Autopilot, Mind Wandering, and the Out of the Loop Performance Problem
Jonas Gouraud, Arnaud Delorme, Bruno Berberian

To cite this version:


HAL Id: hal-01729156
https://hal.archives-ouvertes.fr/hal-01729156
Submitted on 12 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Autopilot, Mind Wandering, and the Out of the Loop Performance Problem

Jonas Gouraud1*, Arnaud Delorme2 and Bruno Berberian1

1 Systems Control and Flight Dynamics Department, Office National d’Études et de Recherches Aérospatiales, Salon de Provence, France, 2 Center of Research on Brain and Cognition (UMR 5549), Centre National de la Recherche Scientifique, Toulouse, France

To satisfy the increasing demand for safer critical systems, engineers have integrated higher levels of automation, such as glass cockpits in aircraft, power plants, and driverless cars. These guiding principles relegate the operator to a monitoring role, increasing risks for humans to lack system understanding. The out of the loop performance problem arises when operators suffer from complacency and vigilance decrement; consequently, when automation does not behave as expected, understanding the system or taking back manual control may be difficult. Close to the out of the loop problem, mind wandering points to the propensity of the human mind to think about matters unrelated to the task at hand. This article reviews the literature related to both mind wandering and the out of the loop performance problem as it relates to task automation. We highlight studies showing how these phenomena interact with each other while impacting human performance within highly automated systems. We analyze how this proximity is supported by effects observed in automated environment, such as decoupling, sensory attention, and cognitive comprehension decrease. We also show that this link could be useful for detecting out of the loop situations through mind wandering markers. Finally, we examine the limitations of the current knowledge because many questions remain open to characterize interactions between out of the loop, mind wandering, and automation.

Keywords: out of the loop, mind wandering, automation, performance, vigilance, psychophysiological measures, complacency

INTRODUCTION

To continuously improve system safety, the critical systems industry makes extensive use of automation (Parasuraman, 1987; Billings, 1991; Sheridan, 1992; Degani and Heymann, 2000; Baxter et al., 2012). Automation has been introduced to answer performance and safety requirements in aircraft cockpits (Wise et al., 1994), in cars (Naujoks et al., 2016), and in power plant consoles (Cummings et al., 2010). Since the 1980s, designers integrated multiple modes of automation, allowing pilots to fly in autopilot mode. The automated mode is now able to maintain an altitude, fly to a point, or perform a landing, all without any human intervention (Wiener, 1988). Cars are currently going through the same revolution, as level 2 automation is being deployed—autopilots manage the car’s trajectory while human supervision is still needed. At the same time, the industry is conducting studies of level 3 automation—no human intervention or supervision required...
Unfortunately, if implementing higher levels of automation can improve the efficiency and capacity of a system, it also introduces difficulties for human operators. It is now well-accepted that traditional automation has several negative consequences for performance and safety, a set of difficulties which are called out of the loop (OOTL) performance problem. The OOTL phenomenon corresponds to a deterioration of the operator’s attention when interacting with highly automated system. The terms “total confusion” (Bureau d’Enquête et d’Analyse, 2002, p. 167; National Transport Safety Board, 1975, p. 17), “surprise effect” (Bureau d’Enquête et d’Analyse, 2012a, p. 10, Bureau d’Enquête et d’Analyse, 2016, p. 44) or “no awareness of the current mode of the system” (Bureau d’Enquête et d’Analyse, 2012b, p. 178) indicate a similar process—a mental state where the operator has lost his or her situation awareness and is not able to monitor the system efficiently. OOTL, which constitutes a human-machine miscommunication, has been pointed out as a cause of many accidents of various scales (Billings, 1991; Endsley and Kiris, 1995; Molloy and Parsaraman, 1996). Human-machine miscommunication describe situations where an operator or a machine “obtains an interpretation that she believes is complete and correct, but which is, however, not the one that the other speaker intended her to obtain” (McRoy, 2017). Miscomprehension can create accidents or drive operators to reject automation. For example, power plant operators declared to Andersson (2008) that they generally avoided higher automation level use because they “don’t know what it is doing.” When the Federal Aviation Administration of the United States investigated the accident of the Eastern Airlines L-1011, which crashed during clear weather and with no apparent causes, the investigation concluded that the crew was focused on a red light in the cockpit and didn’t notice that the autopilot had disengaged and that the plane started slowly going down (Federal Aviation Authority, 1972). At an operational level, the OOTL performance problem induces a performance decrease when trying to transfer manual control over the system (Berberian et al., 2012). Amongst other problems, an operator that is OOTL might take longer or be completely unable to detect an automation failure, decide if an intervention is needed, and find the adequate course of action. In the current context of the continued increase in automation, it is crucial to understand the sources of human–system interaction difficulties.

Although the OOTL performance problem represents a key challenge for system designers, it remains difficult to characterize and quantify after decades of research (Bainbridge, 1983; Baxter et al., 2012). Some researchers have pointed out vigilance failure as a key component of OOTL situations (Sarter and Woods, 1995b; Amalberti, 1999). Reports of incidents in aviation have notably illustrated the role of vigilance failure in human error. For example, Mosier et al. (1994) examined NASA’s Aviation Safety Reporting System (ASRS) database and found that 77% of the incidents in which over-reliance on automation was suspected involved a probable vigilance failure. Similarly, Gerbert and Kemmler (1986) studied German aviators’ anonymous responses to questionnaires about automation-related incidents and reported failures of vigilance as the largest contributor to human error. Nowadays, there is some consensus for the existence of a degradation of human operator vigilance in interaction with highly automated system (see, for example, O’Hanlon, 1981; Wiener, 1987; Strauch, 2002).

In this review, we aim to improve our comprehension of the OOTL performance problem and the related vigilance failure. In particular, we aim to explore the relation between the vigilance failures as observed in OOTL and the mind wandering (MW) phenomenon. MW is the human mind’s propensity to generate thoughts unrelated to the task at hand (Christoff, 2012; Stawarczyk et al., 2012). MW is a fuzzy concept referring to the human mind’s propensity to experience a variety of thoughts, which can be categorized along several dimensions. We will here use the term “mind wandering” to point out guided/unguided, internally/externally generated and spontaneous/intended thoughts unrelated to the task at hand. Regardless the exact properties of these thoughts, the MW phenomenon diverts attention from immediate goals while the subject can be aware of it or not (Golchert et al., 2016; Seli et al., 2016). An individual who is MW is at least partly decoupled from his or her environment and show little to no reaction to external stimuli (Schooler et al., 2014). In brain imaging studies, MW is characterized by the activation of the Default Mode Network, a widely distributed brain region comprised of medial prefrontal cortex and the posterior cingulate cortex (Mason et al., 2007; Christoff et al., 2009; Christoff, 2012; Konishi et al., 2015). Even though MW is thought to facilitate prospection, introspection and problem solving (Smallwood and Schooler, 2006), performance drops in numerous tasks has been observed during MW episodes (He et al., 2011; Galera et al., 2012; Schad et al., 2012; Bastian and Sackur, 2013; Casner and Schooler, 2013, 2015; Yanko and Spakek, 2014; Berthié et al., 2015). Several aspects outline a possible role of MW in OOTL in highly reliable automated environments. This paper reviews the literature related to both MW and OOTL performance problem as it relates to automation. We investigate the possibility of a link between MW and OOTL by reviewing how features of both phenomena bridge the two together. Far from being only theoretical, we highlight how such a link could help both MW and OOTL research in practice. Finally, we analyze perspectives to go further toward understanding and detecting both phenomena.

MIND WANDERING TO COMPLETE OOTL THEORIES

Multiple studies have showed that MW affects us all. The time we spend experiencing MW varies from 24 up to 60% depending on the study—40% in Schad et al. (2012); 47% in Killingsworth and Gilbert (2010); 24 and 31% in Bixler and D’Mello (2014); 30% in Kane et al. (2007); and 60% in Kam et al. (2011). This phenomenon has three major features: it is experienced by everybody (Killingsworth and Gilbert, 2010), it influences our behavior and attention toward external stimuli (He et al., 2011), and it can take place either intentionally or unintentionally (Smallwood and Schooler, 2006; Seli et al.,
2016). All of those aspects pose a safety risk for any critical task requiring sustained attention, such as supervising automated systems.

MW is sensitive to multiple task characteristics. MW appears when the subject performs monotonous tasks (Eastwood et al., 2012). Familiar stimuli have been shown to increase MW (Bastian et al., 2017). While easier or longer tasks were also associated with more frequent MW episodes (Thomson et al., 2014; Smallwood and Schooler, 2015), MW might actually help to cope with boredom (Schoolder et al., 2014). Boredom arises when people are unable to engage in satisfying activities while blaming their environment for it (Cummings et al., 2015). Several studies by Cheyne et al. (Cheyne et al., 2006; Carriere et al., 2008) point to the relationship between MW and boredom. Using questionnaires, they found a significant increase in everyday attentional failures for individuals more prone to boredom. Interestingly, Cummings et al. (2015) recently warned about a possible increase in boredom when integrating higher levels of automation. Moreover, MW related to automation was recently observed in automated systems. Casner and Schooler (2015) conducted a study where pilots were instructed to handle the approach—flight phase before landing—in a simulator by following beacons at altitudes given by the air traffic controller (ATC) officer. Probes inquired about their state of mind at predetermined times while the pilots had to report their position to the ATC officer. They observed that pilots were more prone to MW for higher levels of automation when they had no interaction with the system. Instead of planning the flight ahead, the pilots were inclined to think about unrelated matters. Although multiple studies have shown that monitoring is stressful and requires high levels of cognitive resources (Warm et al., 1996, 2008; Helton and Warm, 2008), vigilance theories do not explain such an increase in MW. On the contrary, could MW theories give a rational explanation in a monitoring environment?

**Complacency as a Possible Link between OOTL Vigilance Failure and MW**

Automation technology has changed the very nature of operators' work. Pilots are now required to monitor systems for possible failures. Monitoring tasks request a constant attention from the subject in order to detect seldom and unpredictable events over prolonged periods of time. This fundamental function is called the sustained attention (Manly et al., 1999). Interestingly, several studies show that efficient sustained attention over hours cannot be achieved (e.g., Methot and Huitema, 1998). If research on vigilance suggests that time on task significantly decreases our ability to discriminate infrequent and unpredictable signals (Mackworth, 1948; Teichner, 1974; Parasuraman, 1979; Warm, 1984), then vigilance failures also encompass another reality when dealing with automation—that is, the complacency experienced by operators dealing with highly reliable automated systems (Parasuraman et al., 1993a; Cummings, 2004).

Overreliance or complacency is created by an uncritical reliance on the system leading to thinking of it as more competent than it actually is (Lee and Sec, 2004; Bahner et al., 2008). Operators working with systems that fail once every 10 million hours of use tend to underestimate the possibility of automation errors and overtrust the system (Amalberti, 2001; Parasuraman and Wickens, 2008). Because they have the feeling that the system does not require them to work efficiently, they instinctively lower cognitive resources allocated to monitoring (Thackray and Touchstone, 1989; Morrison et al., 1993). The first empirical evidence was the study by Parasuraman et al. (1993a). They tested non-pilot participants on a flight simulation task made of 2D compensatory tracking, fuel management, and system monitoring. In the multiple-task condition, the participants performed the tracking and fuel management tasks manually while the automation handled the system monitoring. In the single-task condition, the participants only had to supervise the automation in the system monitoring task. In both conditions, automation reliability was variable. The participants were responsible for detecting these failures, and they had to take over when there was a failure. Parasuraman et al. (1993a) observed that participants had a detection rate of over 70% when performing the engine status task manually (a baseline condition). Their detection rate substantially declined when performing the task in the multitask condition. Interestingly, the effect was absent when they were in the single task condition, suggesting that the allocation of cognitive resources plays a role in the complacency effect (Moray and Inagaki, 2000; Bailey and Scerbo, 2007). Congruently, operators make fewer eye movements to the raw information sources when using automation than under manual control (Metzger and Parasuraman, 2001; Bagheri and Jamieson, 2004; Wickens et al., 2005), reflecting an allocation of attention to other concurrent tasks. Furthermore, operators tend to less frequently visualize parameters in automation mode than under manual mode, thus blindly trusting the automation diagnosis (Lorenz et al., 2002; Manzey et al., 2006). In a low probability signal context, Manly et al. (1999) used a sustained attention to response task (SART, a GO/NOGO task), to demonstrate a striking positive correlation between signal probability and detection rate.

These results indicate that complacency could be closely linked to MW, as both complacency and MW divert cognitive resources away from the task at hand. Supervising ultra-reliable systems seems to encourage a decrease in cognitive resources allocated to the monitoring task. In this context, resources saved by automation, which should normally be used to plan the flight, would instead be directed toward task-unrelated thoughts. Therefore, complacency might lead operators to free cognitive resources and reallocate them to unrelated thoughts. This assertion is supported by an observed increase in MW in a low probability signal environment (Galera et al., 2012; Berthié et al., 2015; Casner and Schooler, 2015) and as one has been on task for a longer period of time (Teasdale et al., 1995; Smallwood et al., 2003, 2006; McVay and Kane, 2009; Thomson et al., 2014). Nevertheless, the exact direction of this link remains to be assessed. MW could also occur prior to complacency and modify its emergence, for example by lowering the level of confidence needed for the operator to become complacent. Further data is needed to take position.
OOTL and MW: Some Similarities

Issues with Decoupling of Human Observer from the Task at Hand

When designers integrate automation in systems, they often believe that it will only be a substitute to the human operator (i.e., substitution myth, see Woods and Tinapple, 1999). However, an important part of the literature has accumulated evidence against this view. Automation does not only simply perform tasks that were previously handled by humans. It also changes the complexity of the task and creates new issues, thus transforming the nature of human work. Operators give up their direct control over the system for a monitoring role in the supervisory control loop (Moray, 1986; Sheridan, 1992). These changes are far from trivial—direct control involves manual functions including process planning, decision making, selecting responses, and implementing strategies. At the same time, passive information monitoring only requires information sources to be scanned and compared to previously learned references. In an automated environment, operators can experience loss of manual skills (Baxter et al., 2012), a decreased sense of control (Berberian et al., 2012), and a feeling of distance from the system (Bainbridge, 1983). This distance disturbs the operator’s involvement in the task. The same phenomenon of decoupling from the task is observed in MW. The operators’ attention during MW is shifted from the immediate task toward unrelated concerns (Schooler et al., 2011). In other words, although the impact of MW and OOTL on operators’ experience seems different, both influences start with a decoupling from the task. Moreover, both are equally threatening to safety in critical systems. For example, MW leads operators to forget to report as instructed (Casner and Schooler, 2015) and slows their adaptation to original tasks (Mooneyham and Schooler, 2013), whereas OOTL makes operators less responsive (Endsley and Kiris, 1995) and lowers their failure detection rate (Parasuraman and Riley, 1997).

Sensory Attenuation Problem

As defined by Endsley and Kiris (1995), OOTL is defined as the loss of one or more levels of situation awareness, which are perception (perceiving what is happening), comprehension (understanding the meaning of observed events), and projection (being able to think ahead). Given that perception guides both higher levels, its failure impacts the whole cognition. Several studies have shown a longer reaction time and lower detection rate following long automated periods. Endsley and Rodgers (1998) found that ATC officers showed poor performance in detecting conflicts within recorded traffic when they were passively monitoring the traffic. Willems and Truitt (1999) exposed that, in the same condition, ATC officers were slower to answer questions regarding traffic awareness and they recall less information as traffic load increases. In operational conditions, a lack of detection has led to tragic consequences. For example, the crash of the Mont-Saint-Odile (France) was due to a misunderstanding between the system and the pilots (Bureau d’Enquête et d’Analyse, 1992). During the landing procedure, the pilots selected the wrong units for the glide path, leading to a far steeper slope than expected. The cause was that the unit was not shown on the display but on the selection button. This accident demonstrates how operators can be impacted by OOTL and do not perform the usual checks on common procedures.

Similarly, MW involves a reduction in perceptual awareness of the task-relevant environment that lowers the subjects’ ability to detect signals (Merat and Jamson, 2008; He et al., 2011; Blanchard et al., 2014), particularly when dealing with automation (Thackray and Touchstone, 1989). O’Connell et al. (2009) used a Sustained Attention to Response Task to demonstrate that alpha waves were higher during MW episodes in occipital scalp sites. Tasks analyzing selective attention, where one has to inhibit attention to parts of the environment in order to efficiently perform a task, suggest the involvement of alpha activity as a sensory suppression mechanism (Foxe et al., 1998; Foxe and Snyder, 2011), or similarly as reflecting pulsed-inhibition of ongoing cortical processing (Mathewson, 2011). Recently, both electroencephalography (EEG) and magnetic resonance imagery (MRI) have found alpha wave increases in supposedly deactivated regions by manipulating both the level of internally directed attention and the level of self-generated thought (Benedek et al., 2014, 2016), thus supporting the idea of alpha waves being a marker of inhibition. Taken together, these findings rule out the possibility that these effects could rely on sensory (bottom-up) processing of the cue and they suggest an endogenous inhibitory effect (top-down). During this time, the system and environment may change, hence increasing risks to the operator of having an out of date model of the situation. Without a proper perception of feedback and system modes, humans can lack the understanding that is mandatory to operate.

A Human-Machine Interface Communication Problem

In addition to perception, cognitive comprehension may also be impacted by both phenomena. When automation fails or behaves abnormally, the operator is required to handle the difficulties alone. These cases have been well-documented in various domains, most notably flight deck and operating room automation (e.g., Sarter and Woods, 1995a,b; Degani and Heymann, 2000). Several fatal crashes and other incidents have been attributed to problems in the flight crew–automation interface (see for example Federal Aviation Authority, 1995). Sarter et al. (1997) referred to this as automation surprises, a point where the system behaves differently from what the operator expects. In laboratories, Wickens and Kessel (1977) demonstrated that operators removed from the system control show slower reactions and poor response accuracy. Carmody and Gluckman (1993) demonstrated that for complex task models, higher level of automation induced heavy losses of understanding. Taken together, these findings demonstrate that automation failures lead to a critical situation where the operator is OOTL and cannot initiate proper recovery actions.

Interestingly, similar understanding issues have been observed for MW. The subjects experience unconscious working memory transfer from the task at hand toward unrelated thoughts. Participants reading a text exhibited comprehension drops (Smallwood et al., 2008; Schad et al., 2012) and less reactions to text difficulties (Feng et al., 2013) during MW. Brain studies have shown activity uncorrelated to the environment during the same periods (Konishi et al., 2015). A decrement in external
stimuli processing is particularly true within monotonous and uninteresting environments (Mosier et al., 1994). In the operational context, studies point to MW as a possible cause of many driving accidents (Galera et al., 2012), plane crashes (Casner and Schooler, 2013), and medical errors (van Charante et al., 1993), maybe due to a lack of a proper model of the situation in critical moments. Smallwood et al. (2007, 2011) developed the cascading model of inattention in order to offer an explanation. They suggest that the superficial deficit in information processing induced by MW would cascade and impair a deeper level of understanding and negatively impact the construction of an accurate situation model. The poor-quality model would then decrease the ability of the environment to hold the operator's attention, which in turn would decrease the quality of the model, and so on. Therefore, MW episodes would progressively impair the operator's situation model and their capability to handle seldom events. This degraded context could favor OOTL apparition and reveal MW important impact in critical situations.

The Exact Nature of the Link between MW and OOTL

After comparing MW and OOTL on multiple aspects, a question arises: how can they be linked? Casner and Schooler (2013) highlighted the blurry situation of pilots left with spare time and no guidance about how to actively monitor the automation. This spare time could encourage the operators to think about unrelated concerns and this would drive them away from important matters, such as their current position or the mode of the system. Without knowledge of the situation, OOTL risk rises, and threatens operations.

We suggest that MW and OOTL could interact through working memory. When experiencing MW, task-unrelated thoughts flood working memory (McVay and Kane, 2010). Depending on the individual's working memory capacity, MW thoughts might fully occupy working memory capacity, preventing new resources from being allocated to the ongoing task. As the observed vigilance decrement will lower available working memory, full capacity may be reached even more quickly within highly automated environments. At the same time, complacency could drive the operator to lower the amount of working memory capacity allocated to the task. The working memory capacity freed by complacency would be promptly used for more unrelated thoughts. Our framework is supported by various results examining the relations between MW—working memory and OOTL—working memory. Examining the trial-by-trial co-occurrence of MW and performance declines during a working memory span task, Schooler et al. (2014) found that MW precedes poor performance. Our framework states that when filled with task-unrelated thoughts, working memory capacity cannot cope with new cognitive needs. Then operators experience a drop in performance. Similarly, maintaining a good situation awareness—closely linked to whether one is OOTL (Kaber et al., 2000)—requires working memory capacity through the active manipulation and use of information (Durso et al., 1999). When executive resources are used by MW, the individual will see her situation awareness decrease, leading to a higher risk of being OOTL.

Nevertheless, the link between MW and OOTL remains unclear. Characterizing its features could help to both better define OOTL and understand some of the situations that have led to tragic accidents. To achieve this goal, MW markers could help study OOTL situations. We highlight some possible directions for research in the following sections.

MW MARKERS TO STUDY OOTL

The Need for Online Measures of OOTL

One of the biggest difficulties associated with automation is its insidious effect on situation awareness (SA) and performance. Several solutions have been designed to avoid OOTL. Among them, adaptive automation proposes to dynamically change the level of automation according to the value of a parameter. Workload and vigilance levels have already been used as automation triggers, with convincing results on SA and overall performances (Parasuraman et al., 1996; Mikulka et al., 2002). A possibility would be to directly use markers of SA to adapt the level of automation and avoid OOTL situations. Salmon et al. (2009) identified different categories of SA assessment methods, including freeze probe recall techniques, real-time probe techniques, post-trial subjective rating techniques, observer rating techniques, process indices, and performance measures. However, they are poorly suited for online use in operational environments. Most of them are disruptive and either necessitate task freezing (Endsley, 1988), post-trial assessment (Taylor, 1990), reports by an observer (Matthews and Beal, 2002), or direct questions to the subject (Durso et al., 1999). For example, one of the most used measures, the Situation Awareness Global Assessment Technique (Endsley, 1988), requires the pilot to halt the simulation and blank all displays. The pilot is then asked a series of questions to determine his knowledge of the current situation to determine his SA. The QUASA is another widely used measure of SA. The operator has to answer regular true/false probes followed by rating scales about his own confidence. Although this measure does not freeze the simulation, it diverts the operator's attention toward matters unrelated to the task. Critical systems cannot tolerate this impact on performances in real situations. Recent developments support psychophysiological markers to negate vigilance decrement, particularly within adaptive automation (Prinz et al., 2003; Freeman et al., 2004). They create little intrusion for the subject, can record continuously (Eggemeier and Wilson, 1991; Kramer, 1991), and have already demonstrated a capacity to diagnostic on multiple levels; that is, arousal, attention and workload (Hancock and Williams, 1993; Harris et al., 1993; Parasuraman et al., 1993b; Boucein et al., 2007). To help achieve better detection, recent findings in MW literature could help track OOTL. Many psychophysiological markers have already been extensively used in MW studies, covering a wide range of detection tools—from brain imaging to heart-rate and sudation, including oculometry. Therefore, it is necessary to examine the possibilities of tracking OOTL situations using MW markers.
Self-Report Measures

MW markers are sorted using the triangulation classification among self-reports, physiological, and behavioral measures (Smallwood and Schooler, 2015). Self-reports regroup all of the subjective measures of MW. Most experiments use probes to determine periods when subjects are on-task or in MW (Smallwood et al., 2004; Gilbert et al., 2007; Braboszcz and Delorme, 2011; Uzzaman and Joordens, 2011; Feng et al., 2013). Although subjective reports have their limitation (Overgaard and Fazekas, 2016; Tsujiya et al., 2016), they remain widely used to define an interval as MW or focused. Whereas, it may prove difficult for someone to report their level of vigilance, MW reports have demonstrated a high correlation with neurophysiological measures (Smallwood et al., 2008; Cowley, 2013). This robustness could prove to be useful when studying OOTL situations in laboratories but it would not be useful in operational environments. Nevertheless, other markers have demonstrated promising results and could be used with satisfying detection rates in the near future.

Behavioral Measures

Behavioral markers of MW come in a wide variety. Within this category, reaction time measurements take an important place. Multiple studies highlighted the progressively faster reaction time during MW, linking it to impulsive behavior (Smallwood et al., 2003; 2004; 2011; Cheyne et al., 2011). This parameter allows us to track the subject’s attention without disturbing them. It contains much information, such as omissions—subject does not react to a stimulus although they were instructed to (see Bastian and Sackur, 2013)—and anticipations—reaction lower than 100 ms (see Hu et al., 2012). Cheyne et al. (2009) proved the robustness of the coefficient of variability—on a given interval, mean reaction time divided by its variability—to study MW in details (Bastian and Sackur, 2013; Esterman et al., 2013). Parallel to those results, subject accuracy is extensively used, whether it during trial to trial tasks (Braem et al., 2015; Durantin et al., 2015; Konishi et al., 2015) or during continuous monitoring, such as in a car simulator (He et al., 2011; Cowley, 2013; Yanko and Spalek, 2014).

On the whole, behavioral markers can highlight performance decrements induced by MW in many different tasks. They can also be used for OOTL characterization; for example, reaction time to take manual control over a system (de Winter et al., 2014) or accuracy to detect automation failures (Metzger and Parasuraman, 2001). Unfortunately, these measures are also of limited use outside the laboratory. Reaction time is useful when the participants have to perform actions regularly whereas OOTL is mainly a problem when supervising highly automated systems where actions are seldom required. Given that accuracy measures the participants’ shift from the goal, it is also limited to situations where the operator is already OOTL. Therefore, physiological measures could be useful to detect the dynamics of the problem.

Oculometric Measures

Oculometric measures allow us to derive different markers for potential use in detecting attentional lapses occurring during both MW and OOTL. Researchers demonstrated that, during visual tasks, pupil dilation occurs when subjects experience MW (Lowenstein and Loewenfeld, 1962; Yoss et al., 1970; Mittner et al., 2014). This behavior is correlated with norepinephrine activity in the locus coeruleus (i.e., the LCNE system) and is thought to be linked with the role of surprise (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010; Jepma and Nieuwenhuis, 2011). MW is also accompanied by changes in gaze position (Grandschamp et al., 2014), a change in eye movement pattern (Smiley et al., 2010; He et al., 2011), blink count (Uzzaman and Joordens, 2011), and saccades (Bixler and D’Mello, 2014). Reading tasks highlighted differences in on and off-text fixations (Reichle et al., 2010; Bixler and D’Mello, 2015), reading speed (Franklin et al., 2011; Feng et al., 2013), especially related to text difficulty (Schad et al., 2012), within-word fixations, and reading regression (or going back a few words if one did not understand the sentence) (Uzzaman and Joordens, 2011). Given that vision is how we acquire most of our information, it is only logical that our eyes are highly influenced by lapses of attention. These advantages could contribute to make oculometry a necessity for OOTL detection.

ECG and Skin Conductance Measures

Heart rate and skin conductance have been used for a long time to detect periods of boredom (Smith, 1981) and they continue to be part of the latest developments. Their robustness allowed Pham and Wang (2015) to create a classifier which accurately identified lapses of attention during learning. They have also shown promising results when used to determine pilots’ vigilance in real-time (Boucsein et al., 2007). The effects of boredom over amplitude and variability were assessed on both markers. Interestingly, Smallwood et al. (2004) reported similar effects when studying MW. Since MW may favor OOTL situations, heart rate, and skin conductance could also be used to study OOTL. Regrettably, it is possible that MW influence over the signal would be lost within operational environment because stress, movement, and temperature can also play a role in heart rate and skin conductance variations. Consequently, more studies will be required in this field.

Neural Markers

Neural markers of attentional lapses are used to both detect MW and reveal its dynamics. Researchers have mostly used EEG or functional MRI (fMRI) to study those markers, with the notable exception of the HbO2 concentration using functional near-infrared spectroscopy (fNIRS) (see Durantin et al., 2015). EEG activity has a high temporal resolution and a relatively low cost (Luck, 2014), allowing its extensive use for MW research. MW influence on brain waves was suggested by EEG data with an accent on the alpha band (8–14 Hz), although the direction of the influence is still debated (O’Connell et al., 2009; Braboszcz and Delorme, 2011), and event related potentials (ERPs). Sensory attenuation has been observed on the visual component P1 and the auditory component N1 (Kam et al., 2011), while the lack of stimulus processing was shown using P3 (Schoeler et al., 2011), N400 (O’Connell et al., 2009), and fERN (Kam et al., 2012). By contrast, IMRI has a fine spatial resolution but a poor temporal resolution and may be used to detect neuronal networks involved in MW in order to build a map of the wandering
mind. Several studies have highlighted brain regions differently involved in the phenomenon, such as the default mode network (Mason et al., 2007; van den Heuvel and Hulshoff Pol, 2010), the executive network (Christoff et al., 2004, 2009) and the task-positive network (Mittner et al., 2014). Compared to other markers, neural markers of MW could not only answer the question of “when” OOTL occurs, but also the “why” and “how”. This could provide the OOTL problem with the physiological definition that it lacks.

MW research has identified an important set of markers to detect its occurrence. Due to the proximity with psychophysiological measures recently used in automation studies, these markers may also prove to be useful for OOTL research. However, many unknowns still remain regarding some aspects of both phenomena and the feasibility of their study within operational environments is uncertain.

**LIMITS TO CURRENT APPROACHES**

The use of MW findings could be a huge step toward understanding and countering OOTL’s deleterious effects on human performance. MW physiological aspects are for now far better apprehended than for OOTL, while its influence over performance is more precisely assessed, even though many parts of MW remain largely unknown and could limit the transposition.

**Different Levels of MW**

Generally, studies postulate that MW is a binary state, for example when questionnaires ask if the subject is in MW or focused (Braboszcz and Delorme, 2011; Smallwood et al., 2011; Bastian and Sackur, 2013; van Vugt et al., 2015). By contrast, the inattention hypothesis suggested by Smallwood (2011) proposes a gradual view of MW. They manipulated a corpus of text by inserting different types of errors, from pseudowords (lower level errors) to inconsistent statements (higher level errors). During the experiment, those participants who experienced MW exhibited progressive gaze pattern modification depending on error level, supporting a graded nature of the phenomenon. This is in line with findings concerning response time, which mention a progressive acceleration of response times before MW reports (Smallwood et al., 2008; Smallwood, 2010). Cheyne et al. (2009) proposed a three-level model of MW by postulating that response time degradation—slowing, anticipation and omissions—could each correspond to a different level. This hypothesis is empirically confirmed by our ability to perform everyday tasks accurately in spite of MW. For example, driving is still possible with MW (Lerner et al., 2015; Qu et al., 2015) even though it does affect performance. This could also explain why operators can experience MW without systematic OOTL problems. Investigating this possibility will require changing paradigms. Whereas, the probes so far have asked the subject to report their state of mind in a binary fashion, we need to use a scale and compare its results to the evolution of psychophysiological markers. Eventually, taking this parameter into account could allow us to develop systems that are able to discriminate between levels of MW.

**Mind Wandering and Cognitive Fatigue**

It is now clear that MW during driving and piloting tasks decreases short-term performance, especially when the operator is moved to a supervising role. However, the long-term consequences of MW have not been assessed. We experience this on a daily basis—if it was detrimental to survival, there is little doubt that evolution would have removed it (Schooler et al., 2014). Therefore, what are the advantages of such a state of mind? Several papers have highlighted the benefits of MW for curiosity, social skills (McMillan et al., 2013), and creative problem solving (Zedelius and Schooler, 2015, 2016). Another possible advantage of MW could be linked to cognitive fatigue. Humans experience high levels of cognitive fatigue and stress when facing monitoring tasks in monotonous and repetitive environments (Thackray and Touchstone, 1989; Sarter et al., 1997; Warm et al., 2008). At the same time, it has been established that MW propensity increases as the task lasts (Esterman et al., 2013; Pham and Wang, 2015). Therefore, MW may be a mechanism that has been built to decrease cognitive fatigue. Boredom studies mentioned daydreaming as a strategy to cope with boredom within monotonous environments, such as driving, monitoring, or piloting (Davies, 1926; Harris, 2000). The best paradigm to investigate this theory would be to perform real-time tracking and suppress MW as soon as it is detected. Observing the results on mood, fatigue, and arousal could provide precious information about MW’s advantages. Unfortunately this protocol is not possible for now due to MW low detection rates. However, the outcomes would be systems that are able to discriminate between intrusive MW episodes and useful ones, depending on the situation, such as flight phase or traffic density. These systems would reduce OOTL risks while benefitting from MW.

**Real-Time Detection of MW**

When talking about MW research, a straightforward question is to ask if researchers can assess one’s state of mind at a given moment. Such as, whether or not he or she is in MW? This possibility would offer countless possibilities to the study MW. It could, for example, highlight its triggers, assess its benefits, study its dynamic, and define the precise influence of environmental conditions. Recently, studies trying to perform such detection have flourished. They tend to use classifiers, programs that gather information to compare them to a reference and assess if the subject is MW or focused (Delorme et al., 2010). Detection rates are reported through kappa, which is a metric comparing an observed accuracy with an expected one, and included between 0 (random chance) and 1 (exact prediction). Given that reading is an activity where participants do not move much but interact extensively with their environment, it has been the first context used to perform MW detection. Using previous findings (Smallwood et al., 2004) on the influence of MW over galvanic skin conductance, Blanchard et al. (2014) reached a kappa of 0.22. The same kappa was obtained by Pham and Wang (2015) with heart rate variability. Finally, Bixler and D’Mello (2014, 2015) used oculometry during reading to build a classifier which reached 0.31. However, reading is not the only paradigm used for MW detection. Mellinscak et al. (2014) asked the participant to pay attention or ignore some kinesthetic sensation. They
developed a classifier using a passive brain-computer interface (BCI) with a kappa of 0.33, which is the best result so far among MW classifiers. Although using neuroimaging to monitor the participants’ attention seems promising, artifacts on the EEG signal make online processing difficult.

**Multimodal Classifiers**

It is worth noting that, to our knowledge, all studies trying to perform MW or OOTL online detection did so with only one kind of measure—whether it was heart rate, oculometry, or EEG signal—with the notable exception of Bouceïn et al. (2007). It may prove useful to research multimodal classifiers to see if the success rate can be increased. Nevertheless, combining measures would not necessarily result in better detection. Indeed, the main difficulty is to not only design accurate classifiers in order to obtain good prediction but also ensure that the classifiers are sturdy enough for it to be generalized across subjects and conditions. More particularly, high intra- and inter-subject variability make it difficult to build a robust classifier. Intra-subject variability describes the differences observed on one subject depending on their environment. Time, fatigue, and interest are parameters that could influence MW episodes frequency, length and deepness (Smith, 1981; Smallwood et al., 2004; Cummings et al., 2015). Grandy et al. (2013) demonstrated that each human has a stable alpha wave frequency that is independent from cognitive interventions. On the other hand, they observed important differences between subjects in this frequency band. Inter-subject variability often prevents us from building a robust model that is able to be generalized across subjects. One solution is to have the model adapt itself to the user, and then use markers and thresholds that are specific to each individual. However, this model would have a high cost, shortening its range of applications.

**Use MW Detection within Operational Environment**

Although experiments performed in laboratory conditions (e.g., reading and simulators experiments) have produced useful results, they were all performed in a controlled environment. Bixler and D’Mello (2014) have shown the possibility of performing experiments on actual users instead of experimental subjects, although only in a reading tasks. Within an operational environment, systems need to minimize any disruption from the detectors, especially in safety-critical environments. Mkrtchyan et al. (2012) described an ATC interface designed to detect and counter lapses of attention using EEG, thanks to the officer sitting and the stable environment. However, it can be extremely difficult to achieve for pilots and drivers. Not only does the subject variable increase the difficulty to build robust classifiers but conditions of measures can also introduce much noise.

Some systems have recently been designed to overcome these issues. Addressing ease of implementation, dry electrodes measure EEG signal without need for skin preparation (Taheri et al., 1994). Although the signal-to-noise signal is lower and requires further improvement, it could be implemented in operational environments with little disruption for the user, especially if they already wear a helmet, such as jet pilots. Mullen et al. (2015) used this technology to design a wearable EEG system for online neuroimaging with promising results. Recent advances in high-tech industry could produce interesting results in a near future, such as MindRDR (This Place, 2016) or OpenBCI (OpenBCI, 2016). Proving that EEG is not the only promising brain signal measure, Khan and Hong (2015) used fNIRS recorded with a BCI to detect drowsiness with a success rate of ∼84%. Oculometry has also been substantially improved over the past decade, producing efficient, small, and cheap devices. Systems have been proposed with several designs—head-mounted or deployed—and they can be integrated in almost any preexisting system with efficient results. Scanella et al. (2015) showed that flight phases could be differentiated using an eye tracker while demonstrating a remarkable independence regarding inter- and intra-subject variability. Closer to vigilance research, Dehais et al. (2008, 2010) found that an embedded eye tracker allowed detection of gaze features during flight in both nominal and degraded conditions. Several studies have demonstrated the possibility of using EEG for vigilance monitoring in operational environments (Dussault et al., 2005; Jeroski et al., 2014). Cabon et al. (1993) gathered data from ECG put on long range aircrews and train drivers with the device attached on the seatbelts. Bouceïn et al. (2007) recorded the same information—with a more invasive system—to design a flight simulator interface using adaptive automation. Their system could accurately react to varying levels of vigilance. However, the acceptability—which is defined as the capacity of the system to fulfill user’s needs and be accepted for a regular use—was not evaluated during the experiment. Still, these results demonstrate the possibility of building better human-machine interfaces, which could potentially prevent many vigilance related accidents.

**CONCLUSION**

The OOTL phenomenon has been involved in many accidents in safety-critical industries, as demonstrated by papers and reports that we have reviewed. In the near future, the massive use of automation in everyday systems will reinforce this problem. MW may be closely related to OOTL—both involve removal from the task at hand, perception drop, and understanding problems. More importantly, their relation to vigilance decrement and working memory could be the heart of their interactions. Still, the exact causal link remains to be demonstrated. Far from being anecdotal, such a link would allow OOTL research to use theoretical and experimental understanding accumulated on MW. The large range of MW markers could be used to detect OOTL situations and help us to understand the underlying dynamics. On the other hand, designing systems capable of detecting and countering MW might highlight the reason why we all mind wander. Eventually, the expected outcome is a model of OOTL–MW interactions which could be integrated into autonomous systems.

This system description echoes recent advances toward adaptive and communicative automation (Cassell and
Vilhjálmsson, 1999; Sarter, 2000; May and Baldwin, 2009). Adaptive systems could detect and react to operators’ state of mind, including mood, motivation, fatigue, or arousal. The signals sent, information displayed, and levels of automation could be adjusted by the system to maximize situation awareness and vigilance. These systems could detect MW and decide whether it should be stopped or allowed depending on the situation and the characteristics of the episode. Thus, the operator could benefit from MW’s advantages while having a reduced risk of going on to OOTL. The benefits of keeping an operator always in the loop could demonstrate that humans can still be useful in safety favoring industries.

REFERENCES


作者贡献

所有作者都对工作做出了实质性的、直接的和智力的贡献，并同意该文的发布。

致谢

我们感谢法国国家航空组 (DGA) 为他们对我们的第一作者的大力支持。这项工作已经通过 ANR (Young researcher program—ANR-15-CE26-0010-01) 支持。作者也感谢 Sylvain MADEC 为他的宝贵见解和 Isabelle SANTOS 为她的英语校对。

引用


Endsley, M. R., and Rodgers, M. D. (1998), Distribution of attention, situation awareness & workload in a passive air traffic control task. Air Traffic Control Q, 6, 21–44. doi: 10.2514/1.clear.6.1.21


