]	AT-CDT Mean tLMS (s)	.12	.07	2.51	-36.65	0.36	0.22	0.35	1.59	.130
	- All Trials									
	AT-CDT Mean tLMS (s)	.02	-.03	0.37	-39.58	0.21	0.34	0.14	0.61	.551
	- Low ATD									
	AT-CDT Mean tLMS (s)	.07	.02	1.36	-33.88	0.26	0.22	0.26	1.17	.259
	- Medium ATD									
SOT Mean RT (corrected) [s]	AT-CDT Mean tLMS (s)	.37**	.33**	10.47**	-35.79	0.58**	0.18	0.61**	3.24**	.005
	- High ATD									
	AT-CDT Mean tLMS (s)	.07	.02	1.43	-34.80	0.30	0.25	0.27	1.19	.250
	- All Trials									
	AT-CDT Mean tLMS (s)	.003	-.05	0.05	-36.68	0.09	0.38	0.05	0.23	.824
	- Low ATD									
SOT Mean RT (corrected) [s]	AT-CDT Mean tLMS (s)	.03	-.02	0.65	-32.15	0.20	0.25	0.19	0.81	.430
	- Medium ATD									
	AT-CDT Mean tLMS (s)	.33**	.29**	8.78**	-34.95	0.60**	0.20	0.57**	2.96**	.008
	- High ATD									

Note. ** $p < .01$ (two-tailed). β represents standardized regression coefficient. Adj. = Adjusted. ATD = Air Traffic Density. I/C = Intercept. RT = Response Time. tLMS = time to loss of minimum separation.

.017 for categorical comparisons (between the three ATD conditions and between the three TB). Table 5 shows all pairwise differences that were found to be significant based on these post-hoc comparisons. The findings showed that participants responded the slowest on the second ($M = -19.81$, $SE = 2.65$) and sixth ($M = -19.41$, $SE = 2.63$) test trials, on which they experienced the medium ATD condition for the first and the hard ATD condition for the second time, respectively. Specifically, responses on these two trials were significantly slower than responses on the fifth and seventh trials ($p_s < .001$, see Table 5, rows 3, 4, 9, 10 after header), which showed the medium ATD condition for the second time and the low ATD condition for the third time, respectively. On the other hand, participants responded the fastest on the seventh trial ($M = -39.41$, $SE = 2.31$). Responses on this trial was not only significantly faster than those from the second and sixth trials (as aforementioned), but also significantly faster than those from the first, third, eighth, and ninth trials ($p_s < .001$, see Table 5, rows 1, 7, 5, 11 after header).

In addition, categorical comparisons between the three ATD conditions showed significant differences for all three possible comparisons. On average, participants responded much slower on the high ATD trials than on the medium ($p = .005$) and low ATD trials ($p < .001$). They also responded slower on the medium ATD trials than on the low ATD trials ($p < .001$). On the other hand, participants responded the slowest within the first TB, and comparatively faster within the second and third TB ($p_s = .001$).

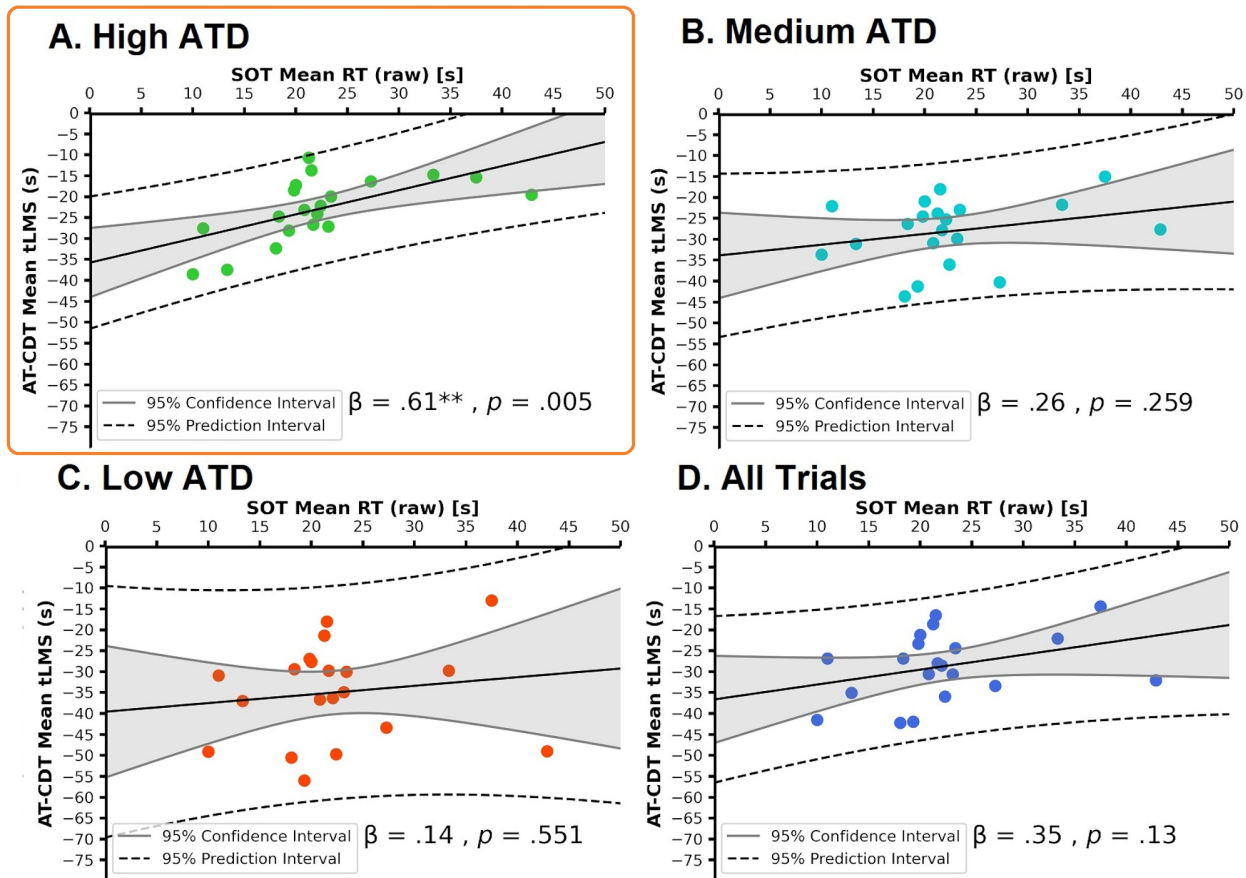


Figure 6. Scatter-plots showing the linear relationships between SOT mean raw RT and AT-CDT mean tLMS under each of (A–C) three air traffic density (ATD) conditions and over (D) all AT-CDT test trials. Panel A (outlined in orange) shows the significant correlation emanating from the high ATD condition. 95% confidence and prediction intervals are shown. In all plots, no data points lie beyond the 95% prediction interval, showing the absence of outliers.

Discussion

The main goal of this study was to investigate how well a newly developed SOT¹⁹ could predict performance on an AT-CDT that presented simulations of different air traffic scenarios within a sample of young adults who can apply to become ATCOs. Through the use of temporal measures recorded from both tasks, we found that this predictive relationship applied only to a high ATD condition that featured 14 aircraft. MRM-T accuracy scores did not correlate with AT-CDT performance but correlated moderately with SOT pointing errors (both mean and *SD* errors). These significant correlations replicated previous findings by Friedman et al. (2020)¹⁹ and demonstrated once more that the SOT possessed concurrent validity with respect to the MRM-T. In addition, RM ANOVA of AT-CDT mean tLMS values showed that participants responded slower, on average, in selecting the correct pair of conflicting aircraft in the high ATD than in the low and medium ATD conditions. These findings corroborated previous RT-based findings by Remington et al. (2000)⁴⁵ and indicated that increasing the number of aircraft and their associated flight parameters in a systematic fashion exerted greater demands on visuospatial information processing. Furthermore, responses in the second and third blocks of AT-CDT test trials were also faster, respectively, than those in the first block. This main effect of trial block was mainly driven by faster responses to low and medium ATD trials shown in the subsequent trial blocks (second and third) than to trials of the same type shown in the first trial block. These findings showed that a spatial learning effect, an effect commonly seen in virtual navigation experiments^{61–63}, applied to participants' performance in the low and medium ATD conditions. Specifically, this effect means that participants gained some knowledge of the conflict events (i.e., knowledge about the general configuration of the conflict geometry and associated topological relationships, as shown in Fig. 3) and applied it to improve their performance on subsequent trials.

Taking stock of all these findings, perhaps the most surprising was the revelation that SOT mean RTs correlated significantly with AT-CDT mean tLMS in the high ATD condition only. Considering that all participants also responded the slowest, on

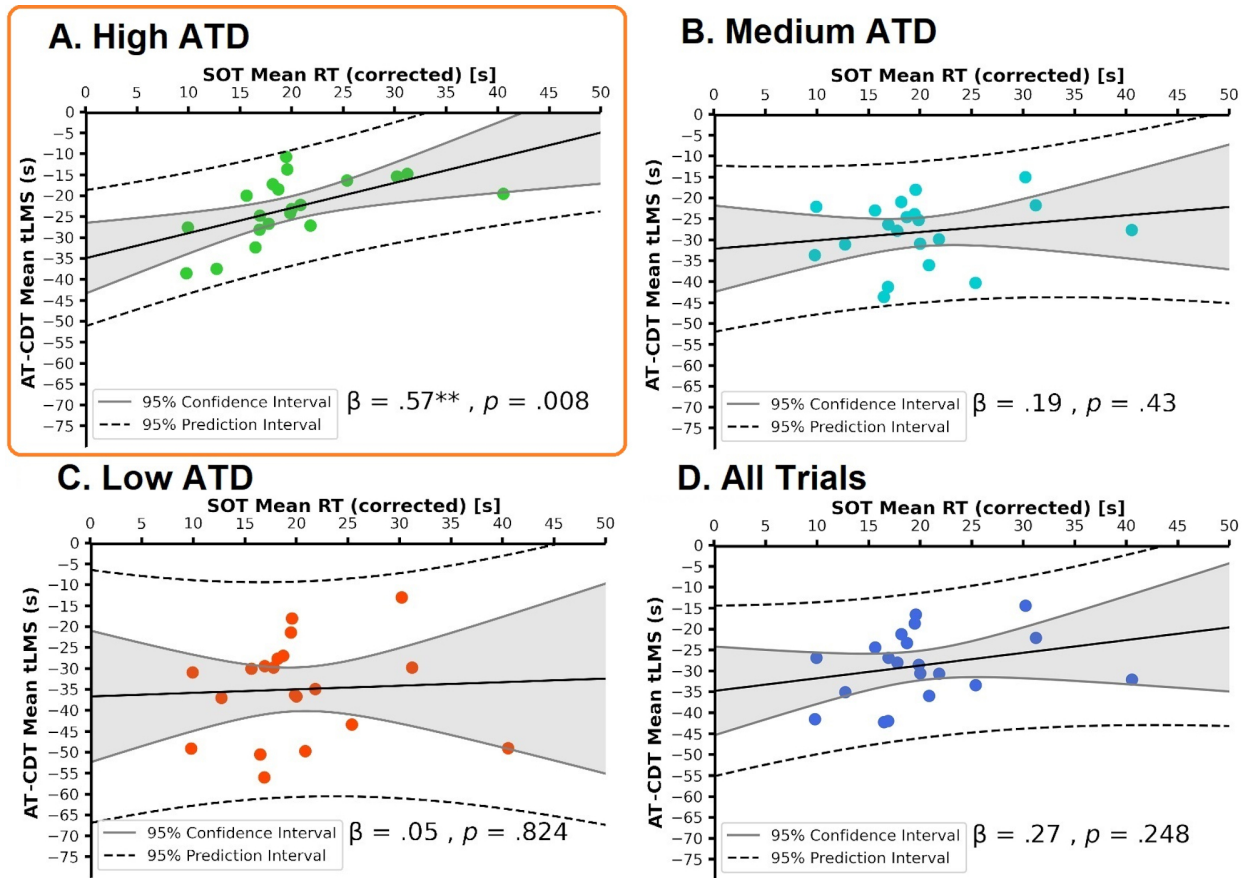


Figure 7. Scatter-plots showing the linear relationships between SOT mean corrected RT and AT-CDT mean tLMS under each of (A–C) three air traffic density (ATD) conditions and over (D) all AT-CDT test trials. Panel A (outlined in orange) shows the significant correlation emanating from the high ATD condition. 95% confidence and prediction intervals are shown. In all plots, no data points lie beyond the 95% prediction interval, showing the absence of outliers.

average, in detecting the correct pairs of conflicting aircraft under the high ATD condition, we postulate that the mental processes inherent to multiobject-directed SOA might have only become relevant for air traffic conflict detection under scenarios with a relatively high number of aircraft. In this experiment, we showed this number to be 14 and interestingly discovered (in a post-hoc observation) that with this number, participants must attend to seven aircraft per conflict event [five on the same flight level – three non-conflicting, two on different flight levels – both non-conflicting]. Coincidentally, the number seven matched the number of distinctive objects forming the object array in the SOT. Henceforth, we postulate that a congruency or relative match in the number of outcome-relevant stimuli between the AT-CDT and SOT might have created the optimal condition for an engagement of mental processes that are commonly activated by both tasks. Within the context of this study, we gave a stipulative definition of "outcome-relevant stimuli" as the average amount of stimuli on a given trial that must be perceived and processed in order to reach a desired or successful outcome. In addition, we considered the possibility that a particular magnitude of *air traffic complexity*, unmeasured by the current AT-CDT, might have created the "optimal condition" for an engagement of shared mental processes. By "air traffic complexity," we take it to mean the amount of perceived difficulty an air traffic scenario presents to an ATCO with respect to ensuring safe and efficient traffic flow⁶⁴.

With respect to the specific mental processes or procedures required for performing the SOT and AT-CDT, we inferred three component processes as essential for accurate performance on both tasks in light of existing research: (i) *visual scanning* [i.e., of 2D object locations and on-screen text instruction (SOT)^{18,19,21,23,24}], of aircraft locations and flight level numbers in the data blocks (AT-CDT)⁴²], (ii) *selective attention* [i.e., to the three objects specified by the on-screen text instruction (SOT)^{18,19,21,23,24}; to clusters of aircraft cruising on the same flight level and the presence of any aircraft heading changes (AT-CDT)^{26,27}], and (iii) *vector computation and mapping* [i.e., visualizing the reference and target directions and mapping them onto the response circle^{17,18,20} (SOT); extrapolating the aircraft flight path or trajectory and mentally mapping them out to check for any potential convergence that may lead to a conflict (AT-CDT)⁴¹]. To elaborate on the interpretations of the findings

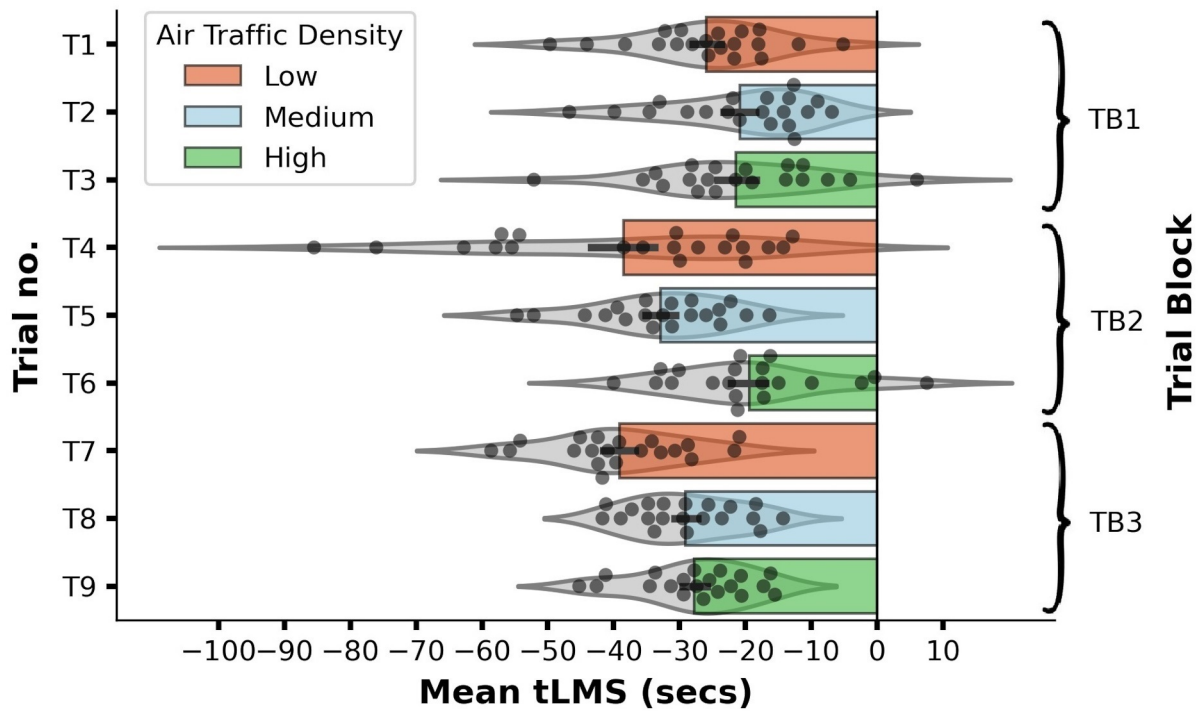


Figure 8. Horizontal bar graph with violin plots showing the distribution of AT-CDT mean tLMS values obtained from each participant on each test trial (T1 – T9). Longer bars represent faster responses associated with quicker detections of conflicts. Each error bar represents $\pm 1 SE$. Zero on the x-axis marks the time-point at which a conflict event occurs. The violin plot overlaid on each bar represents the probability density function (PDF) of the data distribution on each trial. Each PDF was mirrored along the central horizontal axis of each bar. Data distribution on the fourth trial was platykurtic due to the presence of a few fast responders.

stated in the preceding paragraph, it is likely that the cumulative time devoted to executing all these mental processes became comparable between the SOT and AT-CDT in the presence of an approximately equal amount of outcome-relevant stimuli or air traffic complexity in both tasks. We view this as the most logical explanation of the currently observed temporal relationship between SOT and AT-CDT performance and encourage future aviation psychology studies to confirm it with respect to a greater number of outcome-relevant stimuli that are evenly matched between tasks (> 7 per task). Ideally, future studies can employ event-related designs that record the time spent in each stage of information processing for a detailed understanding of the specific type of mental process that is most commonly engaged between tasks.

Together with highlighting these mental processes/procedures, it is also important to evaluate the practical implications of our findings. As mentioned in the introduction, the practical motivation for this study pertains to the goal of selecting ATCOs with suitably high levels of SOA to meet the cognitive demands introduced by FRA implementation, which carries the potential to create more geometrically complex conflicts^{49,50,52}. Here, it is worth noting that the flight paths specified in our AT-CDT did not correspond to fixed routes only and that there were straight-line paths (traversed by non-conflicting aircraft) analogous to free routes that provided shorter flight paths between designated entry and exit points. In this way, the simulated airspace we presented can be seen as a FRA that have aircraft entering from multiple directions and multiple entry points⁵². Seen in this light, we argue that time-based performance on the current SOT can offer a relatively robust prediction of the time taken for correct conflict detections within a FRA with relatively high ATD. Moreover, in view of recent research showing strong inter-relationships between the number of daily flights (related to ATD in a specified airspace), air traffic complexity, and ATCO mental workload⁶⁵, we further argue that the administration of the current SOT can benefit the selection of prospective ATCOs who can detect conflicts accurately under situations with relatively high air traffic complexity and mental workload demands. By "selection of prospective ATCOs," we refer to the identification of ATCO candidates at the initial screening phase in which psychometric tests like the SOT are administered and must be passed in order for the candidate to gain entry to formal ATC training^{10-12,14}. In recognition that becoming a professional ATCO requires passing through multiple rounds of challenging tests and training^{10,13,66}, we must stress that we regard the current SOT as a useful assessment tool at the preliminary selection stage only. This means that we see it as useful for selecting ATCO candidates who can perform well in detecting simulated

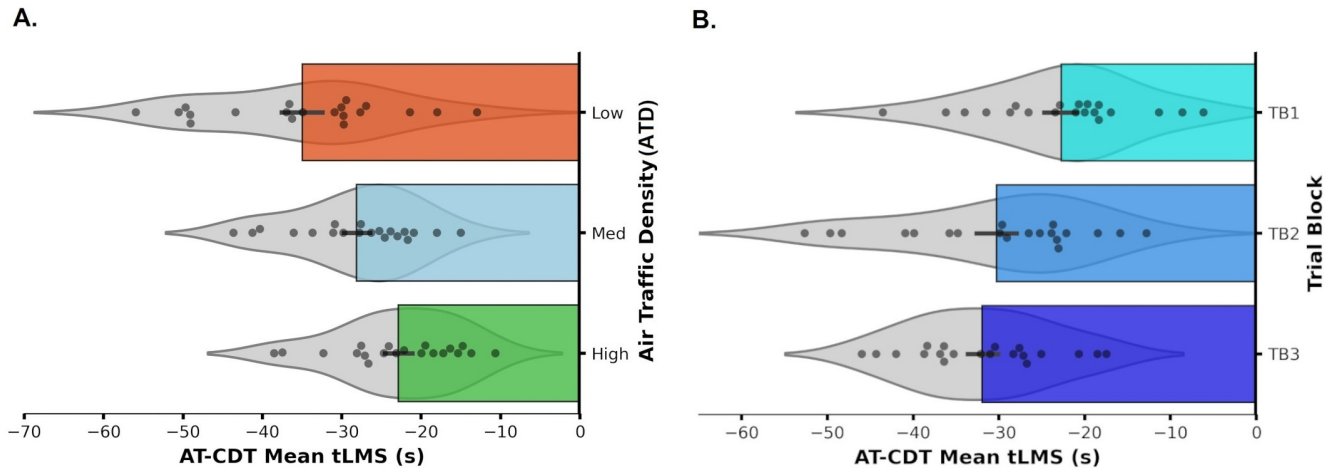


Figure 9. Horizontal bar graph with violin plots showing the distribution of AT-CDT mean tLMS values obtained from each participant under (A) each air traffic density (ATD) condition and within (B) each trial block (TB). Longer bars represent faster responses associated with quicker detections of conflicts. Each error bar represents $\pm 1 SE$. Zero on the x-axis marks the time-point at which a conflict event occurs. The violin plot overlaid on each bar represents the probability density function (PDF) of the data distribution falling within (A) each ATD condition and (B) TB. Each PDF was mirrored along the central horizontal axis of each bar. There was no prominent skewness in data distribution in any ATD condition or TB.

conflicts during ATC training, *not* professionals who can perform well in detecting actual conflicts under real-world operational settings.

In addition, we want to emphasize that our AT-CDT was designed carefully to assess SOA in a controlled fashion without presenting additional variables (e.g., different air speed and flight levels for conflicting aircraft) that could confound the interpretation of the current findings. As such, we recommend future studies to adopt a similar experimental paradigm like ours when designing alternative types of AT-CDTs for assessing other types of cognitive abilities (e.g., spatial attention, processing speed). In practical terms, this means that there must be a systematic control of air traffic variables that can complicate the investigation of common mental processes that are similarly engaged by a predictor task.

An understanding of all these practical implications will not be complete without considering the limitations of this study and we hereby identified a few limitations that can be resolved by future studies. First, owing to the fact that this study was conducted in the midst of the COVID-19 pandemic in 2020, administrative constraints prevented us from collecting a larger and more heterogeneous sample. COVID-19 has hindered participant recruitment in numerous engineering psychology studies⁶⁷ and our study is of no exception. Nevertheless, owing to the fact that our regression analyses involved only one predictor variable and that our scatter-plots showed no outliers or irregularities, we regard a sample size of 20 participants as adequate for showing acceptable statistical power ($\geq 80\%$) and finding significant R^2 values falling within the medium effect range (as shown in Results). For future studies that involve between-subjects analysis and more predictors than ours, a larger sample would be required, and we recommend all researchers who aim to extend our current findings to bear this in mind.

Second, we want to bring attention to the fact that the current SOT does not have a pre-packaged function that records the time one took to complete each trial. Consequently, we could not obtain a precise measurement of the average time taken to complete trials *accurately* within a prescribed pointing error range (e.g., $0^\circ - 90^\circ$). This is also the reason why we computed a corrected mean RT measure for the SOT, a measure that we described as providing an *estimate* of the average time it took to perform a trial accurately. Even though the corrected SOT RT values correlated highly with the uncorrected SOT RT values ($r_{20} = .97$), we deem the use of the corrected RT measure as conceptually important because we wanted to demonstrate the relationship between two sets of temporal measures that factored accuracy into consideration. With all these in mind, we recommend adding codes or functions specifying the recording of the total time elapsed from test onset and the time spent per trial to the current SOT's Java scripts. We also recommend that future versions of SOTs incorporate time-recording as an integral part of their default data recording functions.

Third, we want to mention that ATD is not identical to air traffic complexity, even though a strong positive relationship exists between the two constructs⁶⁵. While ATD can be manipulated easily by increasing the number of aircraft in a designated airspace^{45,57} (as done in this study), air traffic complexity is conceived as a computational index of mental workload and is usually computed using algorithms or computational models based on information related to flight trajectories, number of ATC events, air sector volumes and configurations, etc.^{68,69}. In view of this fact, future aviation psychology studies can consider

Table 5. Post-hoc Tests of Significant AT-CDT Interaction and Main Effects

Sig. Effect	Sig. Post-hoc Pairwise Comparison	$M_{Difference}$ [s]	$SE_{Difference}$ [s]	T_{19}	p (two-tailed)
ATD × TB	Low_ATD ₁₁ > Low_ATD ₇₃	13.14	2.84	4.62***	1.90×10^{-4}
	Medium_ATD ₂₁ > Low_ATD ₄₂	17.67	3.46	5.10***	0.60×10^{-4}
	Medium_ATD ₂₁ > Low_ATD ₇₃	18.25	2.75	6.63***	0.10×10^{-4}
	Medium_ATD ₂₁ > Medium_ATD ₅₂	12.04	2.41	4.99***	0.80×10^{-4}
	Medium_ATD ₈₃ > Low_ATD ₇₃	10.01	1.63	6.13***	0.10×10^{-4}
	^a High_ATD ₃₁ > Low_ATD ₄₂	17.11	4.78	3.58**	.002
	High_ATD ₃₁ > Low_ATD ₇₃	17.69	3.37	5.26***	0.40×10^{-4}
	High_ATD ₆₂ > Low_ATD ₄₂	19.10	4.83	3.93***	0.89×10^{-4}
	High_ATD ₆₂ > Medium_ATD ₅₂	13.48	2.81	4.80***	1.30×10^{-4}
	High_ATD ₆₂ > Low_ATD ₇₃	19.69	2.88	6.84***	0.10×10^{-4}
	High_ATD ₉₃ > Low_ATD ₇₃	11.38	1.72	6.61***	0.10×10^{-4}
ATD	High_ATD _M > Low_ATD _M	12.08	2.14	5.64***	0.20×10^{-4}
	High_ATD _M > Med_ATD _M	5.26	1.66	3.16**	.005
	Medium_ATD _M > Low_ATD _M	6.83	1.36	5.01***	0.80×10^{-4}
TB	TB ₁ > TB ₂	8.01	2.11	3.80***	.001
	TB ₁ > TB ₃	9.71	2.53	3.84***	.001

Note. *** $p \leq .001$. ** $p < .01$. Sig. = Significant. ^aMarginally Significant. ATD = Air Traffic Density. TB = Trial Block. M = Mean. SE = Standard Error. For ATD x TB effect, each subscript indicates trial no. followed by TB no. For ATD main effect, subscript M denotes "mean." For TB main effect, each subscript indicates block no.

using algorithms or computational methods to compute suitable measures of air traffic complexity that allow an experimental manipulation of mental workload in a continuous or parametric fashion.

To summarize, this study is the *first* attempt at using a multiobject-directed, navigationally relevant, and computerized SOT in predicting conflict detection performance in a simulated FRA environment among young adults who can become prospective ATCOs. The key findings are highlighted as follows: (i) We found that SOT RT-based performance predicted AT-CDT tLMS-based performance significantly under a high ATD condition and explained this effect with respect to three component mental processes – visual scanning, selective attention, vector computation and mapping – that were most likely activated optimally under experimental conditions with approximately equal numbers of outcome-relevant stimuli. (ii) On the practical end, we proposed the current SOT to be a useful tool for predicting the time spent on accurate conflict detection and recommended its use in identifying ATCO candidates who have relatively high potential to succeed at conflict detection in a simulated FRA environment during ATC training. (iii) We showed that a SOT response time measure corrected for pointing errors was useful for investigating spatial cognition processes and argued that future versions of SOTs should incorporate time recording as part of their default data recording functions.

In addition, by bringing attention to some noticeable limitations of the current study, we express our hope that they can be addressed in future studies so that greater opportunities can be created for psychometric and human factors research in the ATM domain. In view of the dearth of inter-disciplinary approaches to investigating human spatial cognition and navigation, and virtually no attention to the topic of navigational control in the mainstream spatial navigation literature⁷⁰, we regard our study as an opening up "bridge" of contact between two seemingly divided research domains: spatial navigation and ATM. Henceforth, we hope that this study will become the first of many interdisciplinary attempts at investigating the mental processes involved in both spatial navigation and ATM.

References

1. Trapsilawati, F., Herliansyah, M. K., Nugraheni, A. S. A. N. S., Fatikasari, M. P. & Tissamodie, G. EEG-based analysis of air traffic conflict: Investigating controllers' situation awareness, stress level and brain activity during conflict resolution. *The J. Navig.* **73**, 678–696 (2020). <https://doi.org/10.1017/S0373463319000882>.
2. Goh, S. K. *et al.* Construction of air traffic controller's decision network using error-related potential. In *International Conference on Human-Computer Interaction*, 384–393 (Springer, 2019). https://doi.org/10.1007/978-3-030-22419-6_27.
3. Radüntz, T., Fürstenau, N., Mühlhausen, T. & Meffert, B. Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps. *Front. Physiol.* **11** (2020). <https://doi.org/10.3389/fphys.2020.00300>.
4. Zhang, J., E, X., Du, F., Yang, J. & Loft, S. The difficulty to break a relational complexity network can predict air traffic controllers' mental workload and performance in conflict resolution. *Hum. Factors* 1–14 (2019). <https://doi.org/10.1177/0018720819880646>.
5. Bruder, C. & Hasse, C. Differences between experts and novices in the monitoring of automated systems. *Int. J. Ind. Ergonomics* **72**, 1–11 (2019). <https://doi.org/10.1016/j.ergon.2019.03.007>.
6. Ohneiser, O., Jauer, M., Gürlük, H. & Springborn, H. Attention guidance prototype for a sectorless air traffic management controller working position. In *Proceedings of Deutscher Luft- und Raumfahrtkongress (DLRK)*, DocumentID: 480189, 1–7 (2018). <https://doi.org/10.25967/480189>.
7. EUROCONTROL. White paper on human factors integration and ATM system design. Tech. Rep. 2019, EUROCONTROL (2019). Retrieved from <https://www.eurocontrol.int/publication/human-factors-integration-atm-system-design>.
8. Hubley, A. M. & Zumbo, B. D. Psychometric characteristics of assessment procedures: An overview. In *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*, APA handbooks in psychology, 3–19 (American Psychological Association, 2013). <https://doi.org/10.1037/14047-001>.
9. Zumbo, B. D. & Chan, E. K. Setting the stage for validity and validation in social, behavioral, and health sciences: Trends in validation practices. In Zumbo, B. D. & Chan, E. K. (eds.) *Validity and validation in social, behavioral, and health sciences*, 3–8 (Springer, 2014). https://doi.org/10.1007/978-3-319-07794-9_1.
10. Eißfeldt, H. & Heil, M. C. *Staffing the ATM system: The selection of air traffic controllers* (Routledge, 2017). <https://doi.org/10.4324/9781315242538>.
11. Rathje, H., Golany, Z. & Eißfeldt, H. Pan-European selection test battery for air traffic control applicants. In Goeters, K.-M. (ed.) *Aviation Psychology: Practice and Research*, 169–199 (Ashgate Publishing, 2004).
12. EUROCONTROL. First european air traffic controller selection test (FEAST) package: Information for candidates applying to become student air traffic controllers. Tech. Rep. 2018, EUROCONTROL (2018). Retrieved from https://feast-info.eurocontrol.int/docs/Information_for_Candidates_ClassicPlus_ENGLISH_040419_ANNEX.pdf.
13. Broach, D. Recovery of the FAA air traffic control specialist workforce, 1981–1992. Tech. Rep. DOT/FAA/AM-98/23, Federal Aviation Federation Civil Aeromedical Institute, Oklahoma City, OK (1998). <https://apps.dtic.mil/sti/citations/ADA355135>.
14. Broach, D. *et al.* The validity of the air traffic selection and training (AT-SAT) test battery in operational use. Tech. Rep. DOT/FAA/AM-16/2, Federal Aviation Federation (FAA) Civil Aeromedical Institute, Oklahoma City, OK (2013). <https://apps.dtic.mil/sti/citations/ADA583743>.
15. Kuchar, J. K. & Yang, L. C. A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intell. Transp. Syst.* **1**, 179–189 (2000). <https://doi.org/10.1109/6979.898217>.
16. Brooker, P. STCA, TCAS, airproxes and collision risk. *The J. Navig.* **58**, 389–404 (2005). <https://doi.org/10.1017/S0373463305003334>.
17. Gunalp, P., Moossaian, T. & Hegarty, M. Spatial perspective taking: Effects of social, directional, and interactive cues. *Mem. & Cogn.* **47**, 1031–1043 (2019). <https://doi.org/10.3758/s13421-019-00910-y>.
18. Gunalp, P. N. *Examining perspective taking and its relation to map use and environment learning*. Doctoral dissertation, University of California, Santa Barbara, CA (2020). Retrieved from <https://escholarship.org/uc/item/5n75n3qb>.
19. Friedman, A., Kohler, B., Gunalp, P., Boone, A. P. & Hegarty, M. A computerized spatial orientation test. *Behav. Res. Methods* **52**, 799–812 (2020). <https://doi.org/10.3758/s13428-019-01277-3>.

20. Kozhevnikov, M. & Hegarty, M. A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. & Cogn.* **29**, 745–756 (2001). <https://doi.org/10.3758/BF03200477>.
21. Kozhevnikov, M., Motes, M. A., Rasch, B. & Blajenkova, O. Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Appl. Cogn. Psychol.* **20**, 397–417 (2006). <https://doi.org/10.1002/acp.1192>.
22. Hegarty, M. & Waller, D. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* **32**, 175–191 (2004). <https://doi.org/10.1016/j.intell.2003.12.001>.
23. Zhong, J. Y. *Three types of environmental representations and individual differences in spatial navigation*. Master's thesis, National University of Singapore, Singapore (2013). Retrieved from <http://scholarbank.nus.edu.sg/handle/10635/47243>.
24. Lun, W. M., Zhong, J. Y. & Kozhevnikov, M. *The role of spatial skills and processes in navigation performance*. Undergraduate honors thesis, National University of Singapore, Singapore (2013). Retrieved from <https://doi.org/10.13140/RG.2.2.24302.84805/1>.
25. Goeters, K.-M., Maschke, P. & Eißfeldt, H. Ability requirements in core aviation professions: Job analysis of airline pilots and air traffic controllers. In Goeters, K.-M. (ed.) *Aviation Psychology: Practice and Research*, 99–119 (Ashgate Publishing, 2004).
26. Kissing, D. S. & Eißfeldt, H. What about the next generation? assessing experts' judgements of human abilities required for working in a future ATC environment. In Ahram, T., Karwowski, W. & Marck, T. (eds.) *Proceedings of the 5th International Conference in Applied Human Factors and Ergonomics (AHFE)*, 7785–7795 (2014).
27. Eißfeldt, H., Heintz, A. & Kissing, D. S. Selection requirements to work in future ATM systems. In *16th International Symposium on Aviation Psychology*, 315–320 (2011). https://corescholar.libraries.wright.edu/isap_2011/62.
28. Yang, J. & Zhang, K. The impact of spatial ability on air traffic controller situation awareness and mental workload. In *Logistics: The Emerging Frontiers of Transportation and Development in China*, 4460–4467 (2009). [https://doi.org/10.1061/40996\(330\)651](https://doi.org/10.1061/40996(330)651).
29. Grasshoff, D., Hasse, C., Bruder, C. & Eißfeldt, H. On the development of a monitoring test for the selection of aviation operators. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, 335–344 (Springer, 2015). https://doi.org/10.1007/978-3-319-20373-7_32.
30. Ackerman, P. L. & Kanfer, R. Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *J. Appl. Psychol.* **78**, 413–432 (1993). <https://doi.org/10.1037/0021-9010.78.3.413>.
31. Ackerman, P. L., Kanfer, R. & Goff, M. Cognitive and noncognitive determinants and consequences of complex skill acquisition. *J. Exp. Psychol. Appl.* **1**, 270–304 (1995). <https://doi.org/10.1037/1076-898X.1.4.270>.
32. Ackerman, P. L. & Cianciolo, A. T. Ability and task constraint determinants of complex task performance. *J. Exp. Psychol. Appl.* **8**, 194–208 (2002). <https://doi.org/10.1037/1076-898X.8.3.194>.
33. FEAST test practice: Master the FEAST 1 test with these examples and through prep and advance to the next step. <https://www.jobtestprep.com/feast-test-practice>. Accessed: 2020-11-14.
34. SkyTest® preparation software for European ATCO screenings - FEAST 1: Spot the side. <https://www.skytest.com/modules/spot-the-side>. Accessed: 2021-01-07.
35. Hegarty, M., Burte, H. & Boone, A. P. Individual differences in large-scale spatial abilities and strategies. In Montello, D. R. (ed.) *Handbook of behavioral and cognitive geography*, 231–246 (Edward Elgar Publishing, 2018). <https://doi.org/10.4337/9781784717544.00022>.
36. Holmes, C. A., Marchette, S. A. & Newcombe, N. S. Multiple views of space: Continuous visual flow enhances small-scale spatial learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 851–861 (2017). <https://doi.org/10.1037/xlm0000346>.
37. Muffato, V. & Meneghetti, C. Learning a path from real navigation: The advantage of initial view, cardinal north and visuo-spatial ability. *Brain Sci.* **10**, 1–16 (2019). <https://doi.org/10.3390/brainsci10040204>.
38. Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F. & Epstein, R. A. Variations in cognitive maps: Understanding individual differences in navigation. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 669–682 (2014). <https://doi.org/10.1037/a0035261>.
39. Zhong, J. Y. & Kozhevnikov, M. Relating allocentric and egocentric survey-based representations to the self-reported use of a navigation strategy of egocentric spatial updating. *J. Environ. Psychol.* **46**, 154–175 (2016). <https://doi.org/10.1016/j.jenvp.2016.04.007>.

40. Jong, S. *Rethinking spatial abilities: A multimethod examination of its context-dependent nature and whether tests require increased conceptual, contextual, and perceptual similarity to usage context*. Doctoral dissertation, Carleton University, Ottawa, ON, Canada (2015). <https://doi.org/10.22215/etd/2015-10721>.
41. Loft, S., Bolland, S., Humphreys, M. S. & Neal, A. A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. *J. Exp. Psychol. Appl.* **15**, 106–124 (2009). <https://doi.org/10.1037/a0016118>.
42. Rantanen, E. M. & Nunes, A. Hierarchical conflict detection in air traffic control. *The Int. J. Aviat. Psychol.* **15**, 339–362 (2005). https://doi.org/10.1207/s15327108ijap1504_3.
43. He, C. & Hegarty, M. How anxiety and growth mindset are linked to navigation ability: Impacts of exploration and GPS use. *J. Environ. Psychol.* **71**, 101475 (2020). <https://doi.org/10.1016/j.jenvp.2020.101475>.
44. Simpson, M. *Scale and space: Representations in immersive virtual reality*. Doctoral dissertation, Pennsylvania State University, State College, PA (2020). Retrieved from <https://etda.libraries.psu.edu/catalog/17663mbs278>.
45. Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M. & Romera, M. Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Hum. Factors* **42**, 349–366 (2000). <https://doi.org/10.1518/001872000779698105>.
46. Schmitt, D.-R., Werner, K., Morlang, F., Hampe, J. & Kaltenhäuser, S. Satellite based ADS-B for commercial space flight operations. In *35th Space Symposium, Technical Track, Colorado Springs, CO, USA*, 1–7 (2018). Retrieved from https://www.spacefoundation.org/tech_track_papers/satellite-based-ads-b-for-commercial-space-flight-operations/.
47. Aneeka, S. & Zhong, Z. NOX and CO₂ emissions from current air traffic in asean region and benefits of free route airspace implementation. *J. Appl. Phys. Sci.* **2**, 32–36 (2016). <https://doi.org/10.20474/japs-2.2.1>.
48. Bucuroiu, R. A free route airspace for Europe. In *EUROCONTROL Workshop On Free Route Airspace Marked By New South East Common Sky Initiative: EUROCONTROL*, 1–12 (2017). Retrieved from <https://www.worldatmcongress.org/Uploads/2017%20Presentations/2%20FREE%20ROUTE-Razvan.pdf>.
49. Antulov-Fantulin, B., Juričić, B., Radišić, T. & Rogošić, T. Free route airspace for efficient air traffic management. *Eng. Power: Bull. Croat. Acad. Eng.* **15**, 10–17 (2020). <https://hrcak.srce.hr/244895>.
50. Gaxiola, C. A. N., Barrado, C., Royo, P. & Pastor, E. Assessment of the North European free route airspace deployment. *J. Air Transp. Manag.* **73**, 113–119 (2018). <https://doi.org/10.1016/j.jairtraman.2018.08.008>.
51. Pejović, T., Lazarovski, A. & Pavlovic, G. Impact of free route airspace implementation on safety performance: Ex-post analysis of northern europe free route airspace (NEFRA). 1–7 (9th SESAR Innovation Days, 2019). Retrieved from https://www.sesarju.eu/sites/default/files/documents/sid/2019/papers/SIDs_2019_paper_49.pdf.
52. Schäfer, D. & Modin, E. A human factors perspective on free routing and airborne separation assurance in the Mediterranean airspace. In *5th USA/Europe Air Traffic Management Research and Development Seminar*, 1–11 (EUROCONTROL, 2003). Retrieved from https://atmseminar.eurocontrol.fr/past-seminars/5th-seminar-budapest-hungary-june-2003/papers/paper_064/view.
53. Money, J., Alexander, D. & Walker, H. T. A standardized road-map test of direction sense: Manual (Johns Hopkins University Press, 1965).
54. Civil Aviation Authority of Singapore (CAAS): Career as an air traffic control officer. <https://www.caas.gov.sg/who-we-are/careers/opportunities/career-as-an-air-traffic-control-officer>. Accessed: 2020-12-30.
55. Elman, I. *et al.* Neurological soft signs in individuals with pathological gambling. *PLoS ONE* **8**, e60885 (2013). <https://doi.org/10.1371/journal.pone.0060885>.
56. Elman, I. *et al.* The money road map test (RMT). *PLoS ONE* (2015). <https://doi.org/10.1371/journal.pone.0060885.g003>.
57. Brookings, J. B., F.Wilson, G. & R.Swain, C. Psychophysiological responses to changes in workload during simulated air traffic control. *Biol. Psychol.* **42**, 361–377 (1996). [https://doi.org/10.1016/0301-0511\(95\)05167-8](https://doi.org/10.1016/0301-0511(95)05167-8).
58. Ramseyer, G. C. Testing the difference between dependent correlations using the Fisher Z. *The J. Exp. Educ.* **47**, 307–310 (1979). <https://doi.org/10.1080/00220973.1979.11011698>.
59. Faul, F., Erdfelder, E., Buchner, A. & Lang, A. G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **41**, 1149–1160 (2009). <https://doi.org/10.3758/BRM.41.4.1149>.
60. Kimura, D. *Sex and cognition* (The MIT Press, 1999). <https://psycnet.apa.org/record/1999-02755-000>.
61. Zhong, J. Y. & Moffat, S. D. Age-related differences in associative learning of landmarks and heading directions in a virtual navigation task. *Front. Aging Neurosci.* **8** (2016). <https://doi.org/10.3389/fnagi.2016.00122>.

62. Zhong, J. Y. *et al.* The application of a rodent-based Morris water maze (MWM) protocol to an investigation of age-related differences in human spatial learning. *Behav. Neurosci.* **131**, 470–482 (2017). <https://doi.org/10.1037/bne0000219>.
63. Reynolds, N. C., Zhong, J. Y., Clendinen, C. A., Moffat, S. D. & Magnusson, K. R. Age-related differences in brain activations during spatial memory formation in a well-learned virtual Morris water maze (vMWM) task. *NeuroImage* **202**, 116069 (2019). <https://doi.org/10.1016/j.neuroimage.2019.116069>.
64. Hilburn, B. Cognitive complexity in air traffic control - A literature review. Tech. Rep. 2004-007, EEC Note, vol. 4, EUROCONTROL (2004). Retrieved from https://www.eurocontrol.int/eec/public/standard_page/DOC_Report_2004_007.html.
65. Pejovic, T., Netjasov, F. & Crnogorac, D. Relationship between air traffic demand, safety and complexity in high-density airspace in Europe. In Castán, J. A. P. & Álvaro Rodríguez Sanz (eds.) *Risk Assessment in Air Traffic Management*, chap. 2 (IntechOpen, Rijeka, 2019). <https://doi.org/10.5772/intechopen.88801>.
66. Conzelmann, K., Heintz, A. & Eißfeldt, H. A large scale validation study on air traffic controller selection and training – Design, challenges and results. In *16th International Symposium on Aviation Psychology*, 375–380 (2011). Retrieved from https://corescholar.libraries.wright.edu/isap_2011/52/.
67. Feil-Seifer, D., Haring, K. S., Rossi, S., Wagner, A. R. & Williams, T. Where to next? the impact of COVID-19 on human-robot interaction research. *ACM Transactions on Human-Robot Interact.* **10**, 1–7 (2020). <https://doi.org/10.1145/3405450>.
68. Suárez, N. *et al.* Quantifying air traffic controller mental workload. In *4th SESAR Innovation Days*, 1–6 (2014). Retrieved from https://www.sesarju.eu/sites/default/files/SID_2014-03.pdf.
69. Prandini, M., Blom, H. A. P. & Bakker, G. J. Air traffic complexity and the interacting particle system method: An integrated approach for collision risk estimation. In *Proceedings of the 2011 American Control Conference*, 2154–2159 (2011). Retrieved from <https://doi.org/10.1109/ACC.2011.5991305>.
70. Ekstrom, A. D., Spiers, H. J., Bohbot, V. D. & Rosenbaum, R. S. *Human Spatial Navigation* (Princeton University Press, 2018).

Acknowledgements

This research is supported by the National Research Foundation, Singapore, and the Civil Aviation Authority of Singapore, under the Aviation Transformation Programme. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and the Civil Aviation Authority of Singapore. We thank Dennis K. L. Ng (Manager, ATMRI radar simulation lab) for technical advice and assistance in the design and programming of the AT-CDT and H. Phuoc Dang (Research Associate, ATMRI) for providing us with real-world data about air traffic density. We also thank Wei Yao and Mohamed Ashqar (undergraduate research assistants) for assistance with participant recruitment and programming the extraction and organization of AT-CDT data.

Author Contributions

J.Z. and S.K. drafted the manuscript. J.Z., S.K., and S.A. conceived the experiment. J.Z., S.K., and C.J. conducted the experiment, J.Z. analyzed results, generated all tables and some figures. C.J. organized statistical data and generated the majority of figures. All authors reviewed, edited, and approved the manuscript for submission.

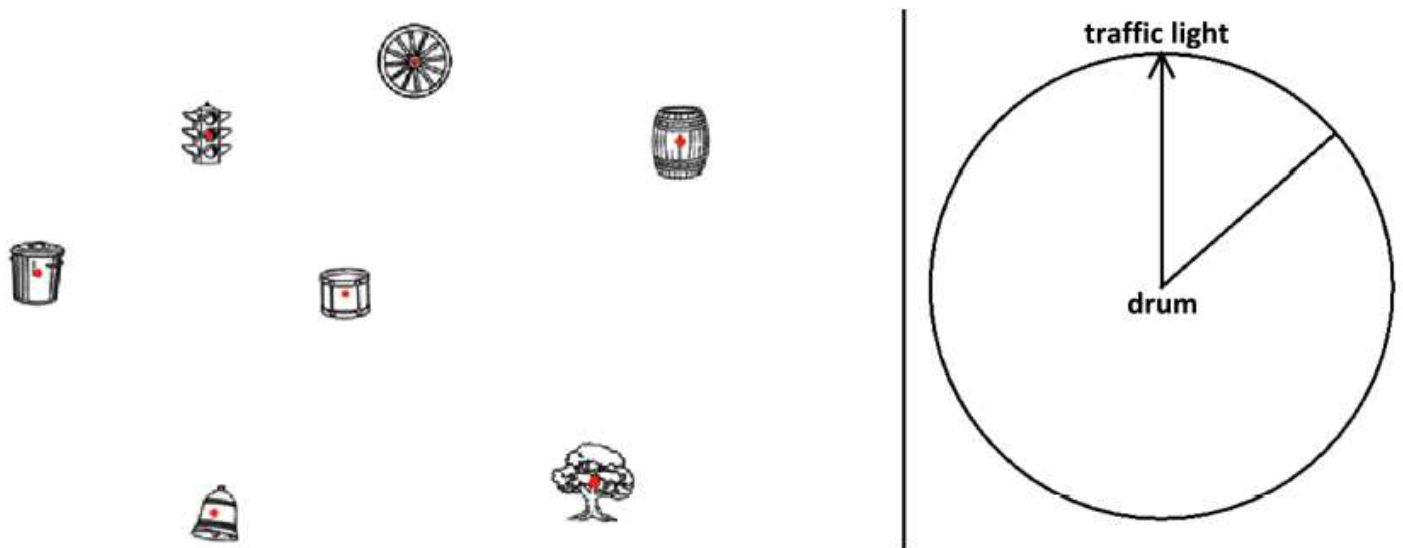
Competing Interests

All authors are employees of Nanyang Technological University (NTU). There are no competing interests concerning the research, authorship, and/or publication of this article. This study was approved by the NTU IRB (Reg. no. 200604393R) and conducted in the absence of any personal, commercial or financial relationships that could be construed as a potential conflict of interest. All methods were carried out in accordance with relevant ethical and institutional guidelines.

Additional information

Correspondence and requests for materials should be addressed to J.Z. and S.A.

Figures



Imagine you are standing at the **drum** and facing the **traffic light**. Point to the **wheel**.
Please press ENTER when finished.

Figure 1

Screenshot of a practice trial in Friedman et al.'s (2020) SOT. For clarity of presentation, each 2D object are shown in higher contrast, verbal instructions and object labels are also enlarged. The target direction in the upper right quadrant of the response circle (derived from the computer mouse actions of the lead author) shows the estimated position of the wheel relative to an initial heading that is to be imagined. [Reproduced with permission under the Creative Common Attribution 4.0 license.]



Figure 2

2K radar monitor (2048 pixels x 2048 pixels) used for showing simulated air traffic data. The air traffic scenario shown was taken from the third AT-CDT test trial – the first high high air traffic density (ATD) condition with 14 aircraft. The left panel shows the 20 mm ruler (boxed on bottom left corner of the radar display) that indicates 5.0 nm separation between a pair of aircraft while the right panel shows details from an enlarged data block of an aircraft (boxed on the right side of the radar display) appearing in the eastern sector of the controlled airspace. The university logo and the full name of the research institute were displayed at the top right corner of the radar display.

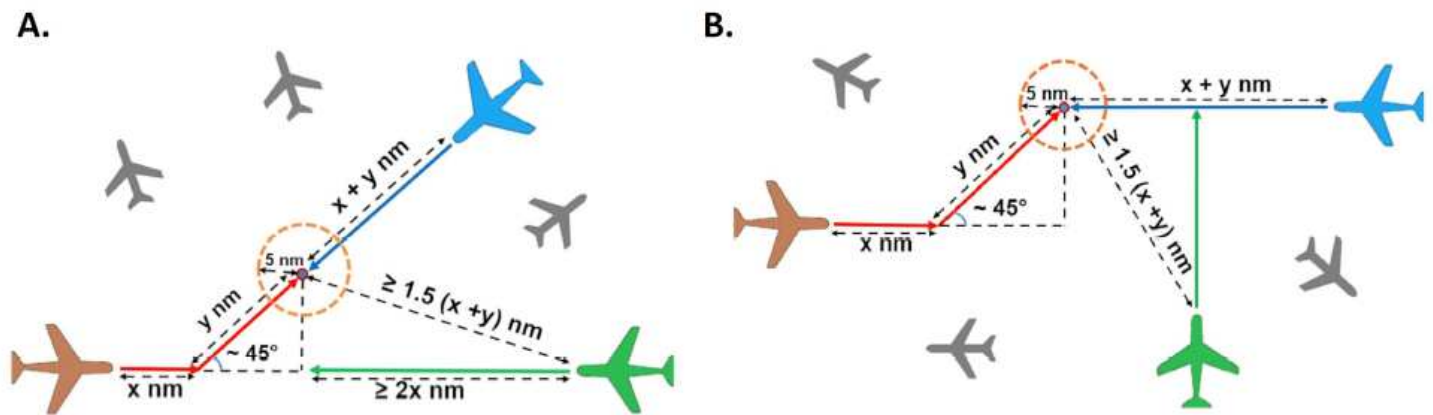


Figure 3

Conflict event models A (left panel) and B (right panel). Algebraic symbols and equations specify the length of the path segments of the air routes (not drawn to scale) and their topological relationships. These routes were kept hidden from participants' view. Colored aircraft represent aircraft cruising on the same flight level while grey aircraft in the background represent aircraft cruising at flight levels different from that of the former cluster. On the same flight level, a non-conflicting aircraft in each model (colored in green) was placed initially at a safe distance away from the conflict zone, which is represented by the dotted circle with a radius of 5.0 nm. The total number of non-conflicting aircraft located in the vicinity of two conflicting aircraft that constituted each conflict event increased from one in the low air traffic density (ATD) condition to three and five in the medium and high ATD conditions, respectively (see text, for more details).

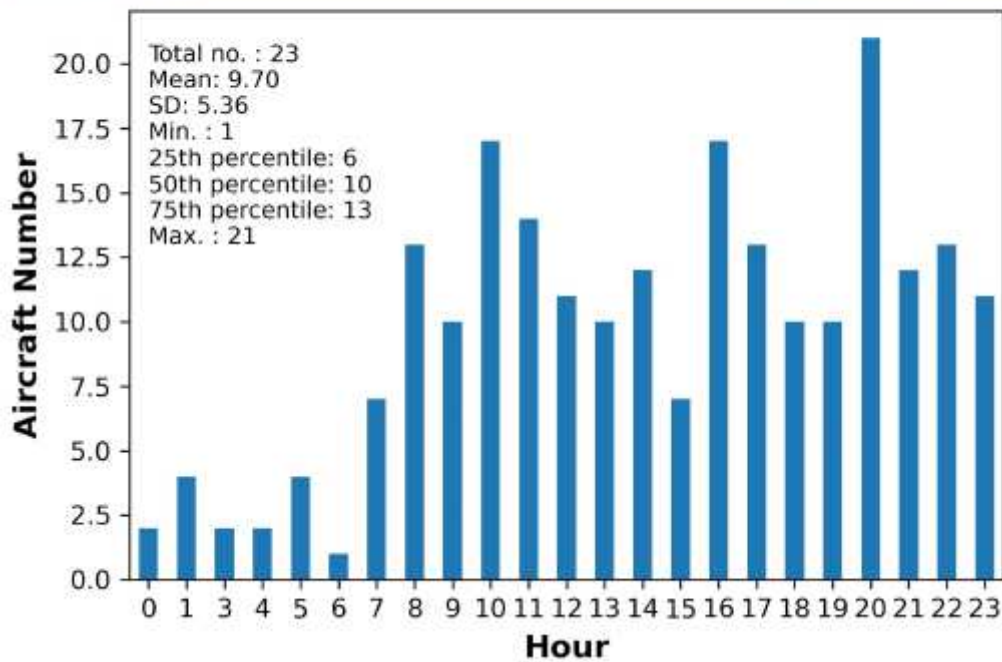
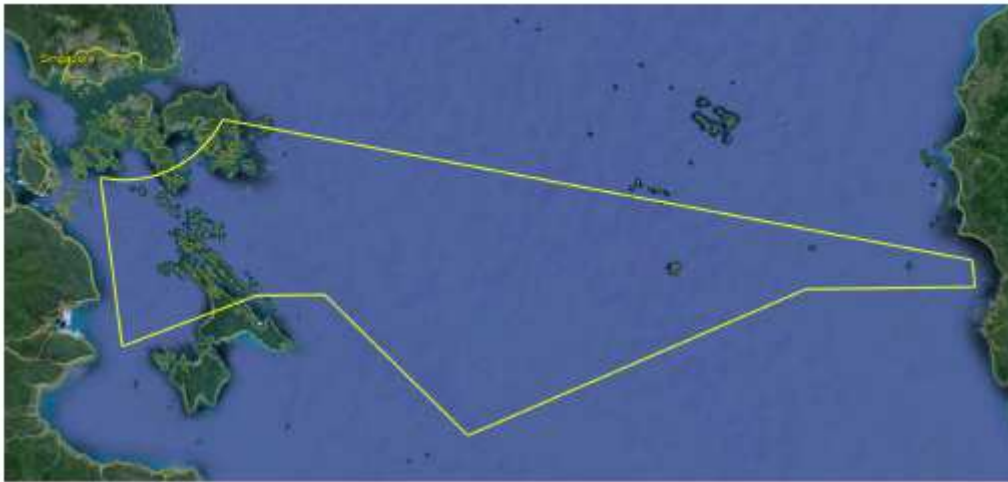


Figure 4

Figure showing the boundaries of an air traffic control (ATC) sector (colored in yellow) covering the airspace in the southeastern region of Singapore island (top panel, generated using Google Earth) and the aircraft number recorded on a randomly selected 24-hour working day in 2019 before COVID-19 outbreak (bottom panel, descriptive statistics shown on the top left corner; SD = Standard Deviation. Min. = Minimum. Max. = Maximum). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

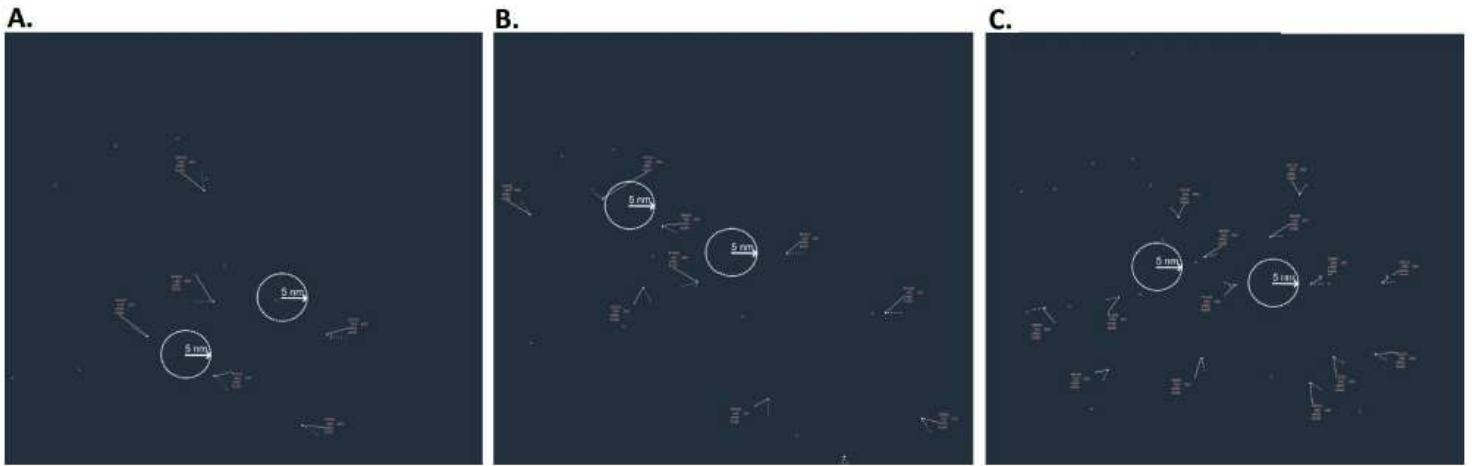


Figure 5

First block of three simulation trials showing three levels of air traffic density (ATD): (A) low (six aircraft), (B) medium (10 aircraft), and (C) high (14 aircraft). White circles, each with a radius of 5.0 nm, delineate the boundaries of the conflict zones. Such circles were not shown during simulation. Through a Latin square design, the pairwise combinations of conflict sites recurred in two different, non-overlapping sequences across the three ATD conditions over the next two blocks of trials (three trials each). The initial arrangement of aircraft in each subsequent ATD condition differed from the arrangement of aircraft shown within the first trial block and varied from each other.

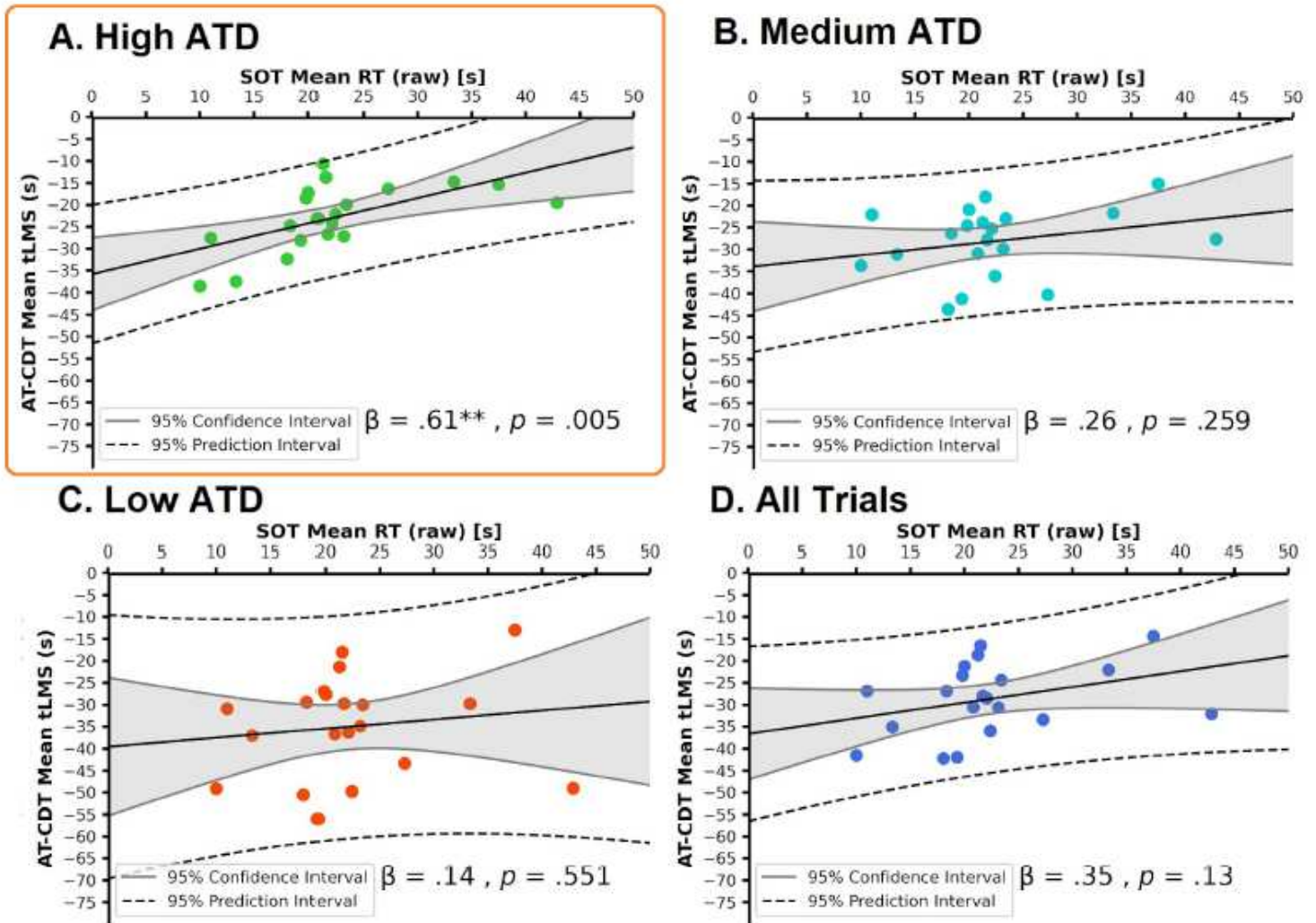


Figure 6

Scatter-plots showing the linear relationships between SOT mean raw RT and AT-CDT mean tLMS under each of (A–C) three air traffic density (ATD) conditions and over (D) all AT-CDT test trials. Panel A (outlined in orange) shows the significant correlation emanating from the high ATD condition. 95% confidence and prediction intervals are shown. In all plots, no data points lie beyond the 95% prediction interval, showing the absence of outliers.

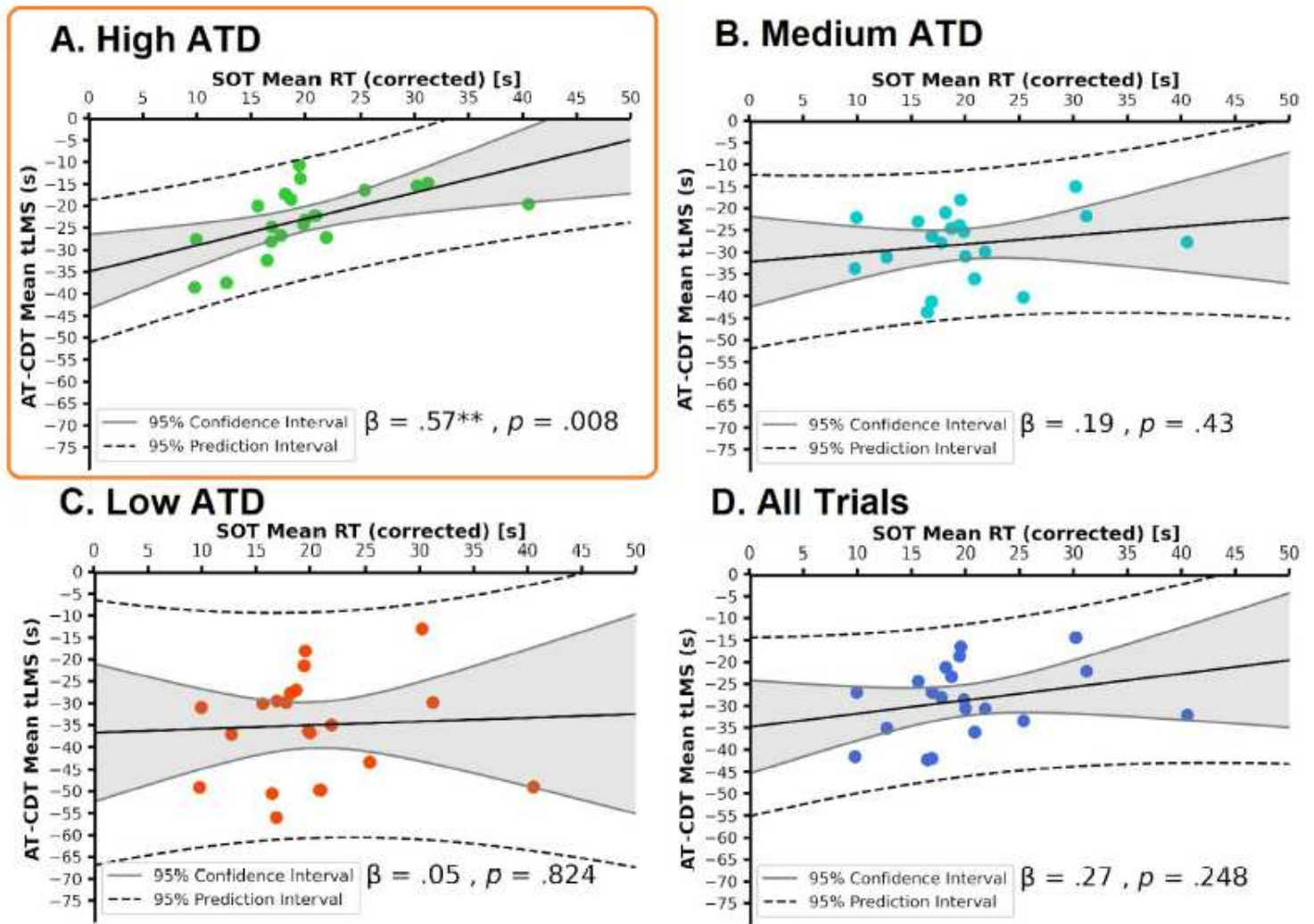


Figure 7

Scatter-plots showing the linear relationships between SOT mean corrected RT and AT-CDT mean tLMS under each of (A–C) three air traffic density (ATD) conditions and over (D) all AT-CDT test trials. Panel A (outlined in orange) shows the significant correlation emanating from the high ATD condition. 95% confidence and prediction intervals are shown. In all plots, no data points lie beyond the 95% prediction interval, showing the absence of outliers.

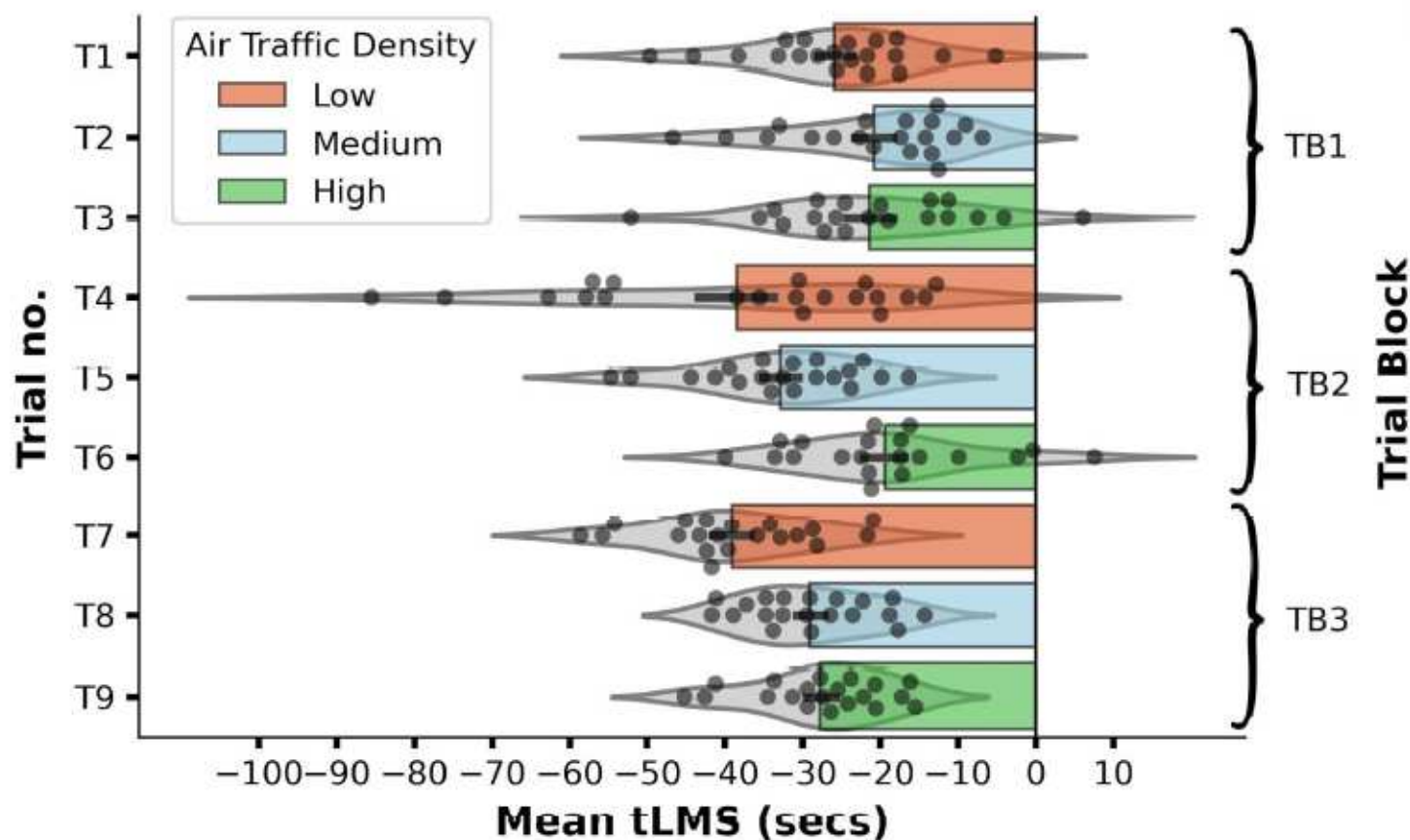


Figure 8

Horizontal bar graph with violin plots showing the distribution of AT-CDT mean tLMS values obtained from each participant on each test trial (T1 – T9). Longer bars represent faster responses associated with quicker detections of conflicts. Each error bar represents ± 1 SE. Zero on the x-axis marks the time-point at which a conflict event occurs. The violin plot overlaid on each bar represents the probability density function (PDF) of the data distribution on each trial. Each PDF was mirrored along the central horizontal axis of each bar. Data distribution on the fourth trial was platykurtic due to the presence of a few fast responders.

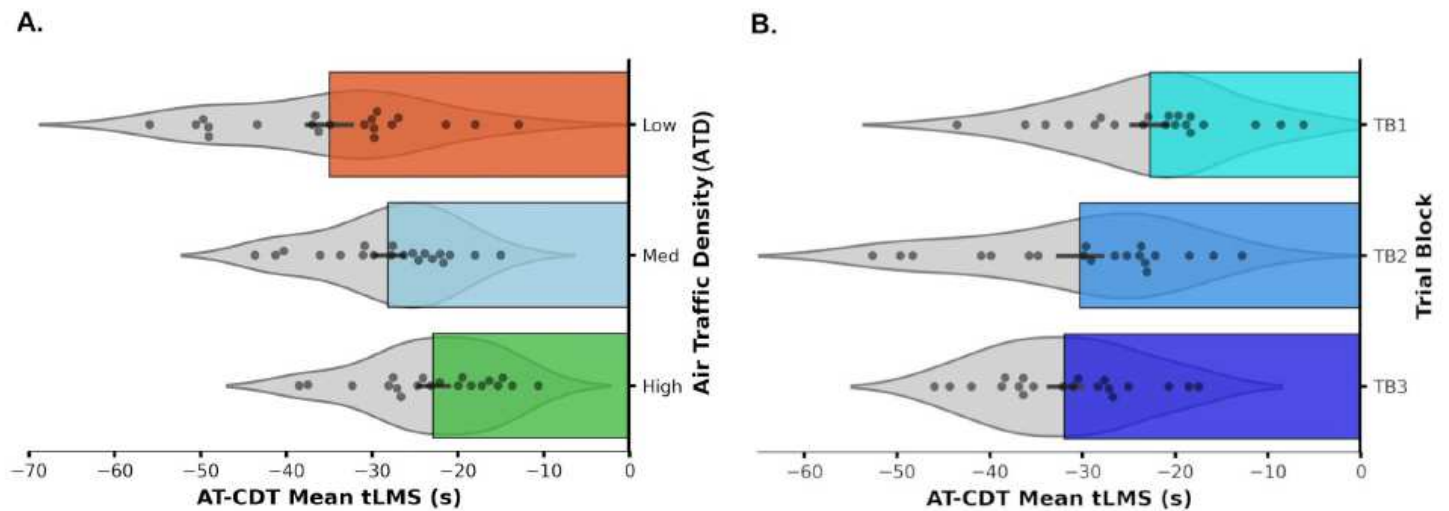


Figure 9

Horizontal bar graph with violin plots showing the distribution of AT-CDT mean tLMS values obtained from each participant under (A) each air traffic density (ATD) condition and within (B) each trial block (TB). Longer bars represent faster responses associated with quicker detections of conflicts. Each error bar represents ± 1 SE. Zero on the x-axis marks the time-point at which a conflict event occurs. The violin plot overlaid on each bar represents the probability density function (PDF) of the data distribution falling within (A) each ATD condition and (B) TB. Each PDF was mirrored along the central horizontal axis of each bar. There was no prominent skewness in data distribution in any ATD condition or TB.