An fMRI investigation of trust-related decision-making associated with the use of short term conflict alert (STCA)

Jimmy Y. Zhong
Air Traffic Management Research Institute (ATMRI)
School of Mechanical and Aerospace Engineering (MAE)
Nanyang Technological University (NTU), Singapore

[FOR SELF-REFERENCE ONLY. DO NOT DISTRIBUTE.]

Abstract

The current proposal aims at a neuroscientific investigation of the magnitudes of trust air traffic controllers (ATCOs) show when using short term conflict alert (STCA) systems with different levels of reliability that can elicit high and low levels of human-automation trust. STCA is an automated warning system used by all ATCOs for the purpose of conflict detection and designed for the primary purpose of ensuring safe separation between any pair of surveillance tracks. The operational use of STCA depends a great deal on the ATCO's trust in the system, and this trust is in term dependent on the perceived system reliability. As different levels of system reliability will engender different levels of uncertainty or mistrust in its use, this proposal adopts an operational definition of trust that involves decision-making under situations with uncertainty and vulnerability. This means that human-automation trust, in the context of STCA use, relates to how well the system can facilitate successful conflict detection under circumstances where uncertainty or unreliability lies in its use.

I. Introduction: Human-machine Trust Defined

Contemporary technology advances in automation, artificial intelligence (AI), and computational algorithms has created great benefits for human operators, along with increases in adverse or unexpected consequences [1], [2], [3]. To ensure safe and seamless human-machine interaction (HMI), trust between humans and machines (in the form of AI or automated tools) has become a widely discussed topic over the past three decades [4], [5], [6]. In particular, within the air traffic management (ATM) domain, human-machine trust has long been recognized as essential for complex ATM systems to deliver the proposed capacity and safety benefits [7], [8].

Irrespective of the types of interactive agents involved, trust, in general terms, is a psychological state that can manifest itself as an intervening variable between particular external situations or experiences (e.g., social interactions) and the human behaviors inherent to such situations [7]. More specific and operational definitions of human-machine trust that has been widely accepted in the current human factors literature concern the perception of trust as: (i) "*the extent to which a user is willing to act on the basis of the information, recommendations, actions, and decisions of a computer-based tool or decision aid*" (p. 11 in [7], adapted from p. 1 in [9])) or (ii) "*the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability*" (p. 6 in [10]). Here, it is worth noting that according to the second operational definition, trust is characterized as decision-making heuristic that machine operators use in situations that are too uncertain or changing rapidly [3], [2].

The current study aims at a neuroscientific investigation of the magnitudes of trust air traffic controllers (ATCOs) show when using short term conflict alert (STCA) systems with different levels of reliability that can elicit high and low levels of human-automation trust. STCA is an automated warning system used by all ATCOs for the purpose of conflict detection and designed for the primary purpose of ensuring safe separation between any pair of surveillance tracks [11]. The operational use of STCA depends a great deal on the ATCO's trust in the system, and this trust is in term dependent on the perceived system reliability of the system [11], [12]. As different levels of system reliability will engender different levels of uncertainty or mistrust in its use, this proposal adopts the second operational definition of trust mentioned above. This means that human-automation trust, in the context of STCA use, relates to how well the system can facilitate successful conflict detection under circumstances where uncertainty or unreliability lies in its use.

II. Research Background: Behavioral and Neuroimaging Studies on Human-Automation Trust

Over the past three decades, numerous studies in the human factors and ergonomics domain had investigated trust between human users and automated tools, which is simply termed as human-automation trust [8], [7], [5], [6]. As AI tools and systems are currently undergoing rapid development and refinements in the ATM domain, we have yet to witness a published ATM study that made a systematic investigation of human-AI trust. Henceforth, this section will focus exclusively on human-automation trust in both ATM and non-ATM domains. Note that by "automation," we refer to the "machine execution of functions"

(p. 286 [13]) and "technology that actively selects data, transforms information, makes decisions, or controls processes" (p. 50 [10]).

Trust has been conceptualized as an essential mechanism for automated ATM systems to deliver the proposed capacity and safety benefits [14]. However, research on trust in ATM systems has been scarce and and has almost all been done using questionnaires that gave a subjective measure of trust [8], [7], [5], [6]. For instance, a popular questionnaire that has been used widely in the ATM domain is the SHAPE Automation Trust Index (SATI) [15], which contains rating scales measuring ATCOs' overall level of trust and the constituents of trust such as reliability, predictability, and understandability. While SATI has been found to be easy to administer, it has been criticized for being unable to provide a multidimensional measure of trust [16]. This means that personality- and individual history-related aspects of trust cannot be assessed by SATI.

Apart from knowing that trust is commonly measured through self-report questionnaires, it is worth noting that a common trust-related issue in the use of automated ATM tools or systems relates to the phenomenon of *complacency*, which occurs when an ATCO places a high level of trust in an automated tool to the negligence or failure of monitoring the "raw" sources of information that provide inputs to the automated system [8], [5]. In the event of overcomplacency or overconfidence in the functionalities of the automated system, system errors may go about undetected [17]. As for the measurement of complacency, the presence of complacency is usually inferred with respect to human performance data reflecting lower levels of attention and shorter monitoring when using automation [18], [19], [20]. For instance, operators have been shown to scan raw information sources less often when using automation than when performing the task manually and when automation reliability is higher rather than lower [19], [20]. In the same vein, when operators were given a tool that contains automated settings for extracting raw information sources, they used it much less often – indicative of faster data extraction – under automation than under manual control [18].

On another end, mistrust of an automated system, the polar opposite of complacency, can also pose significant risks for ATCO performance. Mistrust can either lead an ATCO to "over-monitor" the system in the event of errors [8] or ignore error warnings given by the system entirely [21]. The latter phenomenon has been dubbed the "cry wolf effect" [22] and has led to users of automated tools taking longer to to re-engage the use of automated services after experiencing system failures [23]. Moreover, mistrust is a negative psychological phenomenon that can persist even in the presence of a perfectly reliable automated tool [8]. For instance, using an automated conflict detection advisory in a simulated trajectory-based operation (TBO) airspace, Metzger and Parasumaran (2005) [19] showed that trust ratings of the automated tool were not high even though it was 100% reliable.

In addition to the aforementioned studies, current state-of-the-art studies of human-automation trust have incorporated neuroimaging technologies and techniques. Unlike questionnaires that provide subjective measures of trust, neuroimaging provide objective and real-time measurement of trust and trust-related psychophysiological signals emanating from the human brain and peripheral nervous system [24]. Specifically, these techniques encompass: (i) electroencephalogram (EEG), (ii) functional magnetic resonance imaging (fMRI), (iii) functional near-infrared spectroscopy (fNIRS), (iv) electrocardiogram (ECG), (v) electrodermal activity (EDA) recording, (vi) electromyography (EMG), (vii) electrooculography (EOG), (viii) eye-tracking, and (ix) photoplethysmography (PPG) [25], [24]. Among these, the first three techniques relate to the recording of signals emanating from the brain reflective of changes in underlying neural activity while the latter five techniques involve the recording of signals from the peripheral or autonomic nervous system (i.e., from the heart–ECG, PPG; from the skin and musculature–EDA, EMG; from the eyes–EOG, eye-tracking).

The goal of the current proposed study is to understand the brain mechanisms of human-automation trust in ATM, and hence we shall provide a review of notable brain imaging studies. In recognition of the fact that EEG has been the overwhelming favorite brain imaging technique used by researchers of human-automation trust and that we plan to harness fMRI for this study, extant human-automation studies that utilized these two types of brain imaging techniques shall be the focus of this review. As such, the contributions of EEG and fMRI to human-automation interaction and trust were highlighted in the subsections below. Specifically, because we aimed at learning and adopting the newest state-of-the-art experimental methodologies in our future projects, we concentrated our efforts on a review of notable human-automation trust studies utilizing EEG and fMRI studies over the five years only [2016 – 2021 (present)].

Recent EEG Studies

Over the past five years, there has only been a limited set of EEG studies that investigated human-related trust. We found six noteworthy EEG studies and tabulated their key features – experimental task(s) and design, data analysis technique(s), main findings and implications – in Table I below. A phenomenon common to all these studies pertains to the use of experimental design or tasks that involved the modulation of reliability (or dependability) offered by an automated tool (Table I, column 2). Reliability refers to how consistent an automation can perform in providing accurate information and was largely manipulated through the programming and presentation of automated agents or advisories that can provide information with different probabilities of accuracy [26], [1], [27], [3] or risk-taking tendencies [28], [29].

More importantly, with respect to the brain regions that were activated during human-automation trust-related decision making, these recent EEG studies pinpointed anterior regions such as the (i) lateral prefrontal cortex [29], (ii) anterior cingulate

cortex (ACC) [1], [3], and posterior regions such as the (iii) occipital cortex [29] and (iv) fusiform gyrus (also known as the occipitotemporal gyrus) [3]. Two studies further showed that power increases or variation in the beta frequency band (12 Hz – 35 Hz) were associated with increased levels of trust [27] and making discriminatory judgments between trustworthy and untrustworthy stimuli [26]. Figure 1 shows the brain regions from which trust-related decision making EEG signals were recorded. Figure 1A shows the effective connectivity network uncovered by Sanders et al. (2019) [3] during a participant's detection of mistakes made by the automated advisory while Figure 1B shows the locations of the EEG electrodes, mapped out by Wang et al. (2018) [29], at which significant trust-related activations were found.

Table I: Summary of Key EEG Studies of Human-Automation Trust done over the past Five Years (2016-2020)

Authors (Year of publication)	Experimental Task(s) and Design	Data Analysis Technique(s)	Trust-relevant Findings and Implications
Oh, Seong, Yi, & Park (2020)	2D racing game with varying levels of automated control. Participants' self-reported trust ratings were strongly associated with the levels of automated control provided. Higher trusted emanated from higher automated control.	Power spectral density (PSD) analysis	Power of alpha and beta waves increased during trials with high levels of trust (> 90%). Power of gamma waves increased during trials with low levels of trust (≤ 30%). Observed power increases are collected from electrodes over the entire scalp. Regions of interest were not reported. **Implication:** Alpha waves are related to meditation and anxiety reduction. Beta waves are associated with conscious, attention-demanding thinking processes. Gamma waves are related with anxiety and higher-order information processing.
Jung, Dong, & Lee (2019)	Decision-making task with human- and machine-face agents with different risk-taking personalities. Risk-taking personality of each machine agent was computed using an equation that factored in reward size, probability of correct response, and risk-taking.	Time-frequency analysis EEG power variation analysis	EEG average power variation is higher for both agent's correct (AC) and wrong (AW) decisions when the agent had a human face compared to when it had a machine (robot) face. This pattern of results held in both high and low trust conditions. **Implication:** Participants' neural responses were enhanced by an agent's external human-likenesses. Participants were less sensitive to the participation of less trusted agents if they were externally less human-like.
Sanders, Choo, ... Fitts (2019)	Multi-Attribute Test Battery (MATB) of the US Air Force In the MATB, participants monitored the extent to which four needles of a gauge fell within a nominally acceptable range. Four types of 'algorithms,' varying across two levels of credibility (novice, expert) and two levels of reliability (low - 60\%, high - 80\%) were presented to participants based on a 2 x 2 design over many trials. Experimental design adapted from that used by de Visser et al. (2018) [see row below].	Effective connectivity analysis (ECA), which shows the causal flow of dynamic time-frequency EEG data, were performed on independent components (also known as maximally independent time source series) located in the **fusiform gyrus**, **cingulate cortex**, and **prefrontal cortex**.	Strong flow of time-frequency information from the source node of **fusiform gyrus (FG)** [source node] to the **dorsal anterior cingulate cortex (dACC)** and **posterior cingulate cortex (dPCC)** [terminal nodes] when participants observed the automation to be working successfully under the high reliability condition. **Implication:** ECA findings supported extant notions of the ACC and the PCC as the critical sites for conflict monitoring and error detection and suggested them to be important for calibrating one's trust to the perceived trustworthiness of an automated device.
de Visser, Beatty, ... McDonald (2018)	Flanker task: Popular task used for measuring attention and executive functioning. Participants viewed how a Flanker task was performed based on pre-configured 'algorithms' with low and high degrees of reliability.	Event-related potential (ERP) component analysis	Observational error positivity (oPe), an event-related potential (ERP) arising putatively from the anterior cingulate cortex (ACC), was detected in participants' frontocentral scalp region when they detected erroneous responses made by the 'algorithms.' The amplitude of oPe was higher in the presence of error responses than in the presence of correct responses and this power difference was larger when observing errors made by a high reliability 'algorithm' than when observing errors made by a low reliability 'algorithm.' Interestingly, oPe correlated with questionnaire-based ratings of human-automation trust to a relatively high degree (R^2 = .65). **Implication:** The presence of oPe probably represented a conscious orientation of attention toward unexpected events (i.e., error responses in

			the high reliability 'algorithm's' case). With respect to the correlational findings, this form of conscious awareness could be modulated by trust-related mechanisms.
Wang, Hussein, … Abbass (2018)	Trustworthiness task with three types of artificial agents: (i) low profit/risk-taking, (ii) medium profit/risk-taking, (iii) high profit/risk-taking. 8 scenarios, each with 30 rounds/trials. Number of trustworthy agents varied from 0 to 3 across the scenarios.	Autoregressive modelling of time-dependent EEG dynamics, estimation of EEG complexity using the entropy measure, power spectrum analysis using discrete Fourier transform.	Participants showed higher neural responses in the **frontal** (left, fronto-central, right) and **occipital** cortices with higher perception of trustworthiness of the artificial agent. **Implication:** Frontal activations in general reflected executive functioning and decision-making. Fronto-central activation potentially reflected conflict monitoring, error detection, and decision making about risk and reward. Occipital lobe reflected processing of visual information.
Akash, Hu, Jain, & Reid (2018)	Obstacle detection task. An automated road obstacle sensor was programmed with 50%reliability on half of the experimental trials and 100% reliability on the other half. Brain signals related to trust were collected from the 100% reliability trials while brain signals related to distrust were collected from the 50% reliability trials.	Machine learning (ML)-based modelling classification of EEG and galvanic skin response features.	Two types of "trust-sensor models" (one general and one customized) were created using a common set of EEG and GSR features derived from the entire sample of participants and a smaller set of EEG and GSR features derived from comparisons of significant features between each participant. feature selection. EEG mean frequencies and power variation over the frontocentral region of scalp contributed to the validation of the trust-sensor models. Power variation in the beta band found over the **parietal cortex** (posterior scalp region) was largely observed in the contrast between reliable and unreliable advisories, suggesting that parietal beta waves were involved in the processing of trust-related perceptual information.

Note. Names of brain regions were bolded.

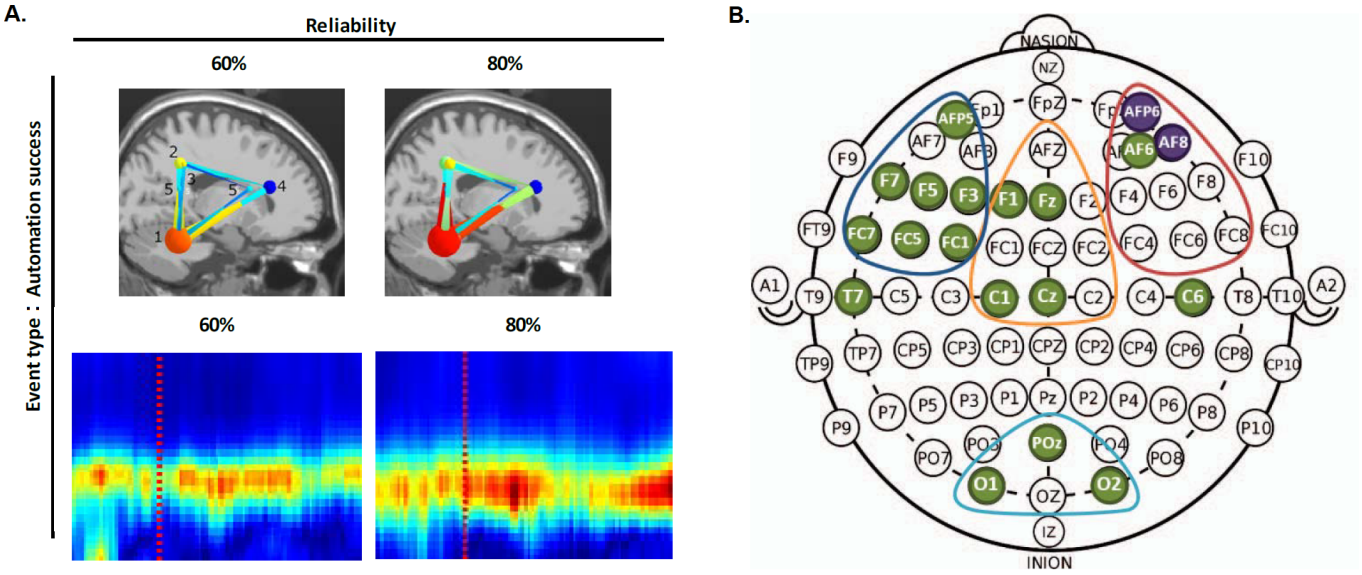

Figure 1: (**A**) EEG-based effective connectivity findings taken from one participant. There was strong flow of time-frequency information (as indicated by the bright red color) from the source node of fusiform gyrus (FG) [source node] to the dorsal anterior cingulate cortex (dACC) and posterior cingulate cortex (dPCC) [terminal nodes] when the automation was observed to be working successfully under the high reliability condition. Lighter colors in the low reliability condition indicate relatively weaker flow of time-frequency information. These findings were time-locked to the occurrence of the observational error positivity (oPE, see Table 1, row no. 4) 200 milliseconds after the onset of the automated event. [Source: Fig. 3, Sanders et al. (2019) [3]] (**B**) Map of EEG electrodes that recorded significant human-automation trust-related signals from the frontal and occipital cortices. [Source: Fig. 5, Wang et al., 2018 [29]]

Recent fMRI Studies

Over the past five years, there has only been three notable studies on human-automation trust [30], [31], [32]. Two investigated brain activations and effective connectivity between regions of interest through the use of a X-ray luggage screening task that presented automated advisories with varying proportion of false alarms [30] and misses [31]. [Note that a false alarm refer an emergency alarm (or alert) that is activated under an situation that does not require it [30], [7] while a miss refer to the absence of an alarm under an emergency situation that truly requires it [31].] The remaining and more recent study was a proof-of-concept study that investigated human-automation trust in the form of a quantum-inspired mathematical model [32] – to date, it was the only published trust-related study in the ATM domain. Table II lists the experimental tasks and design, data analysis techniques, main findings and implications tied to these three studies. From a quick overview of the brain-related findings, several facts are worth highlighting: (i) Even though the behavioral response associated with advice utilization can be the same for both human and machine agents, the brain activation patterns in response to the advice given by these agents can be different [31]. (ii) Greater neural resources, as shown through brain activations observed from contrast analyses, during decision making in response to advice given by a human agent than to advice given by an automated agent – and these activation patterns differ based on the proportion of false alarms (see Figs. 2 A and B) and misses (see Figs. 2 C and D) experienced by the user [30], [31]. (iii) Human-to-machine-agent contrasts done by Goodyear et al. (2016, 2017) [30], [31] showed activations in the ACC and insula; the same findings were uncovered by Pushparaj et al. (2019) [32], who postulated ACC activation to reflect reciprocal trust and insular activation to reflect pure distrust. (iv) ACC activation has also been shown in human-human (interpersonal) trust under using an voluntary trust game that require mutual trust between two participants in deciding the amount of financial payoffs [33]. This shows that the ACC is vital for representing or inferring the mental states of another agent, irrespective of whether it is a human or an automated tool.

With these facts in mind, it is worth mentioning that there is current movement toward the treatment of human-automation trust as a decision-making process [2]. Specifically, this process frames trust in the form of *value-based decisions* that required the human user to continuously weigh the expected personal value of letting the automation to complete the task versus performing it manually [2], [34], [35]. In this way, any trust-related assessment must be made against a backdrop of risk and/or reward that may come with human-automation cooperation. Through a critical review of the decision neuroscience literature, Drnec et al. (2016) [2] identified an assortment of brain regions, located in the prefrontal, parietal, and limbic regions that can be involved in trust-related value-based decisions. As shown in Table III, these hypothesized value-based regions involve the amygdala, ventral striatum, and ventral medial prefrontal cortex (vmPFC). Currently, to our knowledge, it is unknown as to: (i)

how many of these regions will be activated during human-automation trust-related decision making, (ii) the extent to which they are activated and the scale of connectivity between them, and (iii) how these brain activation and connectivity patterns, if present, change over time and under different physical or social environments. These issues show that there are many unknowns concerning the neural correlates of human-automation trust and that further investigations are definitely needed.

III. Study Rationale and Motivation

The preceding review of the recent neuroscience literature on human-automation trust showed that both EEG and fMRI showed some converging evidence with respect some common regions of activation, particularly those located in the frontal lobe (e.g., PFC and ACC). What is missing, however, is that there is currently no published EEG study that investigated human-automation trust in the ATM context and that there is only one fMRI study that examined this form of trust via a conflict detection task [32]. Since what is currently known about trust-related brain mechanisms in ATM is shown through this standalone fMRI study [32], this proposed study sets forth to extend the scope of the existing fMRI findings by adopting the fMRI task protocol used previously by Goodyear and colleagues [30], [31] and utlizing a compound event-related paradigm used previously by Zhong (2019) [36] (see section below, for details). Importantly, this study follows an operational definition of trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." [10] (as stated in "Introduction"), which characterizes trust as a decision-making heuristic that users of automated tool use in dynamic situations that involve risks and uncertainties [3], [2]. This definition was choosen becuase this study aims to follow up on Pushparaj et al.'s (2019) [32] by using an improved conflict detection task that features STCAs with two different levels of reliabilities would create air traffic control (ATC) scenarios with disparate levels of trust or uncertainties during system use.

STCA use serves as good context for examining trust because STCA is not 100% foolproof or reliable and can set off false alarms [7], [37], particularly in a termainal maneuvering area (TMA) with dense air traffic [38]. In the general use of automated tools, it is worth noting that false alarms can be more damaging to overall operational performance than misses [39]. In the ATM domain, false alarms are formally called "false alerts" and refers to any alert that does not correspond to a situation requiring particular attention or action (e.g. an alert caused by the display of split tracks on radar) [40]. Owing to the presence of STCA false alerts, trust in system reliability among ATCOs has always been cited as a primary concern [7], [11], [12], and and there is a current drive directed at creating enhanced STCA systems that issue lower rates of false alerts through enhanced algorithms (e.g., multiple tracking hypotheses tracking) – as specified in the European ATM Master Plan 2020 [38].

Crucially, this study aims to investigate ATCOs' trust associated with STCA use through fMRI – as done recently by Pushparaj et al. (2019) [32] – because fMRI is non-invasive, easy to administer, and is generally safe for most people. Notably, fMRI offers good spatial resolution of the brain to an accuracy of up to 1.0 mm (which cannot be offered by EEG) and temporal resolution of up to 1.0 second [41]. It renders structural or anatomical images of the brain, as well its functional images, which represent hemodynamic responses to decisions and/or responses made over the period of the brain scan [41]. Based on statistical analyses that commonly operate on the basis of General Linear Models (GLMs), functional brain images are color-coded to convey the regions of significant brain activity [42]. These color-coded regions convey "blood-oxygen-level-dependent (BOLD)" activity, which represents a coupling of cerebral blood flow into a brain region and the neuronal activation in that region [41]. In simple terms, this means that when an area of brain is "activated" by some mental processes, the flow of oxygenated blood into it increases. With fMRI, BOLD activations can be detected in both cortical and subcortical regions and this shall render a fuller picture of the brain regions (compared with the use of EEG) involved in making trust-related decisions during STCA use.

IV. Methods

Research Questions: The review of the recent neuroscience literature on human-automation trust showed that this form trust has been largely investigated in the context of automated tools that are not 100% reliable or trustworthy [26], [1], [27], [3], [28], [29], [30], [31]. The investigation of trust in this fashion presupposes uncertainty in automated tool use and this study follows this notion in the proposed fMRI task design. Specifically, this study aims to administer an improved version of the conflict detection task used by Pushparaj et al. (2019) [32] by incorporating the design principles used previously by Goodyear and colleagues [30], [31]. Like what Goodyear and colleagues did, this study proposes the administration of an automated tool, a STCA system in this study, with a low reliability of 60%. In addition, it is also proposed that a perfectly reliable STCA system that can elicit 100% trust in system use be administered. To our knowledge, there has not been any previous neurimaging study that investigated the brain activations associated with 100% trust in an automated or AI tool/system and this study aims to investigate such activations and compare them with activations associated with mistrust in the same type of tool/system. Following this line of thought, this study aims to address the following research questions:

1. Which brain regions will be activated when ATCOs use a STCA system with two vastly different levels of reliabilities?

2. How do the brain activations observed from the use of these two types of STCAs compare and contrast against each other (i) when observing an air traffic scenario before the onset of a conflict or STCA advice?, (ii) when making decisions after the onset of STCA advice?, and (iii) when reflecting on the decision concerning the acceptance or rejection of STCA advice?

3. To what extent are the significant brain regions of activations involved in trust-related decision-making correlated with each other?

Answering these questions will contribute to the construction of a neuroimaging database that informs neuroscience researchers using fMRI of the patterns of brain activations and connectivity of ATCOs with varying levels of trust for STCA use. By collecting brain-based evidence related to human-automation trust in the ATM context, it is also hoped that such evidence can be utlized as a useful source of reference for the future development of Machine Learning algorithms that can identify trusted automation or AI use based on well-known brain activation and connectivity patterns [25], [43].

Targeted Sample Size: Using a fully within-subjects design, this study aims to recruit a minimum of 10 professional ATCOs.

Conflict Detection Task for fMRI: The design of the fMRI-based conflict detection task (CDT) involving a decision-making event in each trial follows the task design of Goodyear and colleagues [30], [31] while the generation and presentation of trial types (2) for each STCA follows a *compound event-related paradigm* that was used previously in a navigational decision-making study by Zhong (2019) [36] (see subsection below, for details). Figure 3 shows the decision matrices associated with two STCA systems with 60% and 100% reliability, respectively. Note that for the STCA with 60% reliability, 40% of the ATC scenarios with conflicts present will feature false alerts. Figure 4 shows the two types of trials – compound and partial – tied to the presentation of each type of STCA. Both trial types feature an initial observational phase in which air traffic flow is observed. The compound trial differed from the partial trial with the requirement to make decisions to the advice provided by the STCA within a period that is expected to last for an average of 4.0 seconds. Feedback about the correctness of the reponse (i.e., whether it is right or wrong to accept or reject the STCA advice) is provided within a 2.0 second period, following by double fixation crosses before the start of a new trial. Temporal jittering of the delays between events (i.e., between decision-making and feedback) and trials was important to reduce expectancy and habituation effects that can lower BOLD signal quality [44].

Having a feedback phase is important because it will provide insights into the levels of calibrated trust among ATCO participants. Calibrated trust refers to the condition in which the anticipated or perceived trust of a machine, automation or AI matches the experienced trust associated with the actual use of the tool [10], [1], [3]. Seen in this light, trust can be studied as an evolving process that gets updated with increased information of the operating machine or system [1]. The feedback provided to ATCO participants shall allow them to adjust their self-perceived trust and fine-tune their performance based on updated knowedge of STCA system capability.

EUROCONTROL's Escape Light Simulator [45] will be used to design the ATC scenarios. These ATC scenarios will feature two different numbers of aircraft that will be presented repetitively in different starting locations across the trials. This method of presenting varying numbers of stimuli had been used previously by Goodyear and colleagues [30], [31] and serves the purpose of preventing any habituation effects that can significantly reduced the intensitites of the collected BOLD signals. Each ATC scenario will feature a lateral conflict between a pair of aircraft. There will be eight types of conflict events/geometries and each of these events/geometries will be repeated three times to constitute 24 compound trials. The partial trials will only feature obervations of these eight trials, without the need to make decisions or conduct any responses. The subection below provide details about the numbering and ordering of these two types of trials.

With respect to STCA design, an auditory tone will signal the onset of its activation (Figure 4). At the same time when the tone sounded, the pair of conflicting aircraft will be circled in a brightly color and the participant will have to decide whether or not to accept the advice. Due to the small of view that makes the reading of text information difficult in the fMRI scanner, the conflicts designed will all occur on the same flight level and the STCA will provide alerts with respect to these lateral conflicts only. To ensure that the STCA advice presentation is kept as realistic as possible, EUROCONTROL's guidelines for STCA operations and managment [40] will be followed when designing the STCA.

Compound Event-Related Paradigm: The compound event-related paradigm (also called the "partial-trial design" [46]) is an fMRI protocol that enables the separation of trials incurring two or more distinct processes (e.g., sensory, cognitive, motor) [i.e., compound trials] from trials incurring an initial subset of such processes (i.e., partial trials) [47], [48], [46]. In the context of the proposed experiment, compound trials pertain to the ATC trials featuring the full set of events encompassing air traffic flow observation, decision-making in the presence of STCA advice, and feedback (Figure 4) whereas partial trials pertain to the trials that involved the observation event only. The delineation of these two trial types shall enable an assessment of the brain regions involved in air traffic flow observation and STCA-based decision-making under different conditions of trust.

The time courses of BOLD responses to the compound and partial trials will be estimated using a General Linear Model (GLM) that made no a priori assumption about the shape of the hemodynamic response function (HRF) [47], [48], [49], [50]. This GLM shall involve: (i) jittering of the inter-event temporal intervals, ranging from 1 to 3 TRs (repetition time), in an exponential fashion to ensure a high level of accuracy in estimating the shape parameters of the hemodynamic response

function (HRF) through the sampling of more points on the HRF compared with using a fixed interval design [51], and (ii) inter-mixing of partial and compound trials in a 3:1 ratio to prevent participants from predicting the onset of a particular trial type and to ensure that the time courses of BOLD responses to different cognitive processes could be differentiated from each other [48], [49], [50].

Procedure: The study is proposed to involve a pre-fMRI practice phase and a formal fMRI scanning phase interleaved with subjective trust assessment (see Figure 5). In the initial practice phase, the participants must perform a designated number of practice trials (e.g., 12) featuring compound and partial trials arranged in a 3:1 ratio. Upon completion and before entering the fMRI scanner, they would be told that the advice provided by STCA would not always be 100% reliable and that they must exercise their judgment during advice acceptance or rejection. When making a decision to accept or reject the advice provided by the STCA, they would press one of two buttons on button box or keypad. In the fMRI scanning phase, they participants would perform two runs featuring the same number of compound and partial trials (e.g., 32 trials - 24 compound, 8 partial) but with different levels of STCA reliability. Each run would feature trials with STCA pre-set at a fixed reliability level (either 60% or 100%). These two types of runs will be presented in a counter-balanced order across participants. After each fMRI run, the experimenter will ask each participant a question related to his/her level of trust in the STCA over the microphone: "Overall, to what extent do you trust the STCA's ability to correctly identify the conflict?" on a scale ranging from 0 (not at all) to (completely). [Note: Alternatively, this question can be presented on a powerpoint slide and shown to participants.] This scale is adapted from one used recently by de Visser et al. (2018) [1]. Having a subjective report of trust is important because it will provide a quick assessment of the extent to which a participant is aware of the STCA's functionality experienced in each run. This subjective measure can also be useful for conducting correlational analysis with relevant BOLD parameters estimates derived from GLM analysis.

Table II: Summary of Key fMRI Studies of Human-Automation Trust done over the past Five Years (2016-2019)

Authors (Year of Publication)	Experimental Task(s) and Design	Neuroimaging Analysis Technique(s)	Trust-relevant Findings and Implications
Pushparaj, Ayeni, … Duong (2019)	Air Traffic Conflict Detection Task. Participants viewed five conflict scenarios and decided whether to accept or reject the advice given by an automated advisory tool that presents prompt advisory (text messages). A quantum-inspired mathematical model was used to define trust. Trust was defined as pure trust, pure distrust, reciprocal trust, and reciprocal distrust. Pure trust and pure distrust reciprocal were conceptualized as independent constructs while reciprocal trust and distrust were conceptualized as co-dependent constructs. Based on the model conceived, the aforementioned states of trust can be superimposed in accordance to the principle of superposition.	Seed-based correlation analysis (SCA). SCA explores the functional connectivity brain regions based on correlating the time series of the relevant seed voxels of activation.	When using insula as the seed, insular activation was found to correlate significantly with activations in the insula, anterior cingulate cortex (ACC), and putamen during the third conflict scenario, suggesting a simultaneous activation of pure distrust, reciprocal trust/distrust, and pure trust, respectively. When using ACC as the seed, ACC activation was found to correlate with the activations in the insula, and putamen during the second prompt and the fifth conflict scenario, suggesting a simultaneous activation of reciprocal trust/distrust, pure distrust, and pure trust, respectively. **Implication:** The fMRI findings provided the first findings for validating the quantum-inspired trust model. The main findings of simultaneous activations of two or more brain regions associated with the different states of trust suggested that mental representations of different trust states can be superimposed upon each other when monitoring air traffic and prior to deciding on the acceptance of an automated advisory.
Goodyear, Parasuraman, … Krueger (2017)	Same X-ray screening task as used by Goodyear et al. (2016), see row below) Bad advice refers to "misses" (knife = present, advice = absent) rather than "false alarms (knife = present, advice = absent)." Good advice: Hits (knife = present, advice = present; 60% of trials) Bad advice: Misses (40% of trials) Same three-way mixed model factorial design as used by Goodyear et al. (2016).	Same analyses as used by Goodyear et al. (2016)	Behavioral findings were largely consistent with previous findings shown by Goodyear et al. (2016). Main effects of trust and advice utilization percentage were found. Participants gave lower trust ratings after the fMRI scan than before the fMRI scan. Advice utilization lower in the presence of bad advice than in the presence of good advice. Unlike previous findings by Goodyear et al. (2016), no main or interactive effects of agent were found. During the decision phase (to search or to clear the misses" (passage of the luggage), brain activations (human > machine agent) were found in: **left anterior precuneus (L. aPreC)** **right lingual gyrus (R. LG)** **left cuneus (L. CUN)** During the feedback phase (slide showing whether a knife was present or absent), brain activations (machine > human agent) were found in: **right precentral gyrus (R. PrG)** **right inferior parietal lobule (R. IPL)** **left putamen (L. Pu)**

			Granger causality analysis (GCA) showed that during the decision phase, **R. LG** was the source of a brain network with output connections to target ROIs **R. anterior cingulate cortex (R. ACC)**, **L. PreC**, and **L. CUN**. During the feedback phase, **left fusiform gyrus (L. FG)** was found to send out output connections to R. IPL. **Implication:** Replicated previous findings by Goodyear et al. (2016) in showing that greater neural resources were consumed during decision-making in response to advice given by a human agent than to advice given by a machine agent. Effective connectivity findings derived from GCA suggested that the LG modulated visual attention in this decision-making process through a bottom-up processing of task-relevant information.
Goodyear, Parasuraman, … Krueger (2016)	X-ray screening task involving consideration and acceptance of advice from human and machine agents. Advice pertained to whether to search or to clear the passage of a luggage for the presence of a knife. Good advice: Hits (knife = present, 50% of trials) Bad advice: False alarms (knife = present, advice = absent; 40% of trials) Three-way mixed model factorial design. Between-subjects IV: 1. Advice (human vs. machine). Within-subjects IV: 1. Advice (good vs. bad), 2. Time (fMRI run1 vs. run2)	Between-subjects contrast analysis. Effective Connectivity Analysis via multivariate Granger causality analysis (GCA)	Self-reported trust in the human agent, not the machine agent, was lower post-scan when compared with pre-scan. Advice utilization percentage was lower in run 2 under the bad advice condition in the presence of the human agent compared with the machine agent. During the decision phase (to search or to clear the Parasuraman passage of the luggage), brain activations (human > machine agent) found in (Table 1, p. 9): **right posterior insula (R. PI)** **right anterior precuneus (R. aPreC)** **left aPreC** **left posterior cingulate cortex (L. PCC)** **left rostrolateral prefrontal cortex (L. rlPFC)** During the feedback phase (slide showing whether a knife was present or absent), brain activation (good advice > bad advice) in the **left dorsomedial prefrontal cortex (L. dmPFC)** was found in the presence human agent advice only: During the decision phase (run 1, bad advice), GCA showed that **L. aPreC** and **R. PI** were the sources of a brain network with outputs connections to target ROIs R. aPreC, PCC, and L. rlPFC. These effective connectivity path weights were stronger in the human-agent group compared to the machine-agent group. **Implication:** Greater neural resources were engaged when making decisions in response to bad advice given by a human than by a machine. The aPreC and PI were most probably involved in an integration of social evaluations (judgments about others' intentions and personal traits) with the recruitment of brain physiological responses.

Table III: Brain regions with putative roles in decision making. [Source: Table 1, Drnec et al. (2016)]

Neural substrate	Role in decision making	Reference
Amygdala	Processes/computes the value of negative stimuli	Yacubian et al. (2006) and Basten et al. (2010)
Ventral striatum	Processes/computes the value of positive stimuli	Yacubian et al. (2006), Basten et al. (2010) and Lim et al. (2011)
Ventral medial prefrontal cortex (vmPFC)	Calculates the difference of value signals from amygdala and ventral striatum in value based decisions	Basten et al. (2010) and Philiastides et al. (2011)
Dorsolateral prefrontal cortex (dlPFC)	Calculates the difference of signals from amygdala and ventral striatum in perceptual decisions	Basten et al. (2010) and Philiastides et al. (2011)
Lateral intraparietal cortex (LIP)	Accumulates and integrates the value of evidence processed by the vmPFC (evidence largely from monkeys)	Platt and Glimcher (1999); Platt (2002); Basten et al. (2010) and Rorie et al. (2010)
	A cortical area involved in gaze fixation, saccade, and attention, underlying evidence accumulation	Coe et al. (2002) and Goldberg et al. (2006)

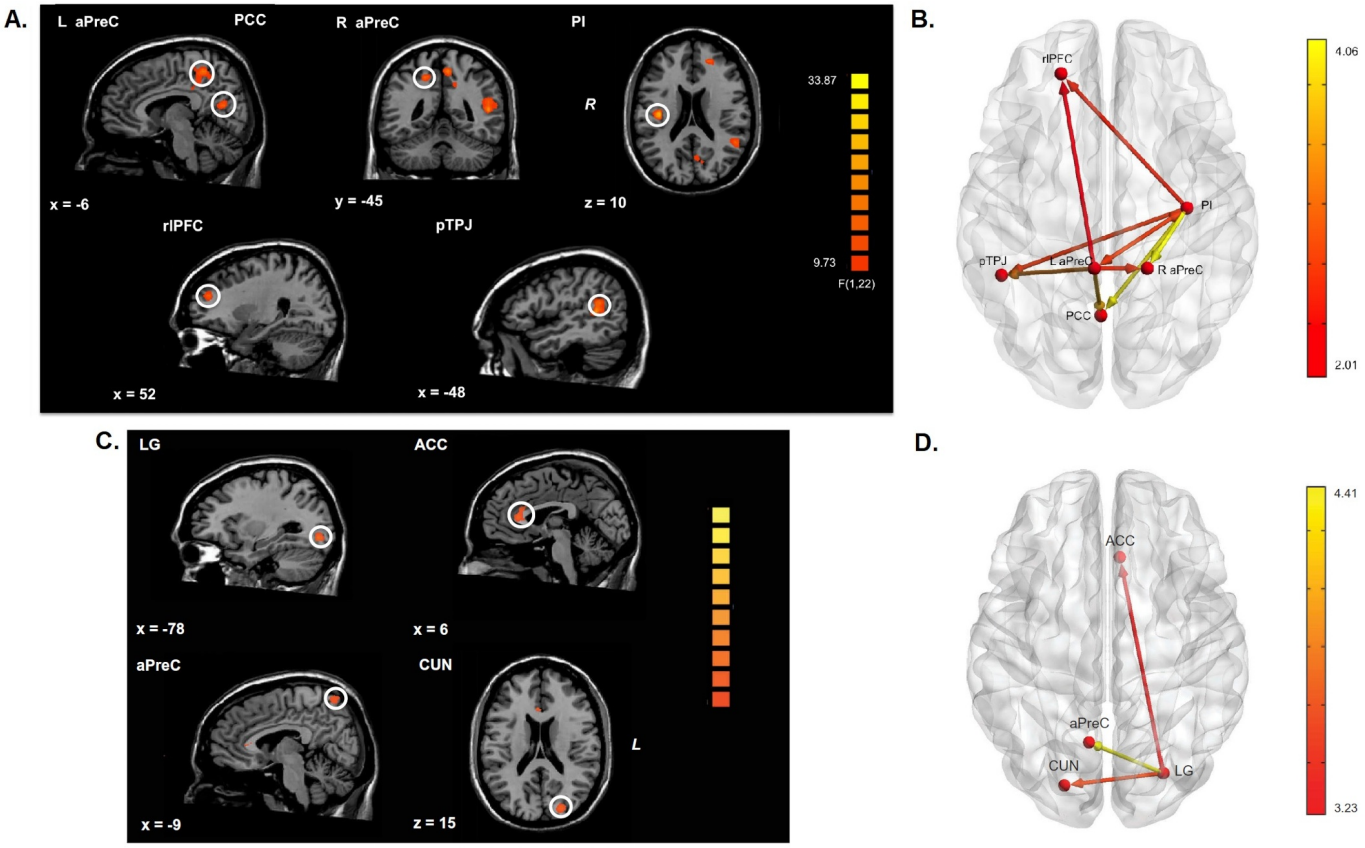

Figure 2: (**A**) Brain regions activated during the decision phase in a X-ray luggage scanning task with 40% false alarms based on human-to-machine agent contrast analysis. (**B**) Effective connectivity network during the decision phase in a X-ray luggage scanning task with 40% false alarms based on human-to-machine agent contrast analysis. [Source for panels A and B: Figs. 5 and 7, Goodyear et al. (2016) [30]]. (**C**) Brain regions activated during the decision phase in a X-ray luggage scanning task with 40% misses based on human-to-machine agent contrast analysis. (**D**) Effective connectivity network during the decision phase in a X-ray luggage scanning task with 40% misses based on human-to-machine agent contrast analysis. [Source for panels C and D: Figs. 3 and 5, Goodyear et al. (2017) [31]]

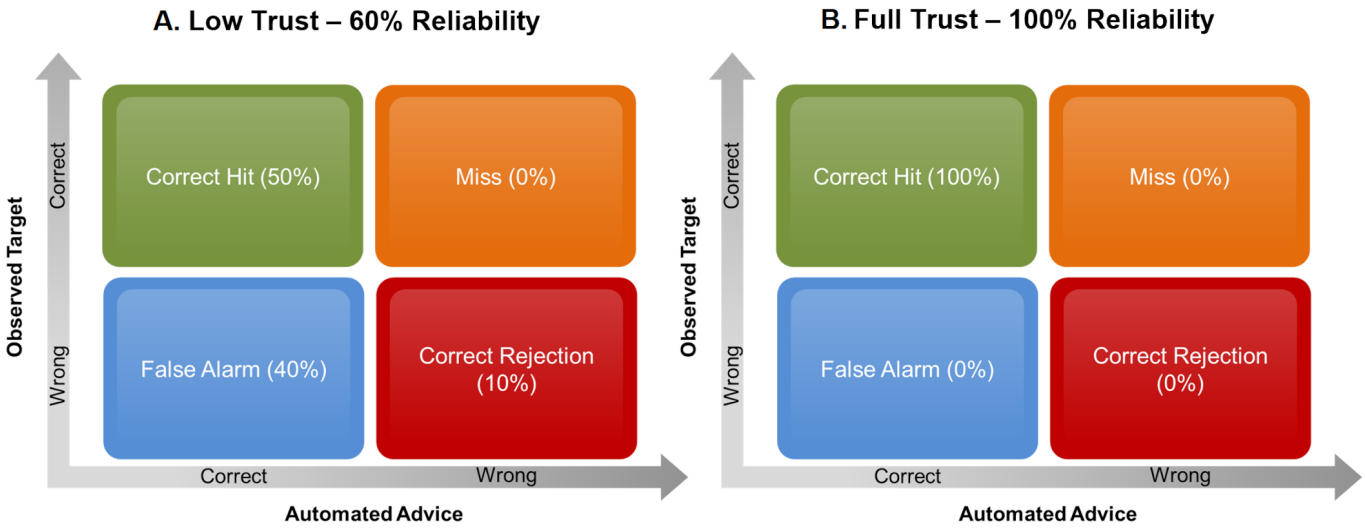

Figure 3: **A** Decision matrix for STCA system with 60% reliability. **B** Decision matrix for STCA system with 100% reliability.

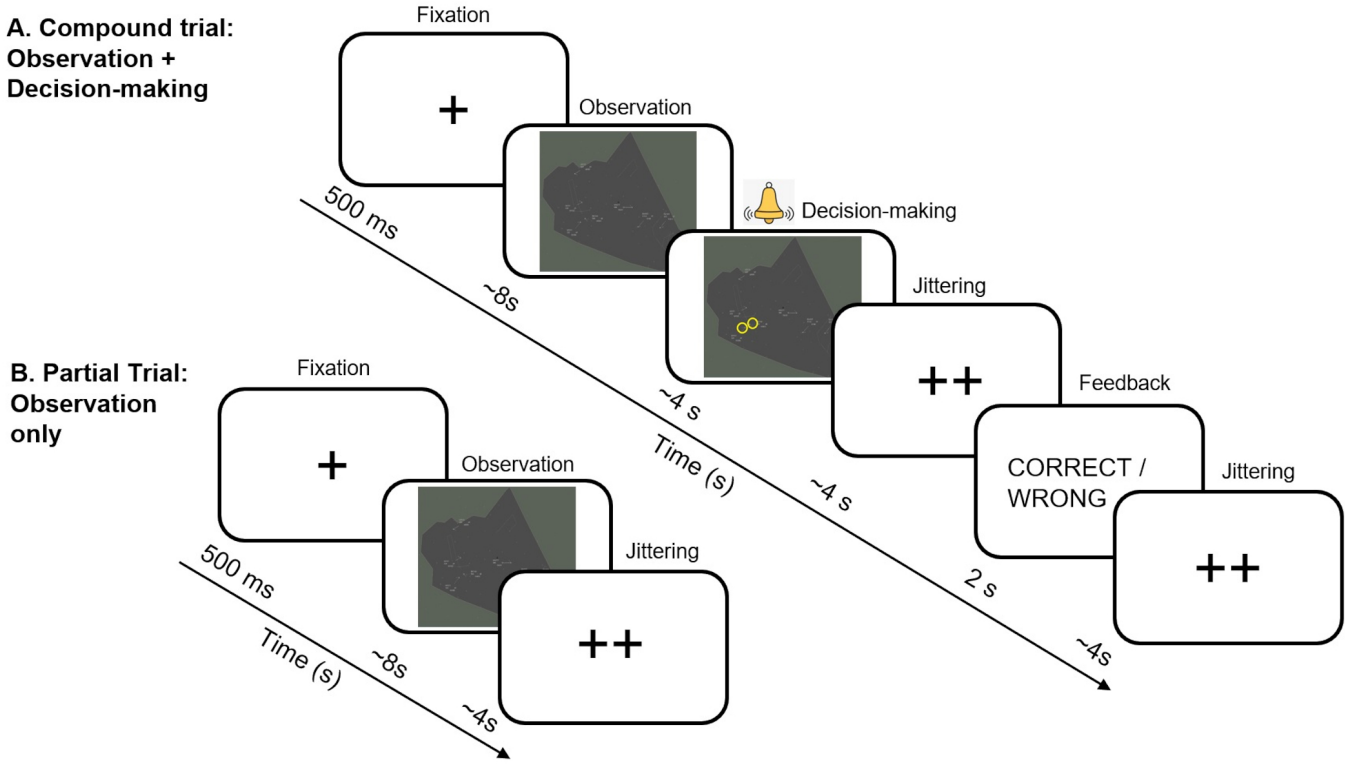

Figure 4: **A** Events in a compound trial that feature observation of air traffic flow, decision-making tied to acceptance or rejection of STCA advice, and feedback tied to the executed response. **B** A partial trial that features an observation of air traffic flow after the fixation period.

Figure 5: Flowchart diagram showing the linear ordering of experimental phases. Note that the same one-item trust survey will be used directly after each fMRI run.

REFERENCES

[1] E. J. de Visser, P. J. Beatty, J. R. Estepp, S. Kohn, A. Abubshait, J. R. Fedota, and C. G. McDonald, "Learning from the slips of others: Neural correlates of trust in automated agents," *Frontiers in Human Neuroscience*, vol. 12, no. 309, pp. 1–15, 2018, https://doi.org/10.3389/fnhum.2018.00309.

[2] K. Drnec, A. R. Marathe, J. R. Lukos, and J. S. Metcalfe, "From trust in automation to decision neuroscience: applying cognitive neuroscience methods to understand and improve interaction decisions involved in human automation interaction," *Frontiers in Human Neuroscience*, vol. 10, no. 290, pp. 1–14, 2016, https://doi.org/10.3389/fnhum.2016.00290.

[3] N. Sanders, S. Choo, N. Kim, C. S. Nam, and E. P. Fitts, "Neural correlates of trust during an automated system monitoring task: Preliminary results of an effective connectivity study," in *Proceedings of the Human Factors and Ergonomics Society*, 2019, pp. 1–5, https://doi.org/10.1177/1071181319631409.

[4] H. A. Abbass, J. Scholz, and D. J. Reid, Eds., *Foundations of trusted autonomy*. Springer, 2018, retrieved from https://link.springer.com/content/pdf/10.1007%2F978-3-319-64816-3.pdf.

[5] R. Parasumaran, T. B. Sheridan, and C. D. Wickens, "Situation awareness, mental workload,and trust in automation: Viable, empirically supported cognitive engineering constructs," *Journal of Cognitive Engineering and Decision Making*, vol. 2, no. 2, pp. 140–160, 2008, https://doi.org/10.1518/155534308X284417.

[6] K. E. Schaefer, J. Y. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human Factors*, vol. 58, no. 3, pp. 377–400, 2016, https://doi.org/10.1177/0018720816634228.

[7] C. Kelly, M. Boardman, P. Goillau, and E. Jeannot, "Guidelines for trust in future atm systems: A literature review," EUROCONTROL, Tech. Rep. HRS/HSP-005-GUI-01, 2003, retrieved from https://www.researchgate.net/publication/311065869_Guidelines_for_Trust_in_Future_ATM_Systems_A_Literature_Review.

[8] J. Langan-Fox, M. J. Sankey, and J. M. Canty, "Human factors measurement for future air traffic control systems," *Human Factors*, vol. 51, no. 5, pp. 595–637, 2009, https://doi.org/10.1177/0018720809355278.

[9] M. Madsen and S. Gregor, "Measuring human-computer trust," in *11th Australasian Conference on Information Systems*, vol. 53. Australasian Association for Information Systems, 2000, pp. 6–8, retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.3874&rep=rep1&type=pdf.

[10] J. D. Lee and K. A. See, "Trust in automation and technology: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004, https://doi.org/10.1518/hfes.46.1.50_30392.

[11] SKYbrary, "Short term conflict alert (STCA)," https://www.skybrary.aero/index.php/Short_Term_Conflict_Alert_(STCA), accessed: 2021-03-18.

[12] EUROCONTROL, "Safety nets: A guide for ensuring effectiveness," https://skybrary.aero/bookshelf/books/2761.pdf, 2017, accessed: 2021-03-18.

[13] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*. IEEE, 2000, pp. 286–297, retrieved from https://doi.org/10.1109/3468.844354.

[14] J. D. Lee and N. Moray, "Trust, control strategies, and allocation of function in human machine systems," *Ergonomics*, vol. 22, no. 10, pp. 1243–1270, 1992, https://doi.org/10.1080/00140139208967392.

[15] P. Goillau, C. Kelly, M. Boardman, and E. Jeannot, "Guidelines for trust in future atm systems: Measures," EUROCONTROL, Tech. Rep. HRS/HSP-005-GUI-02, 2003, retrieved from https://www.researchgate.net/publication/311068551_Guidelines_for_Trust_in_Future_ATM_Systems_Measures.

[16] S. M. Merritt and D. R. Ilgen, "Not all trust is created equal: Dispositional and history-based trust in human–automation interactions," *Human Factors*, vol. 50, no. 2, pp. 194–210, 2008, https://doi.org/10.1518/001872008X288574.

[17] U. Metzger and R. Parasuraman, "The role of the air traffic controller in future air traffic management: An empirical study of active control vs. passive monitoring," *Human Factors*, vol. 43, no. 4, pp. 519–528, 2008, https://doi.org/10.1518/001872001775870421.

[18] D. Manzey, J. E. Bahner, and A. D. Hueper, "Misuse of automated aids in process control: Complacency, automation bias and possible training interventions," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 3, 2006, pp. 220–224, retrieved from https://doi.org/10.1177/154193120605000303.

[19] U. Metzger and R. Parasuraman, "Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload," *Human Factors*, vol. 41, no. 7, pp. 35–49, 2005, https://doi.org/10.1518/0018720053653802.

[20] C. Wickens, S. Dixon, J. Goh, and B. Hammer, "Pilot dependence on imperfect diagnostic automation in simulated uav flights: An attentional visual scanning analysis," ILLINOIS UNIV AT URBANA SAVOY, Tech. Rep. AHFD-05-02, 2005, retrieved from https://apps.dtic.mil/sti/citations/ADA446167.

[21] M. T. Dzindolet, H. P. Beck, and L. G. Pierce, "Adaptive automation: Building flexibility into human-machine systems," in *Understanding adaptability: A prerequisite for effective performance within complex environments*, C. S. Burke, L. G. Pierce, and E. Salas, Eds. Emerald Group, 2006, pp. 213–245.

[22] J. P. Bliss, "Alarm reaction patterns by pilots as a function of reaction modality," *The International Journal of Aviation Psychology*, vol. 7, no. 1, pp. 1–14, 1997, http://dx.doi.org/10.1207/s15327108ijap0701_1.

[23] V. Riley, "A theory of operator reliance on automation," in *Human performance in automated systems: Recent research and trends*, M. Mouloua and R. Parasuraman, Eds. Lawrence Erlbaum, 1994, pp. 8–14.

[24] I. B. Ajenaghughrure, S. D. C. Sousa, and D. Lamas, "Measuring trust with psychophysiological signals:a systematic mapping study of approaches used," *Multimodal Technologies and Interaction*, vol. 4, no. 63, pp. 1–29, 2020, https://doi.org/10.3390/mti4030063.

[25] P. Aricò, G. Borghini, G. Di Flumeri, S. Bonelli, A. Golfetti, I. Graziani, S. Pozzi, J. P. Imbert, G. Granger, R. Benhacene, and D. Schaefer, "Human factors and neurophysiological metrics in air traffic control: a critical review," in *IEEE Reviews in Biomedical Engineering*. IEEE, 2017, pp. 250–263, retrieved from https://hal-enac.archives-ouvertes.fr/hal-01511343.

[26] K. Akash, W. L. Hu, N. Jain, and T. Reid, "A classification model for sensing human trust in machines using EEG and GSR," *ACM Transactions on Interactive Intelligent Systems*, vol. 8, no. 4, pp. 1–20, 2018, https://doi.org/10.1145/3132743.

[27] S. Oh, Y. Seong, S. Yi, and S. Park, "Neurological measurement of human trust inautomation using electroencephalogram," *nternational Journal of Fuzzy Logic and Intelligent Systems*, vol. 20, no. 4, pp. 261–271, 2020, http://doi.org/10.5391/IJFIS.2020.20.4.261.

[28] E.-S. Jung, S.-Y. Dong, and S.-Y. Lee, "Neural correlates of variations in human trust in human-like 3 machines during non-reciprocal interactions," *Scientific Reports*, vol. 9, no. 9975, pp. 1–10, 2019, https://doi.org/10.1038/s41598-019-46098-8.

[29] M. Wang, A. Hussein, R. F. Rojas, K. Shafi, and H. A. Abbass, "EEG-based neural correlates of trust in human-autonomy interaction," in *2018 IEEE Symposium Series on Computational Intelligence*. IEEE, 2018, pp. 350–357, https://doi.org/10.1109/SSCI.2018.8628649.

[30] K. Goodyear, R. Parasuraman, S. Chernyak, P. Madhavan, G. Deshpande, and F. Krueger, "Advice taking from humans and machines: An fmri and effective connectivity study," *Frontiers in Human Neuroscience*, vol. 10, no. 542, pp. 1–15, 2016, https://doi.org/10.3389/fnhum.2016.00542.

[31] K. Goodyear, R. Parasuraman, S. Chernyak, E. de Visser, P. Madhavan, G. Deshpande, and F. Krueger, "An fmri and effective connectivity study investigating miss errors during advice utilization from human and machine agents," *Social Neuroscience*, vol. 12, no. 5, pp. 570–581, 2017, https://doi.org/10.1080/17470919.2016.1205131.

[32] K. Pushparaj, A. J. Ayeni, G. Ky, S. Alam, V. Vijayaragavan, B. Gulyás, and V. N. Duong, "A quantum-inspired model for human-automation trust in air traffic control derived from functional magnetic resonance imaging," in *9th SESAR Innovation Days*, 2019, pp. 1–8, retrieved from https://www.sesarju.eu/sites/default/files/documents/sid/2019/papers/SIDs_2019_paper_55.pdf.

[33] F. Krueger, K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, A. Heinecke, and J. Grafman, "Neural correlates of trust," *Proceedings of the National Academy of Sciences, USA*, vol. 104, no. 50, pp. 20 084–20 089, 2007, https://doi.org/10.1073/pnas.0710103104.

[34] A. Rangel, C. Camerer, and P. R. Montague, "A framework for studying the neurobiology of value-based decision making," *Nature Reviews Neuroscience*, vol. 9, pp. 545–556, 2008, https://doi.org/10.1038/nrn2357.

[35] J. D. Wallis, "Cross-species studies of orbitofrontal cortex and value-based decision-making," *Nature Neuroscience*, vol. 15, pp. 13–19, 2012, https://doi.org/10.1038/nn.2956.

[36] J. Y. Zhong, "The impact of two spatial strategies on entorhinal and hippocampal involvement in visual path integration," Doctoral dissertation, Georgia Insititute of Technology, Atlanta, GA, USA, 2019, retrieved from https://smartech.gatech.edu/handle/1853/61594.

[37] C. D. Wickens, S. Rice, D. Keller, S. Hutchins, J. Hughes, and K. Clayton, "False alerts in air traffic control conflict alerting system: Is there a "cry wolf" effect?" *Human Factors*, vol. 51, no. 4, pp. 441–462, 2009, https://doi.org/10.1177/0018720809344720.

[38] SESAR, "European atm master plan: Digitalizing Europe's aviation infrastructure," 2020. [Online]. Available: https://www.sesarju.eu/masterplan

[39] S. R. Dixon, W. C. D., and J. S. McCarley, "On the independence of compliance and reliance: Are automation false alarms worse than misses?" *Human Factors*, vol. 49, no. 4, pp. 564–672, 2007, https://doi.org/10.1518/001872007X215656.

[40] EUROCONTROL, "Guidelines for short term conflict alert (STCA)," 2017. [Online]. Available: https://www.eurocontrol.int/publication/eurocontrol-guidelines-short-term-conflict-alert-stca

[41] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging*, 3rd ed. Oxford University Press, 2018.

[42] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *Handbook of functional MRI data analysis*. Cambridge University Press, 2018.

[43] T. Radüntz, N. Fürstenau, T. Mühlhausen, and B. Meffert, "Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps," *Frontiers in Physiology*, vol. 11, no. 300, pp. 1–13, 2020, https://doi.org/10.3389/fphys.2020.00300.

[44] F. Richlan, B. Gagl, S. Hawelka, M. Braun, M. Schurz, M. Kronbichler, and F. Hutzler, "Fixation-related fmri analysis in the domain of reading research: using self-paced eye movements as markers for hemodynamic brain responses during visual letter string processing," *Cerebral Cortex*, vol. 24, no. 10, pp. 2647–2656, 2014, https://doi.org/10.1093/cercor/bht117.

[45] EUROCONTROL, "EUROCONTROL simulation capabilities and platform for experimentation," 2021. [Online]. Available: https://www.eurocontrol.int/simulator/escape

[46] H. Ruge, T. Goschke, and T. S. Braver, "Separating event-related bold components within trials: The partial-trial design revisited," *NeuroImage*, vol. 47, no. 2, pp. 501–513, 2009, https://doi.org/10.1016/j.neuroimage.2009.04.075.

[47] J. M. Ollinger, G. L. Shulman, and M. Corbetta, "Separating processes within a trial in event-related functional mri: I. the method," *NeuroImage*, vol. 13, no. 1, pp. 210–217, 2001, https://doi.org/10.1006/nimg.2000.0710.

[48] ——, "Separating processes within a trial in event-related functional mri: Ii. analysis," *NeuroImage*, vol. 13, no. 1, pp. 219–229, 2001, https://doi.org/10.1006/nimg.2000.0711.

[49] G. L. Shulman, J. M. Ollinger, E. Akbudak, T. E. Conturo, A. Z. Snyder, S. E. Petersen, and M. Corbetta, "Areas involved in encoding and applying directional expectations to moving objects," *The Journal of Neuroscience*, vol. 19, no. 21, pp. 9480–9496, 1999, https://doi.org/10.1523/JNEUROSCI.19-21-09480.1999.

[50] M. E. Wheeler, G. L. Shulman, R. L. Buckner, F. M. Miezin, K. Velanova, and S. E. Petersen, "Evidence for separate perceptual reactivation and search processes during remembering," *Cerebral Cortex*, vol. 16, no. 7, pp. 949–957, 2006, https://doi.org/10.1093/cercor/bhj037.

[51] A. M. Dale and R. L. Buckner, "Selective averaging of rapidly presented individual trials using fmri," *Human Brain Mapping*, vol. 5, no. 5, pp. 329–340, 1997, https://pubmed.ncbi.nlm.nih.gov/20408237/.