Consciousness and Cognition 74 (2019) 102777

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Individual differences in higher-level cognitive abilities do not predict overconfidence in complex task performance

Troy A.W. Visser[a,*], Angela D. Bender[a], Vanessa K. Bowden[a], Stephanie C. Black[a], Jayden Greenwell-Barnden[a], Shayne Loft[a], Ottmar V. Lipp[b]

[a] School of Psychological Science, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia
[b] School of Psychology, Curtin University, Kent Street, Bentley, WA 6102, Australia

ARTICLE INFO

Keywords:
Overconfidence
Working memory
Attentional control
Impulsivity
Situation awareness
Air traffic control
Latent-factors

ABSTRACT

Even when people perform tasks poorly, they often report unrealistically positive estimates of their own abilities in these situations. To better understand the origins of such overconfidence, we investigated whether it could be predicted by individual differences in working memory, attentional control, and self-reported trait impulsivity. Overconfidence was estimated by contrasting objective and subjective measures of situation awareness (the ability to perceive and understand task-relevant information in the environment), acquired during a challenging air traffic control simulation. We found no significant relationships between overconfidence and either working memory or attentional control. However, increased impulsivity significantly predicted greater overconfidence. In addition, overall levels of overconfidence were lower in our complex task than in previous studies that used less-complex lab-based tasks. Our results suggest that overconfidence may not be linked to high-level cognitive abilities, but that dynamic tasks with frequent opportunities for performance feedback may reduce misconceptions about personal performance.

1. Introduction

Many everyday tasks, both leisure and job-related, require individuals to make accurate decisions on the basis of limited information. Decades of research on decision making under such conditions suggest that we are often surprisingly inaccurate at estimating our own performance. In particular, we often display an “overconfidence bias” such that we overestimate our abilities relative to others (Alicke & Govorun, 2005; Baboushkin, Hardoon, Derevensky, & Gupta, 2001; Ehrlinger, Mitchum, & Dweck, 2016; Moore & Healy, 2008; Svenson, 1981) and relative to measures of actual performance (e.g., Lichtenstein, Fischhoff, & Phillips, 1977; Stanovich & West, 1998; Ohan & Johnston, 2011; O'Brien & Kardas, 2016).

Overconfidence has been linked to both positive and negative outcomes. With respect to positive outcomes, overconfidence can enhance self-esteem by increasing feelings of knowledge and competency (Blanton, Pelham, DeHart, & Carvallo, 2001), lead to greater innovation in corporate settings (Galasso & Simcoe, 2011; Hirshleifer, Low, & Teoh, 2012), and boost romantic desirability and intrasexual competitiveness (Murphy et al., 2015). There is also evidence that active self-enhancement (a positive bias in a view of ones own capabilities relative to peers) can lead to greater creativity, and resilience (i.e., increased pain resistance in a cold-pressor task; O'Mara & Gaertner, 2017).

On the other hand, overconfidence has also been linked to many serious negative outcomes such as student underachievement

---

* Corresponding author.
E-mail address: troy.visser@uwa.edu.au (T.A.W. Visser).

https://doi.org/10.1016/j.concog.2019.102777
Received 14 February 2019; Received in revised form 21 June 2019; Accepted 24 June 2019

(Dunlosky & Rawson, 2012), risky driving behavior (Horswill & McKenna, 1999; Schlehofer et al., 2010), poor investment decisions (Ronay, Oostrom, Lehmann-Willenbrock, & Van Vugt, 2017), problem gambling (Davis, Sundahl, & Lesbo, 2000), and criminal behavior (Banerjee, Humphery-Jenner, Nanda, & Tham, 2018; Loughran, Paternoster, Piquero, & Fagan, 2013). Furthermore, overconfidence is a prominent feature of disorders such as attention-deficit hyperactivity disorder (Knouse, Bagwell, Barkley, & Murphy, 2005; Ohan & Johnston, 2002, 2011) and narcissism (Campbell, Goodie, & Foster, 2004).

Given the relationship between overconfidence and a variety of both positive and negative outcomes, it is important to develop a better understanding of the mechanisms that lead to overconfidence. Past research in this area has focused on possible contributing factors such as personality traits (Kleitman & Stankov, 2007; Schaefer, Williams, Goodie, & Campbell, 2004), age (Hansson, Rönnlund, Justin, & Nilsson, 2008; Stankov & Crawford, 1996), gender (Barber & Odean, 2001; Huang & Kisgen, 2013; Soll & Klayman, 2004), and the nature of the task completed by participants (Hansson, Juslin, & Winman, 2008). However, relatively less work has directly examined possible links between cognitive ability and overconfidence. To that end, in the present work, we examine the relationship between overconfidence and two high-level cognitive abilities – working memory (WM) and attentional control (AC).

Our decision to concentrate on these particular cognitive abilities derives from both theoretical explanations for overconfidence as well as past research findings. With respect to theory, the heuristics and biases approach suggests that overconfidence is likely to arise from cognitive errors stemming from the inappropriate use of heuristics in decision making (Griffin & Tversky, 1992; Kahneman & Tversky, 1996). One possible corollary of this assertion is that superior cognitive abilities might lead to less reliance on heuristics and thus more accurate judgments about performance. Alternatively, superior AC and WM abilities might lead to the use of more effective heuristics (Raab & Gigerenzer, 2005). In either case, superior cognitive abilities would be expected to ameliorate overconfidence.

Some support for this position can be seen in studies showing a negative relationship between fluid and crystalized intelligence (Horn & Cattell, 1966) and overconfidence (Bruine de Bruin, Parker, & Fischhoff, 2007; Pallier et al., 2002; Stanovich & West, 1998). For example, Stanovich and West (1998) measured overconfidence as the discrepancy between mean percentage accuracy and mean percentage confidence ratings on general knowledge questions. Overconfidence was shown by more than 75% of the sample, and was significantly negatively correlated with a cognitive abilities construct derived from scores on the Raven Advanced Progressive Matrices (Raven, 1962), the Nelson-Denny reading comprehension task (Brown, Bennett, & Hanna, 1981) and the Scholastic Aptitude Test. Results such as these (see also Zakay & Glicksohn, 1992) suggest overconfidence may be negatively related to high-level cognitive functions associated with intelligence and information processing ability.

In addition to work focusing on general intelligence, several studies have looked at the links between overconfidence and more specific high-level cognitive abilities. For example, Camchong, Goodie, McDowell, Gilmore, and Clementz (2007) recruited two participant groups chosen on the basis of scores on an index of problem gambling behavior (which is linked to high levels of overconfidence). On each experimental trial, participants first viewed the names of two American states and were asked to decide which state had a larger population. Participants were then presented with one of the two states shown in the first display and asked to press a response button to indicate whether this response option was a match to the correct answer. Critically, while participants who were not problem gamblers showed an increase in activation in the right temporal-occipital junction when presented with an incorrect response option, this was not seen in problem gamblers. The authors interpreted this as evidence for a link between overconfidence and abnormal activity in the attention system designed to identify task-relevant stimuli.

Molenberghs, Trautwein, Böckler, Singer, and Kanske (2016) presented participants with a short video followed by factual questions and confidence ratings about perceived accuracy. Neural activation, measured by functional magnetic resonance imaging during the confidence ratings, suggested a link between more accurate confidence judgments and reduced activation in the anterior medial prefrontal cortex. Activation in this area, in turn, has been linked to metacognition in prior studies (Fleming, Ryu, Golfinos, & Blackmon, 2014; Fleming, Weil, Nagy, Dolan, & Rees, 2010) as well as episodic memory and reward processing (de la Vega, Chang, Banich, Wager, & Yarkoni, 2016).

In contrast to this work, which suggests potential links between overconfidence and high-level cognition, Del Missier, Mäntylä, and De Bruin (2012) examined the relationship between overconfidence (evaluated in a similar manner to earlier studies by Stanovich & West, 1998, and Bruine de Bruin et al., 2007) and two executive functions: shifting between tasks, and a general "supervision" ability consisting of WM updating and inhibition (Miyake & Friedman, 2012). Using a latent factors analysis that combined results from multiple cognitive tasks to provide a purer measure of shifting and supervision constructs, they found no significant relationships between their measures of executive functioning and overconfidence.

In sum, there are both theoretical and empirical reasons to suggest a link between high-level cognitive functions and overconfidence. However, there is sparse and somewhat conflicting evidence about what specific cognitive functions might be implicated. An additional gap in the literature arises from the use of simple lab-based tasks to assess overconfidence, which have commonly looked at the disparity between knowledge and confidence in the accuracy of that knowledge in a series of discrete trials. Notably, these tasks were characterized by being the sole focus of participants' attention, requiring responses with little or no time pressure, and by the fact that participants did not receive any feedback about performance during the task.

Importantly, these properties of lab-based tasks are likely to differ from real-world tasks, such as driving, where overconfidence is a significant problem. These differences, in turn, may significantly impact the magnitude of overconfidence found in more realistic tasks. For example, many real-world tasks are complex and require rapid and accurate responding to unpredictable environmental events. This might significantly deplete the availability of cognitive resources, limiting the availability of participants to form accurate impressions about their performance, and thus exacerbate overconfidence. Alternatively, real-world tasks also typically provide timely feedback about errors (e.g., other drivers will use their horns, slam on brakes, make rude gestures, etc.). Such feedback provides information that participants might use to make more accurate judgments about their ability, thereby reducing overconfidence. Indeed, Heck and Krueger (2015) examined the impact of performance feedback on self-enhancement effects in a sports knowledge quiz. They found that providing feedback on actual quiz performance reduced self-enhancement amongst those who initially overestimated their performance, consistent with the notion that feedback was incorporated into participant's self assessments.

To explore the issues noted above, the present work looked at the cognitive correlates of overconfidence during a simulated air-traffic control task (ATC; Fothergill, Loft, & Neal, 2009). We chose the ATC task as a broad exemplar of many everyday activities that require visual monitoring, multitasking, and sustained attention, and because it has been widely used by ourselves (e.g., Loft, 2014; Wilson, Farrell, Visser, & Loft, 2018) and other research groups (e.g., Redick et al., 2013) to investigate mechanisms underlying complex task performance. To assess overconfidence, we looked at the difference between an objective and a subjective measure of situation awareness (SA) collected in the course of completing the ATC task. SA is commonly defined as the ability to perceive, understand, and act upon task-relevant information in the environment (2015; Endsley, 1995) and is widely considered to be a key element to task performance in complex, dynamic environments (Endsley & Jones, 2013; Loft et al., 2018; Stanton, Salmon, & Walker, 2015).

This criterion-based approach, although not without its potential drawbacks (Edwards, 1995; Krueger, Heck, & Asendorpf, 2017), is broadly analogous to that employed in many previous studies that operationally defined overconfidence as the difference between accurate responding to knowledge questions and subjective feelings of accuracy on those questions. Our objective measure of SA was the Situation Present Assessment Method (SPAM; Durso, Dattel, Banbury, & Tremblay, 2004), which directly asked participants for specific information about events during the ATC task. This measure of actual knowledge was compared with a second subjective assessment of SA, the Situation Awareness Rating Technique (SART; Taylor, 1990) that was administered at the conclusion of the ATC task and probed subjective SA using a short questionnaire. To calculate an index of overconfidence, we then standardized scores on the two SA measures and then took the difference score between them.

We examined how overconfidence might be related to cognitive performance by administering multiple cognitive tasks that assessed WM and AC. The outcome from a single task is likely to be influenced by measurement error and other statistical factors that make it more difficult to find correlations with other variables, such as overconfidence. To overcome this problem, like Del Missier et al. (2012), we used confirmatory factor analysis (CFA), which is a latent variable technique that statistically extracts the common variance among indicators that are believed to tap the same underlying mechanism. This allows the construction of a 'purer' latent variable (Miyake et al., 2000), that more accurately assesses the underlying cognitive construct to be obtained.

A final issue examined here was the relationship between impulsivity and overconfidence. Impulsivity is commonly conceived as a personality trait (Costa & McCrae, 1992; Eysenck & Eysenck, 1985; Tellegen, 1982) characterized by a proclivity to respond to situations without due consideration of the outcome (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001; Taylor, Visser, Fueggle, Bellgrove, & Fox, 2018). Notably, high levels of impulsivity are related to reduced error monitoring (Stahl & Gibbons, 2007; Taylor et al., 2018) and an increase in risky choices (Kim & Lee, 2011) – behaviours that are also linked to overconfidence. Moreover, high levels of impulsivity are also found in groups that are characterized by overconfidence including problem gamblers (MacKillop et al., 2014) and individuals with elevated levels of narcissism (Campbell et al., 2004). In light of these suggestive parallels, we undertook a more exploratory examination of the potential link between impulsivity and overconfidence. To do this, we assessed trait impulsivity using the Barratt Impulsivieness Scale – Version 11 (BIS-11; Patton, Stanford, & Barratt, 1995). The BIS-11 is a self-report measure that probes long-term evidence for impulsive activities across a variety of contexts. We then examined whether BIS-11 scores predicted overconfidence in SA.

2. Participants

One hundred and eighty-seven participants (67 male, mean age = 21 years, *SD* = 5.36, age range: 17–58 years) were initially recruited from the University of Western Australia's undergraduate participant pool as part of a larger study looking at the relationship between situation awareness and cognitive processing. All participants had normal or corrected to normal vision. The study was approved by the Human Research Ethics Committee of the University of Western Australia and all participants gave informed, written consent. Participants received partial course credit or were compensated AUD10 per hour. They were also able to earn up to AUD20 in bonus payment as an incentive for fast and accurate performance. Data from seven participants were omitted (five who did not attend all experimental sessions; two due to experimenter or computer error on multiple tasks). Fourteen participants did not complete the BIS-11 scale due to experimenter error. This left a final sample of 180 participants for all analyses, except those involving the BIS-11 (N = 166).

3. Materials and procedure

All tasks were presented on a BenQ XL2420T monitor attached to a Pentium computer located in a dimly lit room. Participants were seated approximately 50 cm from the computer monitor. Responses were made by a keyboard and/or mouse attached to the computer, which recorded all response information.

Participants completed the tasks as part of a larger battery administered over 2 two-hour sessions, separated by at least one day (median = 2.00; SD = 9.11). Actual separation time between sessions varied depending on the availability of participants to return to the laboratory. Tasks were either self-paced and/or divided into discrete blocks with rest breaks in between. During the first session, the cognitive tasks were administered in a pseudo-random order with the proviso that tasks assumed to measure the same construct (e.g., working memory) were never administered consecutively and were equally likely to occur at the beginning, middle, and end of the session across participants. During the second session, participants completed the ATC task, with SA measured directly during the task via the SPAM. Perceived SA was assessed with the SART at the end of the ATC scenario, followed by administration of the Barratt Impulsiveness Scale. The order of the tasks in the second session was the same across participants.

3.1. Working memory tasks

3.1.1. Corsi block tapping task (Corsi, 1972)

On each trial, participants were shown a display consisting of nine squares. Two to nine squares then lit up in a pre-arranged sequence, after which participants were prompted to recall the correct sequence by clicking on the squares in the same order they were lit. There were two trials for each sequence length. If at least one of the two trials was completed correctly, the sequence length was increased by one, and the task continued. Otherwise, the task ended. The total number of correctly recalled sequences across the whole task was used as an estimate of WM span (maximum 16).

3.1.2. Operation span task (Unsworth, Heitz, Schrock, & Engle, 2005)

Each trial comprised a series of dyads consisting of a math equation to be solved as quickly and accurately as possible (e.g., $4 * 2 + 1 = ?$), followed by a letter to be encoded. After viewing three to seven of these dyads (randomly inter-mixed; three trials per sequence length), participants were then asked to recall the letters in the correct order. The task began with three practice blocks: (a) letter encoding only, (b) equation solving only, and (c) both letter encoding and equation solving. For the experimental trials, participants were allocated an individualized response window to answer the math equation. This response window calculated as the mean response time on trials in which only the math equation had to be solved, plus 2.5 SDs. The sum of all letters recalled in the correct order was used as an estimate of operation span (maximum score = 75).

3.1.3. Change detection task (Jonides & Yantis, 1988)

In this task, participants were presented with a display of four, six, or eight coloured (black, yellow, purple, white, green, blue, red) squares for 200 ms. This display then disappeared for 900 ms, and was replaced by a single coloured circle presented in one of the locations previously occupied by a coloured square. Participants then reported whether the colour of the circle matched that of the previously-presented square as accurately as possible without time pressure by pressing the “Z” key (coloured squared matched) or “M” (no match) on the keyboard. After the response was made, the circle disappeared, and the next trial began. The experiment consisted of two blocks of 60 trials. The proportion of correct responses across set size 4, 6, and, 8 was used as an estimate of change detection (Tas, Luck, & Hollingworth, 2016).

3.2. Attentional control tasks

3.2.1. Single vs. dual response selection task (Dux et al., 2009)

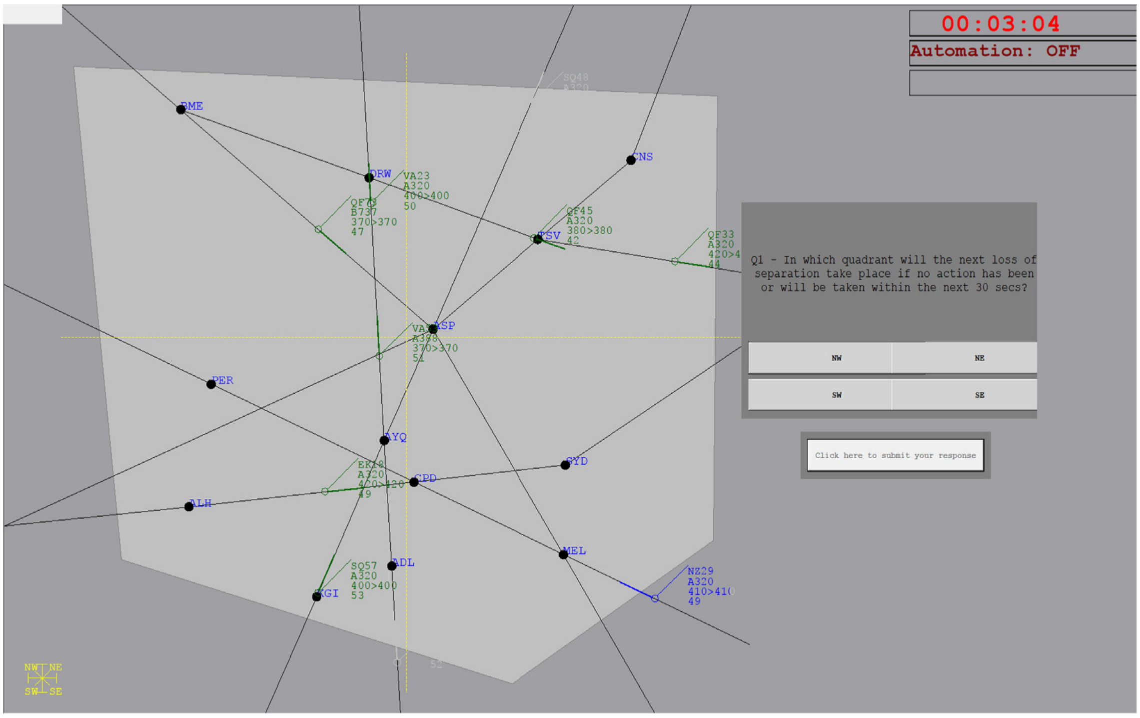
In this task, participants first completed practice blocks comprising two different, two-alternative forced choice (AFC) tasks. In the first block, participants completed 12 trials in which they had to make a speeded judgment about the colour of a circle presented on the display. In the second block, participants completed 12 trials in which they had to make a speeded judgment about the identity of a complex tone (Filmer, Mattingley, Marois, & Dux, 2013). Participants then performed a final practice block, in which they identified as quickly as possible a randomly-selected circle and tone that were presented simultaneously. Response mappings were counterbalanced across keys and hands. The experimental phase consisted of four blocks of 36 trials, with each of the trial types from the practice blocks randomly intermixed. The difference in reaction times between the visual and auditory tasks when performed alone and when performed together was used as an indicator of attentional control, with smaller difference scores indicating better performance.

3.2.2. Psychological refractory period (Van Selst, Ruthruff, & Johnston, 1999)

Participants initially completed three practice blocks (20 trials each), performing two different four-AFC tasks. In the first practice block, participants completed a visual task (Task 1), in which they were asked to identify one of four different symbols (#, &, @, or %) as quickly as possible by pressing a marked key. In the second practice block, participants were asked to identify one of four complex tones as quickly as possible by pressing a different marked key (Filmer et al., 2013). For each task, participants were assigned a response hand and a specific response mapping (H, J, K, or L for right hand responses and A, S, D, or F for left hand responses), with the mapping of hand to task counterbalanced across participants. After practicing these stimulus-response mappings, participants completed a third practice block that combined the first two. Participants were first presented with the symbol (T1) for 200 ms, a short (200 ms) or long (1000 ms) inter-stimulus interval (ISI), and then a complex tone (T2) for 200 ms. Participants had to respond as quickly and accurately as possible to both T1 and T2, without grouping (Tombu & Jolicœur, 2005) the responses. The experimental phase was identical to the final practice block and consisted of 4 blocks of 40 trials, with the two ISIs presented equally often in each block. The difference between mean reaction times at the shorter and longer ISIs was used as an indicator of attentional control, with smaller difference scores indicating better performance.

3.2.3. Attentional blink (Raymond, Shapiro, & Arnell, 1992; Visser, Davis, & Ohan, 2009)

On each trial, participants were presented with a central fixation for 100 ms, followed by a rapid serial visual presentation (RSVP)



Q1 - In which quadrant will the next loss of separation take place if no action has been or will be taken within the next 30 secs?

Fig. 1. Screenshot of the air traffic control task (Fothergill et al., 2009) with a Situation Present Assessment Method (SPAM; Durso et al., 2004*)* question appearing on the right side of the display.

stream of items presented at the centre of the display. The RSVP consisted of two black target letters (excluding I, L, O, Q, U, V, X) presented amongst eight black distractors consisting of keyboard symbols (#, $, %, &) and digits (2–9). Each RSVP item was presented for 200 ms and separated from the next by an 80 ms ISI during which the display was blank. The two target letters were separated by 200 ms (lag 2), 300 ms (lag 3), or 800 ms (lag 8). After the final RSVP item, participants were prompted to identify each target letter at their leisure, and to guess if unsure. Participants completed 24 practice trials, followed by 96 experimental trials, divided into four equal blocks of 24 trials. Attentional blink magnitude, calculated by summing across the average T2|T1 accuracy from lag 3 and lag 8, was used as an indicator of attentional control, with a smaller score indicating better performance.

3.3. *Complex task performance measure*

Air Traffic Control (Fothergill et al., 2009). The ATC task is a simulation of en route air traffic control system that requires participants to ensure all aircraft proceed safely through their control sector. The simulation is designed to be broadly representative of tasks performed by air traffic controllers. Across two 30-minute scenarios (one practice and one testing scenario), participants were presented with a dynamic display (see Fig. 1) consisting of 8–12 aircraft travelling on designated flight paths, at varying altitudes and rates of speed (indicated on the display adjacent to each aircraft). Participants were required to “accept” aircraft into their sector and “handoff” aircraft leaving their sector by selecting the aircraft icon by clicking on it, and then pressing a designated response key (‘A’ and ‘H’ respectively). Aircraft requiring one of these responses began to flash six seconds prior to entering or exiting the sector, and responses were required within 15 s to be considered correctly accepted or handed-off. Participants also needed to monitor for potential “conflicts” between aircraft travelling in their sector – defined two aircraft as being within 5 nautical miles and at the same altitude. If a conflict was detected, participants were instructed to intervene by selecting one of the aircrafts with their mouse and using the keyboard to enter a new designated altitude. Missed conflicts were indicated by auditory and visual (aircraft changed colour) feedback. The response latency to correctly detect a conflict and to accept and hand-off aircraft were combined into a single latent factor as a measure of ATC performance (see Bender et al., in preparation).

3.4. *Situation awareness measures*

3.4.1. *Situation present assessment method (SPAM; Durso et al., 2004)*

To objectively assess SA, a modified SPAM probe technique was employed during the ATC task. Each probe began with a prompt (“Ready for Question?”), paired with an auditory alert tone in the participant’s headset (see Fig. 1). Participants were instructed to click the “Ready” button below this prompt to initiate SPAM questions as soon as their workload allowed it. After clicking the button, the ATC display froze, and a series of four probe questions were presented in the same location as the SPAM prompt. Each probe

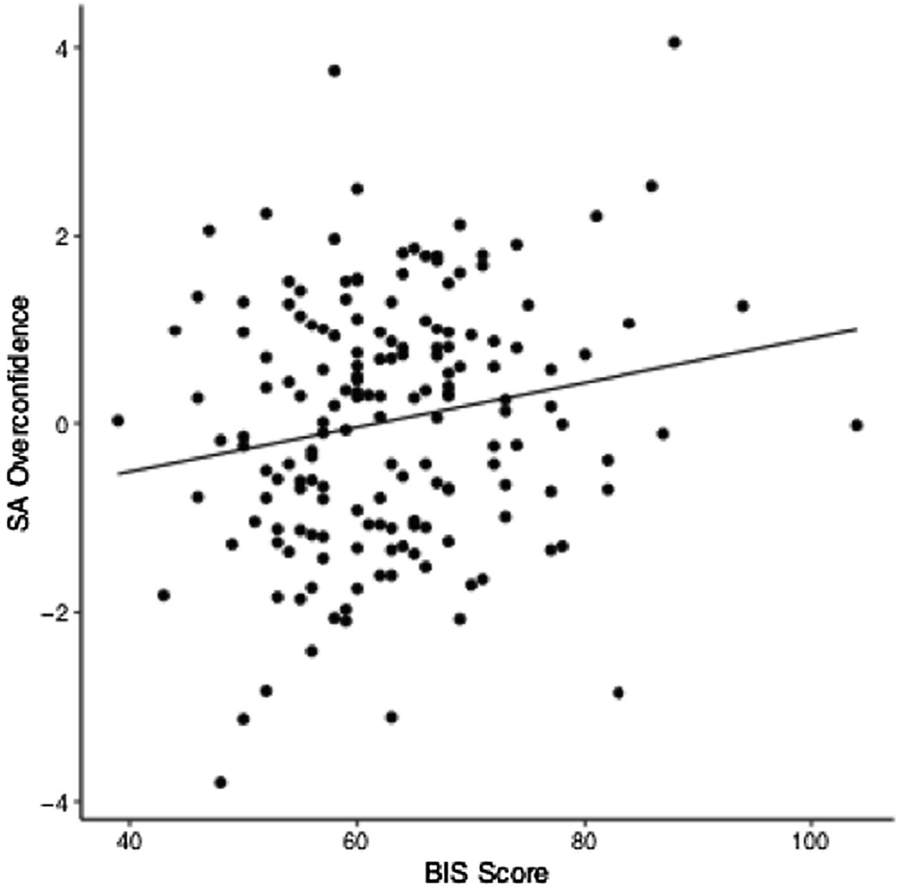


Fig. 2. Scatterplot showing relationship between index of overconfidence ($SA_{overconf}$) and Barratt Impulsiveness Score (BIS).

question had four response options, and participants were asked to choose one of these options as quickly and as possible while trying to maximize accuracy. Once the participants had responded to all probe questions, the ATC scenario resumed. For both practice and test scenarios, the first probe query was presented ∼2–3 min into the scenario and additional probe questions were presented every ∼1.5–2 min (practice scenario = 15 questions, test scenario = 18 questions). If a "Ready for Question?" prompt was not responded to after 30 s, it was removed, and no SA question was presented. SPAM accuracy is typically at ceiling because of task instructions and the fact that the information required to answer each query is visible on the display. On trials where prompts were answered incorrectly, it is difficult to establish the source of error, and thus whether participants were focusing on the SPAM task. For these reasons, mean RT to correctly answered probe questions was used as the objective measure of SA, with faster RTs indicating greater SA. The rationale is that individuals who are maintaining better SA should know, or know where to find, the appropriate information to address the query and are thus able to respond faster to SPAM queries than individuals with poor SA (Chiappe et al., 2016; Durso et al., 2004).

3.4.2. Situation awareness rating technique (SART; Taylor, 1990; Salmon et al., 2009)

The SART questionnaire is a widely used instrument to assess participant's subjective assessment about their SA during a task. We used a version of the questionnaire drawn from Salmon et al. (2009; Fig. 2) consisting of nine questions. Responses to each question were made on a seven-point Likert scale (1 = low; 7 = high) and each question assessed one of three dimension related to SA. The three dimensions were: demand on attentional resources, supply of attentional resources, and understanding of the situation/task. An overall SART score was calculated by employing the following formula: SA = U – (D-S), where U = understanding (two questions), D = demand (four questions), and S = supply (three questions). Thus higher scores indicated greater perceptions of SA.

3.5. Impulsivity measure

Barratt Impulsiveness Scale (BIS-11; Patton et al., 1995*).* Impulsiveness was probed with 30 questions requiring responses on a 4-point Likert scale (1 = rarely/never, 4 = almost always/always). Questions focused on the prevalence of behaviours potentially associated with impulsive behaviour, with questions such as "I do things without thinking", "I get easily bored when solving thought problems", or "I spend or charge more than I earn". The minimum score is 30 (low impulsivity) and the maximum score is 120 (high

Table 1

Descriptive statistics for all measures. O-Span = Operation Span; Corsi = Corsi Block Tapping Task; Ch Det = Change Detection; AB = Attentional Blink; Dual Sin = Single vs. Dual Task Cost; PRP = Psychological Refractory Period; Conflict Det = ATC Conflict Detection measure; Acc-Hand = ATC Acceptance and Hand-off measure; SPAM = Situation Present Assessment Method; SART = Situation Awareness Rating Technique; $SA_{overconf}$ = Situation Awareness Overconfidence Index; BIS = Barratt Impulsiveness Scale. Reliability estimates calculated using either [a]Intraclass correlation coefficient (ICC), [b]Cronbach's alpha.

Measure	*M*	*SD*	Std. Error	Skew	Kurtosis	Reliability
O-Span	57.52	12.21	0.91	−0.908	0.39	0.70[a]
Corsi	9.49	1.84	0.14	0.62	0.07	0.76[a]
Ch Det	70.72	8.28	0.62	−0.61	0.95	0.59[a]
AB	38.88	23.48	1.75	0.65	−0.12	0.68[a]
Dual Sin	434.80	187.40	13.97	0.44	0.16	0.91[a]
PRP	255.2	191.20	14.25	0.95	1.67	0.98[a]
Conflict Det	111.00	38.82	2.89	−0.43	−0.42	0.90[a]
Acc-Hand	2528.00	924.60	68.92	0.81	0.23	0.89[a]
SPAM	14362.82	3482.10	259.5	0.59	0.26	0.90[a]
SART	18.72	5.10	0.38	0.11	0.77	0.72[b]
$SA_{overconf}$	0.00006	1.34	0.10	−0.14	0.21	0.73[b]
BIS	62.91	10.13	0.79	0.75	1.40	0.73[b]

impulsivity).

4. Results

Outlier screening was performed separately for each participant and task. The value of each outcome variable above or below 3 *SD*s was winsorised with the 3 *SD* cut-off value. As can be seen in Table 1, all measures were approximately normally distributed, with values of kurtosis and skewness below generally accepted values (i.e., kurtosis < 4 and skewness < 2; Kline 1998). In addition, confirming our assumptions on the basis of task instructions and the presence of the required information on the display, SPAM accuracy was 90.50% (SD = 8.50), justifying our use of correct SPAM RT in the overconfidence index.

4.1. Index of SA overconfidence

In order to facilitate interpretation of the results, SART scores were first transformed so that lower SART scores reflected better SA, mirroring the relationship between SPAM RT and SA. Difference scores between responses on the objective (SPAM) and subjective (SART) SA scales for each participant were then calculated to derive an overconfidence index. To do this, scores on each measure were first standardized, and then a difference score for each participant was calculated using the following formula: $SA_{overconf} = SART_{z\text{-score}} - SPAM_{z\text{-score}}$, where $SART_{z\text{-score}}$ represents standardized SART performance and $SPAM_{z\text{-score}}$ represents standardized SPAM performance. Thus, a positive $SA_{overconf}$ score reflects overconfidence (i.e., participants evaluated themselves more positively than their actual SA ability), while a negative score reflects underconfidence (i.e., participants evaluated their SA ability more negatively than their actual SA performance).

4.2. Confirmatory factor analysis

Prior to evaluating the relationship between $SA_{overconf}$, AC, WM and complex task performance, we addressed the nature of the AC and WM constructs by employing confirmatory factor analysis to test two separate models and then compared fit statistics to select the best fitting model via multiple fit indices. The fit of each model was evaluated with the chi-square statistic, Akaike's information criterion (AIC), the standardized root-mean-square residual (SRMR). The chi-square statistic (non-significant values indicate a satisfactory fit), and AIC (lower AIC values indicate better fit of the model), both measure the fit between the predicted and observed covariances, while the SRMR (values between 0.05 and 0.08 indicate an acceptable fit (Hu & Bentler, 1998) represents the square root of the average covariance residuals between the predicted and observed model. We also examined the models with the Bentler's comparative fit index (CFI), which compares each model to an independent baseline model. Here values closest to 1.00 are indicative of a good model fit (Hu & Bentler, 1998).

We performed chi-square difference tests on the nested data to examine whether the full model significantly improved model fit compared to the alternative, more restrictive model. To do this, we subtracted the chi-square value and the degrees of freedom of the full model from the restricted model. A non-significant chi-square difference indicates that the restrictive model represents a significantly better fit. Model 1 tested a single factor model in which all of the tasks loaded onto one single factor. The fit of the model was acceptable, X^2 (9) = 14.18, p = .12, RMSEA = 0.06, SRMR = 0.05, NNFI = 0.84, CFI = 0.91, AIC = 3027.61. Model 2 tested a two-factor model in which WM (O-Span, Corsi, change detection) loaded onto one factor and the attentional control measures (PRP, dual vs. single task, AB) loaded onto another factor. These two factors were allowed to correlate. The fit of this model was excellent, X^2 (8) = 4.20, p = .84, RMSEA = < 0.001, SRMR = 0.03, NNFI = 1.00, CFI = 1.00, AIC = 3019.97. The statistical fit of the two-factor model (Model 2) was better than the fit of the single factor model (Model 1), $\Delta\chi^2$ (1) = 9.98, $p < .01$. More importantly, when

Table 2

Pearson correlations between variables of interest. Numbers in brackets represent p-values. ATC = Air Traffic Control performance composite, $SA_{overconf}$ = Situation Awareness Overconfidence index, AC = attention control composite; WM = working memory composite; BIS = Barratt Impulsiveness Scale.

Measure	1	2	3	4	5
1. ATC	–				
2. $SA_{overconf}$	−0.176 (0.018)	–			
3. AC	0.153 (0.040)	−0.146 (0.051)	–		
4. WM	0.311 (< 0.001)	−0.073 (0.327)	0.278 (< 0.001)	–	
5. BIS	−0.025 (0.753)	0.179 (0.021)	−0.122 (0.118)	0.022 (0.780)	–

comparing models, Model 2 had smaller AIC, CFI, RMSEA and SRMR values compared to Model 1. This suggests that performance across the six cognitive tasks could be best explained by two separate constructs.

Finally, we tested a single-factor model to establish the dimensionality for ATC performance variables (Acceptance and Handoff RT and Conflict Detection). The Acceptance and Handoff RT and Conflict Detection measures loaded significantly on the ATC latent variable (0.79, *ps* = < 0.001), indicating that the two ATC tasks measured a common ATC factor.

4.3. Factor composites

We used regression analyses to examine the predictors' contribution to variance in the participants' confidence as measured by $SA_{overconf}$. For the regression analyses, we created factor composites for the cognitive predictors (WM and AC). To facilitate interpretation of the results, the WM indices were transformed, so that lower values indicated better ability across all variables. To form the WM factor composite, the Corsi, O-Span, and Change Detection were included, while the Single vs. Dual, PRP, and AB indicators were included in the attention control factor (as per our CFA results). We also used the factor composite for ATC performance. For all factors, component tasks were entered into an exploratory factor analysis (principal-axis factor extraction) and a one-factor solution was specified. Factor scores for each participant and for each construct were then used in all subsequent analyses.

4.4. Correlations

The correlations among the five variables of interest are shown in Table 2, while a more detailed correlation matrix including all measures can be seen in Table 3. Inspection of Table 2 indicates ATC performance was significantly positively related to AC and WM.

Table 3

Pearson correlations for all measures. Numbers in brackets represent p-values. O-Span = Operation Span; Corsi = Corsi Block Tapping Task; Ch Det = Change Detection; AB = Attentional Blink; Dual Sin = Single vs. Dual Task Cost; PRP = Psychological Refractory Period; Conflict Det = ATC Conflict Detection measure; Acc-Hand = ATC Acceptance and Hand-off measure; SPAM = Situation Present Assessment Method; SART = Situation Awareness Rating Technique; BIS = Barratt Impulsiveness Scale.

	1	2	3	4	5	6	7	8	9	10	11
1. Ospan	–										
2. Corsi	0.289 (< 0.001)	–									
3. Ch Det	−0.126 (0.092)	−0.262 (< 0.001)	–								
4. AB	−0.013 (0.864)	−0.133 (0.076)	0.073 (0.328)	–							
5. Dual Sin	−0.108 (0.149)	−0.264 (< 0.001)	0.066 (0.378)	0.181 (0.015)	–						
6. PRP	−0.080 (0.284)	−0.133 (0.075)	0.162 (0.029)	0.165 (0.027)	0.257 (< 0.001)	–					
7. Conflict Det	−0.189 (0.011)	−0.192 (0.010)	0.127 (0.089)	−0.065 (0.387)	0.139 (0.063)	0.022 (0.767)	–				
8. Acc-Hand	−0.263 (< 0.001)	−0.206 (0.005)	0.250 (< 0.001)	0.028 (0.708)	0.172 (0.021)	0.147 (0.049)	0.338 (< 0.001)	–			
9. SPAM	−0.130 (0.081)	−0.248 (< 0.001)	0.130 (0.082)	0.100 (0.181)	0.223 (0.003)	0.151 (0.043)	0.210 (0.005)	0.396 (< 0.001)	–		
10. SART	0.137 (0.067)	0.142 (0.057)	−0.078 (0.301)	0.109 (0.147)	−0.030 (0.693)	−0.121 (0.107)	−0.115 (0.124)	−0.106 (0.159)	−0.101 (0.179)	–	
11. BIS	0.012 (0.875)	−0.014 (0.858)	0.066 (0.397)	−0.030 (0.701)	−0.157 (0.043)	−0.038 (0.623)	−0.002 (0.982)	−0.039 (0.615)	−0.048 (0.541)	−0.189 (0.015)	–

Table 4

Summary of Hierarchical Regression Analysis using Barratt Impulsiveness Scale (BIS), Attentional Control (AC) Composite Factor Scores and Working Memory (WM) Composite Factor Scores to predict overconfidence (* denotes p < .05).

	b	*SE*	*β*	*t*	R^2	ΔR^2
BIS	0.024	0.010	0.179	2.332	0.032	0.032*
AC composite	−0.230	0.152	−0.116	−1.511	0.045	0.013
WM composite	−0.091	0.135	−0.054	−0.679	0.048	0.003

ATC performance was also significantly negatively correlated with $SA_{overconf}$, such that individuals with greater overconfidence performed better on the ATC task (lower RTs in the ATC task reflect better performance). Finally, as can be seen in Fig. 2, $SA_{overconf}$ was significantly positively correlated with BIS, such that people with greater overconfidence tended to be also more impulsive.

While there is some literature to suggest that overconfidence can be related to poorer performance (Dunlosky & Rawson, 2012; Ronay et al., 2017), the link between overconfidence and better ATC performance found here may reflect the fact that our participants largely showed accurate judgement about their SA ability. Ninety-four participants (52.2%) judged their perceived SA ability to be higher than their actual SA, while the mean overconfidence score across participants was not significantly different from zero ($t(179) < 0.01$, $p > .99$). By comparison, many previous studies that used conventional lab tasks found high levels of overconfidence present in the majority of participants (e.g., Stanovich & West, 1998). In addition, earlier studies found suggestive evidence that the link between overconfidence and performance was mediated by inappropriate levels of risk aversion (Dunlosky & Rawson, 2012) or risk seeking (Ronay et al., 2017). By comparison, there were limited opportunities for risk to influence performance in the ATC task, thus also potentially limiting the possibility for a relationship between overconfidence and task performance.

4.5. Hierarchical regression

To examine the main question of how overconfidence is linked to AC, WM, and impulsivity, we conducted a hierarchical regression analysis. We entered BIS in the first step, AC in the second step, and WM in the third step. As shown in Table 4, examination of the beta coefficients indicates that impulsiveness significantly predicted overconfidence variance ($p = .021$). However, neither AC ($p = .133$) nor WM ($p = .498$) were significant predictors of overconfidence.

5. General discussion

We are often called upon to make accurate decisions in the face of uncertainty. Taking a driving test or an examination, for example, may call upon us to respond in situations where we do not possess sufficient information to be sure about the accuracy of our performance. A common finding in such situations is that people are surprisingly overconfident in their abilities, both relative to their actual performance and relative to the population as a whole. The primary goal of the present work was to examine possible cognitive correlates of such overconfidence.

In a departure from many previous studies, we measured overconfidence relative to performance in a dynamic, simulated ATC task by comparing objective and subjective assessments of SA. This ATC task was designed to provide performance feedback and to impose significantly greater cognitive demands than most tasks used in previous studies of overconfidence. We also assessed individual differences in cognitive abilities using a latent factors approach that yielded more reliable indices of these abilities by using multiple measures of WM and AC.

Several key results emerged that will be discussed below. First, analyses showed no significant relationships between overconfidence and either WM or AC. This is despite earlier studies showing a link between overconfidence and intelligence, and activation in brain regions associated with high-level cognitive functions (Camchong et al., 2007; Molenberghs et al., 2016). Second, levels of overconfidence across our participants were modest, with a relatively small proportion of participants exhibiting overconfidence, and mean overconfidence levels not significantly different from zero. Finally, our results indicated a significant relationship between overconfidence and impulsivity, with participants higher in impulsivity more likely to show overconfidence in their SA.

With respect to our finding of relatively low levels of overall overconfidence, several possible explanations emerge. One particularly important factor may be the availability of feedback both on the ATC task and on the objective SA measure. In the case of the ATC task, failures to accept or handoff aircraft and missed conflict detections were clearly indicated by auditory and visual feedback. In the case of objective SA measures, participants with poor SA would have received indirect feedback, as poor SA would be reflected in slow responses. Such information may have supported more accurate metacognition during the subsequent subjective SA task. This is consistent with earlier work showing a reduction in self-enhancement errors when over-confident participants were provided with performance feedback (Heck & Krueger, 2015)

Another possibility is that our task reduced overconfidence by providing additional clarity about the nature of the data on which to base subjective judgements of SA performance. Previous laboratory-based tasks have been criticized on the grounds that the questions create conflicts between the information given in the task and prior knowledge of participants about the real world (Gigerenzer, 2004; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin, 1994). Our current approach avoided this possibility by using a novel task that participants did not possess any prior knowledge about. The avoidance of such conflicting information may have helped participants to more accurately calibrate their own confidence on the basis of task performance and immediate in-task feedback.

The low levels of overconfidence found in our sample also provide evidence against the possibility that overconfidence would be exacerbated in complex situations requiring multi-tasking and fast responding. We reasoned that such an option might be possible in view of extensive literature showing negative effects of such demands on performance in lab-based (e.g., Pashler, 1994), simulated (Bowden, Loft, Tatasciore, & Visser, 2017), and real-world tasks (Dismukes, 2007). Here, however, there was no evidence of an analogous impact of multitasking demands on metacognition. Of course, it is possible that this null effect is particular to our task parameters. Our performance feedback may have been so salient as to overcome competing attentional demands, or the fact that subjective SA was assessed after the ATC task had finished may have blunted the impact of multitasking requirements during the objective ATC task. These possibilities would be worth exploring in future experiments.

The chief aim of the present work was to determine whether individual differences in levels of WM and AC were related to levels of overconfidence. Despite some suggestive evidence for such a link in past studies, no significant relationships emerged in the present sample. Interestingly, this outcome aligns best with a previous study that also used a latent factors analysis technique to assess cognitive performance (Del Missier et al., 2012). This study found overconfidence was unrelated to executive functions, including task-set shifting, WM and inhibition.

Our failure to find a relationship between either WM or AC and overconfidence could reflect our use of a novel, relatively-demanding simulation task. However, we find this explanation unlikely, given that Del Missier et al. (2012) also failed to find a relationship between executive functioning and overconfidence in the context of a more conventional general knowledge task. It might also be argued that our tasks failed to reliably assess WM or AC. However, we believe this is also unlikely as the latent factors approach, using the same component tasks used here, has been used to reliably assess WM and AC in previous studies (e.g., Miyake & Friedman, 2012; Kane et al., 2004; although see Karr et al., 2018) and typically produces more reliable assessments of underlying constructs than single measures used in previous studies in the overconfidence literature. Further, our component measures have been widely used in previous studies to assess WM and AC.

While we cannot, of course, entirely rule out the possibility that our results reflect an idiosyncratic aspect of our experimental design, task or participants, it is likely that neither WM nor AC are key cognitive skills required for achieving appropriate levels of confidence in simulated ATC task performance. This still leaves open the possibility that such skills are required for appropriately judging confidence in one's ability relative to others in the population – a component of overconfidence that we did not assess here. However, it also suggests that future studies may wish to focus on other cognitive skills associated with intelligence and decision making, such as processing speed (Salthouse, 2005) and numeracy (Dieckmann, 2008) as possible correlates with overconfidence. More broadly, it may also be informative to assess the relationship between overconfidence and attributes other than cognitive skills, such as resilience (Gucciardi, Hanton, Gordon, Mallett, & Temby, 2015) and learning style (Dweck & Leggett, 1988).

In this vein, we found a significant relationship between impulsivity and overconfidence, with greater impulsiveness associated with higher levels of overconfidence. This outcome is consistent with previous work showing a high level of co-occurrence between these attributes in populations such as problem gamblers (MacKillop et al., 2014) and individuals with elevated levels of narcissism (Campbell et al., 2004). It is also consistent with evidence that impulsivity and overconfidence are related to reduced error monitoring (Stahl & Gibbons, 2007; Taylor et al., 2018) and increased risk taking (Kim & Lee, 2011).

The correlative nature of our study prevents us from making strong conclusions about possible causal relationships between impulsivity and overconfidence. One possible option is that overconfidence leads to an increase in impulsive behaviour. That is, if an individual believes they are more skilled than others or that they actually are, they may evaluate situations as less risky than others and behave accordingly (i.e., they would behave "more rashly" than those who are not overconfident). Alternatively, the evidence is consistent with the possibility that both overconfidence and impulsive behaviour jointly arise from other factors yet to be identified. Addressing these possibilities is an important avenue for future research.

In conclusion, the present work investigated whether AC and WM were significantly related to overconfidence in the context of a simulated ATC task that mirrors some of the perceptual, cognitive, and response demands of everyday tasks such as driving. The results showed no relationship between these factors, but intriguingly did find a relationship between impulsivity and overconfidence, as well as low levels of overconfidence across participants compared to lab-based tasks in previous studies. This work opens up interesting avenues for future work that should look at a greater variety of possible predictors of overconfidence, as well as the nature of overconfidence in more realistic real-world tasks.

Acknowledgements

We gratefully acknowledge the The Commonwealth of Australia for supporting this research through the Australian Army and Defence Science Partnerships agreement of the Defence Science and Technology Group (ID 7120), as part of the Human Performance Research network (HPRnet). We would also like to thank Susannah Whitney for thoughtful discussions and a keen eye throughout this and other aspects of our project.

References

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. *The Self in Social Judgment, 1*, 85–106.

Baboushkin, H. R., Hardoon, K. K., Derevensky, J. L., & Gupta, R. (2001). Underlying cognitions in gambling behavior among university students. *Journal of Applied Social Psychology, 31*(7), 1409–1430.

Banerjee, S., Humphery-Jenner, M., Nanda, V., & Tham, M. (2018). Executive overconfidence and securities class actions. *Journal of Financial and Quantitative Analysis, 53*(6), 2685–2719.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics, 116*(1), 261–292.

Bender, A.D., Loft, S.D., Lipp, O.V., Bowden, V.K., Gignac, G.E., Whitney, S.J., Long, D.D., Visser, T.A.W. (in preparation). Cognitive predictors of complex task performance: On the influence of attention control, working memory and situation-specific task models.

Blanton, H., Pelham, B. W., DeHart, T., & Carvallo, M. (2001). Overconfidence as dissonance reduction. *Journal of Experimental Social Psychology, 37*, 373–385.

Bowden, V. K., Loft, S., Tatasciore, M., & Visser, T. A. W. (2017). Lowering thresholds for speed limit enforcement impairs peripheral object detection and increases driver subjective workload. *Accident Analysis & Prevention, 98*, 118–122.

Brown, J., Bennett, J., & Hanna, G. (1981). *The Nelson-Denny Reading Test.* Lombard, IL: Riverside.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92*(5), 938.

Camchong, J., Goodie, A. S., McDowell, J. E., Gilmore, C. S., & Clementz, B. A. (2007). A cognitive neuroscience approach to studying the role of overconfidence in problem gambling. *Journal of Gambling Studies, 23*(2), 185–199.

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making, 17*(4), 297–311.

Chiappe, D., Morgan, C. A., Kraut, J., Ziccardi, J., Sturre, L., Strybel, T. Z., & Vu, K.-P. L. (2016). Evaluating probe techniques and a situated theory of situation awareness. *Journal of Experimental Psychology: Applied, 22*(4), 436–454.

Corsi, P. (1972). Memory and the medial temporal region of the brain. Unpublished doctoral dissertation, McGill University, Montreal, QB.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEOFFI) professional manual.* Odessa (FL): Psychological Assessment Resources.

Davis, D., Sundahl, I., & Lesbo, M. (2000). Illusory personal control as a determinant of bet size and type in casino craps games 1. *Journal of Applied Social Psychology, 30*(6), 1224–1242.

de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *Journal of Neuroscience, 36*(24), 6553–6562.

Del Missier, F., Mäntylä, T., & De Bruin, W. B. (2012). Decision-making competence, executive functioning, and general cognitive abilities. *Journal of Behavioral Decision Making, 25*(4), 331–351.

Dieckmann, N. (2008). Numeracy: A review of the literature (Report No. 08-2). Eugene, OR: Decision Research Notes.

Dismukes, R. K. (2007). Prospective memory in aviation and everyday settings. In M. Endsley, M. R., & Jones, W. (2013). Situation awareness. The Oxford handbook of cognitive engineering, 1, 88–108.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280.

Durso, F. T., Dattel, A. R., Banbury, S., & Tremblay, S. (2004). SPAM: The real-time assessment of SA. *A Cognitive Approach to Situation Awareness: Theory and Application, 1*, 137–154.

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron, 63*(1), 127–138.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256.

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64*, 307–324.

Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94–100.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65–84.

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making, 9*(1), 4–32.

Endsley, M. R., & Jones, W. (2013). Situation awareness. In J. D. Lee, & A. Kirlik (Eds.). *The Oxford handbook of cognitive engineering* (pp. 88–108). Oxford, UK: Oxford University Press.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum Press.

Filmer, H. L., Mattingley, J. B., Marois, R., & Dux, P. E. (2013). Disrupting prefrontal cortex prevents performance gains from sensory-motor training. *Journal of Neuroscience, 33*(47), 18654–18660.

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain, 137*(10), 2811–2822.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329*(5998), 1541–1543.

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab Advanced: An air traffic control simulator with realism and control. *Behavior Research Methods, 41*(1), 118–127.

Galasso, A., & Simcoe, T. S. (2011). CEO overconfidence and innovation. *Management Science, 57*(8), 1469–1484.

Gigerenzer, G. (2004). The irrationality paradox. *Behavioral and Brain Sciences, 27*, 336–338.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology, 24*(3), 411–435.

Gucciardi, D. F., Hanton, S., Gordon, S., Mallett, C. J., & Temby, P. (2015). The concept of mental toughness: Tests of dimensionality, nomological network, and traitness. *Journal of Personality, 83*(1), 26–44.

Hansson, P., Juslin, P., & Winman, A. (2008). The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1027.

Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L. G. (2008). Adult age differences in the realism of confidence judgments: Overconfidence, format dependence, and cognitive predictors. *Psychology and Aging, 23*(3), 531.

Heck, P. R., & Krueger, J. I. (2015). Self-enhancement diminished. *Journal of Experimental Psychology: General, 144*(5), 1003–1020.

Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are overconfident CEOs better innovators? *The Journal of Finance, 67*(4), 1457–1498.

Horswill, M. S., & McKenna, F. P. (1999). The effect of perceived control on risk taking 1. *Journal of Applied Social Psychology, 29*(2), 377–391.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424.

Huang, J., & Kisgen, D. J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics, 108*(3), 822–839.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & psychophysics, 43*(4), 346–354.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*(3), 582–591.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189.

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147–1185.

Kim, S., & Lee, D. (2011). Prefrontal cortex and impulsive decision making. *Biological Psychiatry, 69*(12), 1140–1146.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161–173.

Knouse, L. E., Bagwell, C. L., Barkley, R. A., & Murphy, K. R. (2005). Accuracy of self-evaluation in adults with ADHD: Evidence from a driving study. *Journal of Attention Disorders, 8*(4), 221–234.

Krueger, J. I., Heck, P. R., & Asendorpf, J. B. (2017). Self-enhancement: Conceptualization and assessment. *Collabra: Psychology, 3*(1), 1–11.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. *Decision making and change in human affairs* (pp. 275–324). Dordrecht: Springer.

Loft, S. (2014). Applying psychological science to examine prospective memory in simulated air traffic control. *Current Directions in Psychological Science, 23*, 326–331.

Loft, S., Jooste, L., Li, R., Ballard, T., Huf, S., Lipp, O. V., & Visser, T. A. W. (2018). Using situation awareness and workload to predict performance in submarine track management: A multilevel approach. *Human Factors, 60*, 978–991.

Loughran, T. A., Paternoster, R., Piquero, A. R., & Fagan, J. (2013). "A Good Man Always Knows His Limitations" The role of overconfidence in criminal offending. *Journal of Research in Crime and Delinquency, 50*(3), 327–358.

MacKillop, J., Miller, J. D., Fortune, E., Maples, J., Lance, C. E., Campbell, W. K., & Goodie, A. S. (2014). Multidimensional examination of impulsivity in relation to disordered gambling. *Experimental and Clinical Psychopharmacology, 22*(2), 176.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry, 158*(11), 1783–1793.

Molenberghs, P., Trautwein, F. M., Böckler, A., Singer, T., & Kanske, P. (2016). Neural correlates of metacognitive ability and of feeling confident: A large-scale fMRI study. *Social Cognitive and Affective Neuroscience, 11*(12), 1942–1951.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502.

Murphy, S. C., von Hippel, W., Dubbs, S. L., Angilletta, M. J., Jr, Wilson, R. S., Trivers, R., & Barlow, F. K. (2015). The role of overconfidence in romantic desirability and competition. *Personality and Social Psychology Bulletin, 41*(8), 1036–1052.

O'Brien, E., & Kardas, M. (2016). The implicit meaning of (my) change. *Journal of Personality and Social Psychology, 111*(6), 882.

O'Mara, E. M., & Gaertner, L. (2017). Does self-enhancement facilitate task performance? *Journal of Experimental Psychology: General, 146*(3), 442–455.

Ohan, J. L., & Johnston, C. (2002). Are the performance overestimates given by boys with ADHD self-protective? *Journal of Clinical Child and Adolescent Psychology, 31*(2), 230–241.

Ohan, J. L., & Johnston, C. (2011). Positive illusions of social competence in girls with and without ADHD. *Journal of Abnormal Child Psychology, 39*(4), 527–539.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology, 129*(3), 257–299.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774.

Raab, M., & Gigerenzer, G. (2005). Intelligence as smart heuristics. In R. J. Sternberg, J. Davidson, & J. Pretz (Eds.). *Cognition and intelligence* (pp. 188–207). Cambridge: Cambridge University Press.

Raven, J. C. (1962). 'Advanced progressive matrices', set II. London: H. K. Lewis (Distributed in the United States by the Psychological Corporation, San Antonio, TX).

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142*(2), 359.

Ronay, R., Oostrom, J. K., Lehmann-Willenbrock, N., & Van Vugt, M. (2017). Pride before the fall:(Over) confidence predicts escalation of public commitment. *Journal of Experimental Social Psychology, 69*, 13–22.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics, 39*(3), 490–500.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*, 532–545.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the big five. *Journal of Research in Personality, 38*(5), 473–480.

Schlehofer, M. M., Thompson, S. C., Ting, S., Ostermann, S., Nierman, A., & Skenderian, J. (2010). Psychological predictors of college students' cell phone use while driving. *Accident Analysis & Prevention, 42*(4), 1107–1112.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 299.

Stahl, J., & Gibbons, H. (2007). Dynamics of response-conflict monitoring and individual differences in response control and behavioral control: An electrophysiological investigation using a stop-signal task. *Clinical Neurophysiology, 118*(3), 581–596.

Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences, 21*(6), 971–986.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161.

Stanton, N. A., Salmon, P. M., & Walker, G. H. (2015). Let the reader decide: A paradigm shift for situation awareness in sociotechnical systems. *Journal of Cognitive Engineering and Decision Making, 9*, 44–50.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychological, 47*(2), 143–148.

Tas, A. C., Luck, S. J., & Hollingworth, A. (2016). The relationship between visual attention and visual working memory encoding: A dissociation between covert and overt orienting. *Journal of Experimental Psychology: Human Perception and Performance, 42*(8), 1121.

Taylor, J. B., Visser, T. A. W., Fueggle, S. N., Bellgrove, M. A., & Fox, A. M. (2018). The error-related negativity (ERN) is an electrophysiological marker of motor impulsiveness on the Barratt Impulsiveness Scale (BIS-11) during adolescence. *Developmental Cognitive Neuroscience, 30*, 77–86.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *Situational awareness in aerospace operations* (pp. 1–17). Neuilly Sur Seine, France: NATO-AGARD.

Tellegen, A. (1982). *Multidimensional personality questionnaire manual.* Minneapolis (MN): University of Minnesota Press.

Tombu, M., & Jolicœur, P. (2005). Testing the predictions of the central capacity sharing model. *Journal of Experimental Psychology: Human Perception and Performance, 31*(4), 790.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505.

Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the psychological refractory period effect? *Journal of Experimental Psychology: Human Perception and Performance, 25*(5), 1268.

Visser, T. A. W., Davis, C., & Ohan, J. L. (2009). When similarity leads to sparing: Probing mechanisms underlying the attentional blink. *Psychological Research PRPF, 73*(3), 327–335.

Wilson, M. D., Farrell, S., Visser, T. A. W., & Loft, S. (2018). Remembering to execute deferred tasks in simulated air traffic control: The impact of interruptions. *Journal of Experimental Psychology: Applied, 24*(3), 360.

Zakay, D., & Glicksohn, J. (1992). Overconfidence in a multiple-choice test and its relationship to achievement. *The Psychological Record, 42*(4), 519–524.