

This is a peer-reviewed, final published version of the following document, Wheatcroft, J. M., Jump, M., Breckell, A. L., & Adams-White, J. (2017). Unmanned aerial systems (UAS) operators' accuracy and confidence of decisions: Professional pilots or video game players?. Cogent Psychology, 4(1), 1327628. <https://doi.org/10.1080/23311908.2017.1327628> and is licensed under Creative Commons: Attribution 4.0 license:

**Wheatcroft, Jacqueline M. ORCID logoORCID:  
<https://orcid.org/0000-0001-7212-1598>, Jump, Michael,  
Breckell, Amy L. and Adams-White, Jade E. (2017) Unmanned  
aerial systems (UAS) operators' accuracy and confidence of  
decisions: professional pilots or video game players? Cogent  
Psychology, 4. Art 132762.  
[doi:10.1080/23311908.2017.1327628](https://doi.org/10.1080/23311908.2017.1327628)**

Official URL: <https://www.cogentoa.com/article/10.1080/23311908.2017.1327628>  
DOI: <http://dx.doi.org/10.1080/23311908.2017.1327628>  
EPrint URI: <https://eprints.glos.ac.uk/id/eprint/6097>

#### **Disclaimer**

The University of Gloucestershire has obtained warranties from all depositors as to their title in the material deposited and as to their right to deposit such material.

The University of Gloucestershire makes no representation or warranties of commercial utility, title, or fitness for a particular purpose or any other warranty, express or implied in respect of any material deposited.

The University of Gloucestershire makes no representation that the use of the materials will not infringe any patent, copyright, trademark or other property or proprietary rights.

The University of Gloucestershire accepts no liability for any infringement of intellectual property rights in any material deposited but will remove such material from public view pending investigation in the event of an allegation of any such infringement.

PLEASE SCROLL DOWN FOR TEXT.



## COGNITIVE SCIENCE & NEUROSCIENCE | RESEARCH ARTICLE

# Unmanned aerial systems (UAS) operators' accuracy and confidence of decisions: Professional pilots or video game players?

Jacqueline M. Wheatcroft, Mike Jump, Amy L. Breckell and Jade Adams-White

*Cogent Psychology* (2017), 4: 1327628



Received: 11 December 2016  
Accepted: 02 May 2017  
First Published: 09 May 2017

\*Corresponding author: Jacqueline M. Wheatcroft, Department of Psychological Sciences, Eleanor Rathbone Building, University of Liverpool, Liverpool L69 7ZA, UK  
E-mail: [jacmw@liverpool.ac.uk](mailto:jacmw@liverpool.ac.uk)

Reviewing editor:  
Peter Walla, University of Newcastle, Australia

Additional information is available at the end of the article

## COGNITIVE SCIENCE & NEUROSCIENCE | RESEARCH ARTICLE

# Unmanned aerial systems (UAS) operators' accuracy and confidence of decisions: Professional pilots or video game players?

Jacqueline M. Wheatcroft<sup>1\*</sup>, Mike Jump<sup>2</sup>, Amy L. Breckell<sup>3</sup> and Jade Adams-White<sup>1</sup>

**Abstract:** Unmanned Aerial Systems (UAS) operations have outpaced current training regimes resulting in a shortage of qualified UAS pilots. Three potential UAS operator groups were explored for suitability (i.e. video game players [VGP]; private pilots; professional pilots) and examined to assess levels of accuracy, confidence and confidence-accuracy judgements (W-S C-A) during a simulated civilian cargo flight. Sixty participants made 21 decision tasks, which varied across three levels of danger/risk. Scales of Tolerance of Ambiguity, Decision Style and NEO-PIR were also completed. Professional pilots and VGPs exhibited the highest level of decision confidence, with VGPs maintaining a constant and positive W-S C-A relationship across decision danger/risk. As decision danger/risk increased, confidence, accuracy and W-S C-A decreased. Decision danger also had a role to play in the confidence expressed when choosing to intervene or rely on automation. Neuroticism was negatively related, and conscientiousness positively related, to confidence. Intolerance of ambiguity was negatively related to W-S C-A. All groups showed higher levels of decision confidence in decisions controlled by the UAS in comparison to decisions where the operator manually intervened. VGPs display less overconfidence in decision judgements. Findings support the idea that VGPs could be considered a resource in UAS operation.

### ABOUT THE AUTHOR

Jacqueline M. Wheatcroft is a Chartered and Forensic Psychologist in the Institute for Psychology, Health & Society at the University of Liverpool, UK and is Chair of the British Psychological Society Division of Forensic Psychology Training Committee. Her research interests are in the enhancement of information, intelligence and evidence with a focus on process and procedural techniques to increase accuracy and appropriate confidence in information forms. She upholds interdisciplinary research and has published widely in areas that relate to security, law enforcement and legal process. Her work has contributed to governments, police, professional investigators and the courts.

### PUBLIC INTEREST STATEMENT

The move to significant automation has been a feature of aviation over the last 40 years. This paper describes three potential unmanned aerial supervisor (UAS) groups; video game players, private pilots and professional pilots who made 21 decisions across three levels of dangerousness. As danger increased levels of confidence, accuracy and the relationship between how accurate the decision was and the level of confidence applied to those decisions decreased. The dangerousness of the decision also affected how confident participants were when choosing to intervene or rely on the automation; confidence was lower when the operator chose to intervene. Understanding which potential supervisory group has the best skills to make the best decisions can help to improve UAS supervision. Overall, video game players were less overconfident in their decision judgements. The outcome supports the idea that this group could be a useful resource in UAS operation.

**Subjects: Engineering & Technology; Social Sciences; Behavioral Sciences**

**Keywords: human factors; automation; decision-making; confidence; accuracy**

## 1. Introduction

Automation is the allocation of functions to machines that have in the past been executed by humans; the term is also used to refer to machines that perform, partially or fully, those functions (Funk et al., 1999). The move to significant automation has been a feature of aviation over the last 40 years. As such, the term automation captures a complex blend of technology interacting with human operators and carrying out a wide range of tasks (Civil Aviation Authority, 2016). From the removal of the flight engineer from the cockpit whose function is now carried out by sophisticated full authority digital engine control (FADEC) computers to the advanced stabilisation, guidance and navigation functions of modern aircraft flight control systems (FCS), the role of the crew in the cockpit has been transformed from “seat-of-the-pants” aviators to being the monitors of those systems to check that they are functioning correctly. This function is considered to be so important that it warranted a study of its own by the UK regulator and airline industries (Civil Aviation Authority, 2013). The rapid development of automated technologies has moved the world of work and systems into emergent automation innovation challenges. The introduction of automation, to what are now referred to as “glass cockpits”, provides numerous benefits, including increased vehicle trajectory precision (Murnaw, Sarter, & Wickens, 2001) and reduced crew workload. These benefits mainly manifest themselves in tasks that do not require the crew to be involved. However, when collaboration and cooperation between the crew and the automated system is required, problems can occur (Woods & Sarter, 2000). One key issue is that crews can become confused about the state and/or behaviour of the automation (Sarter & Woods, 1994, 1995). This can have fatal consequences, as demonstrated in the AF447 disaster when three highly trained pilots were unable to identify that their aircraft was in a stall condition—a basic skill taught at the earliest stages of pilot training—at least partially because of the information from the aircraft’s systems available to them (Bureau d’Enquêtes et d’Analyses, 2012). Crews can also become complacent about the ability of automation and fail to detect failures in the automatic systems (Parasuraman, Molloy, & Singh, 2009).

The challenges are most acute where the correct functioning of automatic systems is safety-critical. For example, many of the basic in-flight functions are carried out automatically with the crew monitoring the health of these functions, and ready to step in should any of them be identified as not performing correctly. The challenge of carrying out this supervisory task increases manifold when the crew are removed from the cockpit. The human loses vital sensory information (for example, engine pitch can be used as a surrogate for engine health or even thrust demand), whilst the aircraft loses a very powerful sensor and information processor. However, this is the very situation that unmanned aerial system (UAS) operators find themselves in when supervising typical missions for these vehicles. The supervisory task and assessment of the suitability of potential UAS operators thereby forms the basis of this paper.

A UAS can be defined as a powered vehicle that does not carry a human operator, can be operated autonomously or remotely, can be expendable or recoverable and can carry a lethal or nonlethal payload (Department of Defense, 2007). Since the 1970s aviation automation technology has proliferated. This has undoubtedly contributed to the continued excellent safety record enjoyed by air travel. However, with new technologies emerge new problems. For example, there has been a corresponding increase in errors caused by human–automation interaction; that is, human error (Prinzel, DeVries, Freeman, & Mikulka, 2001).

It has been recognised that UASs, and in particular, those that have the capability to make certain high-order decisions independently (this agent will be referred to hereafter as “The Executive”), can reduce life cycle cost and serve as a force multiplier within the military and civilian world (Ruff, Calhoun, Draper, Fontejon, & Guilfoos, 2004). The success and growth in the use of automation and UASs does not eliminate humans from the system—instead it transforms the human role from

operator to supervisor. Such transformation means that the workload of the human supervisor is not necessarily reduced but instead requires cognitive resource and skills to be applied across a different set of tasks. For example, anticipating and understanding the automation (Walliser, 2011) to ensure the UAS is free from errors and effectively take control of malfunctions, if necessary (Ross, Szalma, Hancock, Barnett, & Taylor, 2008). Supervisors are thereby responsible for the allocation of functions between automatic and manual control and whether the supervisor chooses to control the system automatically or manually can have an impact on the performance of the system (Lee & Moray, 1992). Human interaction is thus an integral part of UAS operations as part of the human-machine cooperative (Drury & Scott, 2008). However, it is anticipated that the benefits gained through the use of UASs can be increased through reductions in supervisor-system ratio and multiple UASs monitored by one supervisor (Ruff et al., 2004). Automated systems coupled with the desire to operate them in ever-greater numbers with fewer supervisors demands close examination of the human-machine relationship (Cring & Lenfestey, 2009).

One factor that may influence the efficacy of supervision is that of trust; perhaps one of the most important factors that enables automated systems to be used to their full potential (Lee & Moray, 1992). Trust in automation has been defined as the extent to which the supervisor is confident in and accurately willing to act on the basis of the recommendations, actions and decision of an artificially intelligent agent (the UAS Executive; Madsen & Gregor, 2000). Furthermore, trust has been characterised as “the attitude that an agent (the UAS Executive) will help achieve an individual’s (supervisor’s) goals in a situation characterised by uncertainty” (Lee & See, 2004, p. 51). This can be explained by the suggestion that a relationship embodied in trust leads to the effective use of resources, efficient cooperation and improved communication (Tajfel & Turner, 1986), while distrust produces an opposite conceptual framework (Toma & Butera, 2009; see also Tversky & Kahneman, 1974).

Yet, what is discussed less is the necessary concordance between a supervisor’s own level of trust and confidence in decisions as they relate to accuracy. We will return to this point later. Ultimately, the capabilities and limitations of the UAS need to be understood in order that supervisors can effectively recognise and intervene when automation capabilities have been exceeded (Cring & Lenfestey, 2009). Accordingly, Lee and See’s (2004) Appropriate Trust Framework states that trust calibration is essential for achieving appropriate dependence (i.e. where trust calibration refers to the match between the supervisor’s level of trust in the automation and the automated aid’s capabilities). If a supervisor’s trust does not equal the true capabilities of the system then this may result in difficulties. For example, (a) in misuse (e.g. using it when it should not be used), (b) in an overreliance on the automation (e.g. paying less attention to important information) or (c) in disuse, such as the underuse of automation (e.g. ignoring alarms, turning off automated safety systems; Parasuraman & Riley, 1997).

The fatal consequences of the misuse of automation are evident from the crash of Eastern Flight 401 in the Florida Everglades—due to the crew’s failure to notice the disengagement of the autopilot and their poor monitoring of the aircraft’s altitude (National Transportation Safety Board, 1973). Similarly, the consequences of the disuse of automation can be observed in the crash of Air France AF447 in 2009—due to pilot error when the automatic Stall Warner was ignored because of conflicting air speed readings due to icing of the aircraft’s air data system (Martins & Soares, 2012). These catastrophic situations demonstrate the importance of the operator’s need to have appropriate confidence and trust in the automation available to them. However, these examples relate to accidents when the pilots were on board the aircraft. The issue becomes more relevant when a UAS is involved as the supervisor lacks the proprioceptive cues available to pilots of manned aircraft (e.g. changes in engine noise or vibration that can indicate possible engine malfunctions, Drury & Scott, 2008). Indeed, Tvaryanas, Thompson, and Constable (2005) have showed a significant number ( $n = 271$ ) of UAS mishaps have occurred in the last decade due to human factors. Further, an analysis of 16 UAS accidents by Glussich and Histon (2010) showed, in many instances, common human deficiencies directly contributed to the loss of control of the automation and the aircraft.



The framework of automation use by Dzindolet, Beck, Pierce, and Dawe (2001) predicts that cognitive, motivational and social processes work together to cause misuse, disuse and inappropriate trust in automation and, indeed, many factors may affect each of these processes impacting upon automation use. When forming trust judgements, supervisors of automated functions compare the perceived reliability of the automated aid to the reliability of manual operation in order to determine the perceived utility of the aid and the level of automation trust. If the perceived utility of the aid is high, trust in the automation is likely to be high and dependence on the automation expected. Conversely, if the perceived utility is low, trust will also be low and so self-reliance expected. Cognitive biases can impact upon the use of automation. For example, the number of tasks to be performed, intrinsic interest in the task, cognitive overhead, penalties for failure and rewards for completion, and so on, will affect the effort a supervisor will expend on any task and the likelihood of reliance on the automated aid. However, Lee and See (2004) found that high levels of trust in automation do not always result in misuse as long as the trust is appropriate. In support of this, individuals with high levels of trust in automation were more successful at detecting automation failure than those with low levels of trust. Furthermore, the self-confidence of a supervisor significantly influences how they interact with automation and the degree of trust instilled in it (Lee & Moray, 1992; Riley, 1994; Will, 1991). Individual's use and trust automation more when their confidence in own ability is lower than in automation, and vice-versa (Lee & Moray, 1992; Riley, 1994). Thus, biases in self-confidence can have a substantial effect on the appropriate reliance on automation (Lee & See, 2004). That reliance may also be influenced by the degree of confidence one has in the automation and thus some research demonstrates that individuals can tend to over rely on automation (Parasuraman & Manzey, 2010).

Automation bias, as it is termed, occurs when there is overconfidence in the automation system. It has been defined by Mosier and Skitka (1996) as "a heuristic replacement for vigilant information seeking and processing" (p. 205). This tendency to over rely on automation can negatively impact on decision-making. For example, supervisors are likely to approve system decisions even when the system providing the information is unreliable (Cummings, 2004). Three main reasons for the occurrence of automation bias have been highlighted in the literature (Mosier & Skitka, 1996; Parasuraman & Manzey, 2010). First, automation may be deemed less cognitively demanding thus being a preferred choice, as individuals tend to opt for the option of least effort (a cognitive miser effect—Fiske & Taylor, 1991). Second, individuals tend to overestimate of the correctness of automation viewing it as holding superior knowledge to that of their own. For instance, information from automated systems has been rated as more accurate than information provided by humans (Dijkstra, Liebrand, & Timminga, 1998), and supported by research which suggests those with more expertise are less likely to rely on automation (Sanchez, Rogers, Fisk, & Rovira, 2014). Third, individuals may view automation as a diffusion of responsibility resulting in feeling less accountable for the decision (Latane & Darley, 1970). Indeed, Skitka, Mosier, and Burdick (2000) found that increasing accountability reduced instances of automation bias. Hence, as supervisors need to be able to correctly allocate between automated and manual functions (Ross et al., 2008) it would be beneficial to examine what factors influence overconfidence in automation.

It has been suggested that the effect of self-confidence and reliance on automation (i.e. increased confidence) can be moderated by both the skill level of the supervisor and risk associated with the decision to use or not to use automation (Riley, 1994). For instance, experience may impact on how much confidence is placed on a decision. Indeed, Riley, Lyall, and Weiner (1993) found that pilots rely on automation more than novices. Further, decision confidence may also depend on the danger or risk associated with the decision. However, research regarding whether individuals rely on automation more or less with increased risk is mixed. For example, when the risk is low individuals show more confidence in automation, but when the risk is high individuals tend not to rely on automation, suggesting a reduction in confidence in automation when greater risk is involved (Perkins, Miller, Hashemi, & Burns, 2010). However, Lyons and Stokes (2011), where supervisors were provided with the option to use either a human aid or automated tool for decision-making, found that in conditions of high risk the human aid was relied on less, demonstrating a preference to the automated aid in high risk circumstances. Hence, confidence in automation may well vary according to associated

risk. It is necessary therefore that both confidence in automated and manual decisions and self-confidence of potential supervisors of automation be evaluated.

Currently, a wide range of individuals can legally operate a UAS. These range from professional pilots (e.g. Royal Air Force) to enlisted men (e.g. US Marine Corps) to private individuals (e.g. those who qualify for a UK Basic National UAS Certificate (BNUC-S) which allows them to fly aircraft up to 20 kg maximum take-off mass (MTOM) within visual line of sight (VLOS). Certification can vary however depending upon classification. For example, larger systems such as Predator/Reaper or Global Hawk require formal training courses in UAS operations, tactical and theatre operations, battlespace awareness, threats, weapons and sensors. Smaller systems tend to perform less complex missions and require less formal training. However, the tempo of UAS operations, at least for larger (generally military at the time of writing) vehicles has now outpaced current supervisor training regimes resulting in a shortage of qualified UAS pilots. Surrogates need to be found to replace the use of manned aircraft pilots as UAS supervisors; preferably recruits who would learn faster and be easier to train, to accelerate supervisor training, to meet these new and pressing requirements (McKinley & McIntire, 2009). Indeed, the US Air Force has adopted aptitude requirements and a training syllabus (Undergraduate RPA Training or URT) for UAS pilot trainees with little or no prior flying experience (see Carretta, 2013; Rose, Barron, Carretta, Arnold, & Howse, 2014).

Nevertheless, it is possible that the ground control stations of UASs can be compared to traditional video game environments. This comparison can be made in the sense that, in a video game, the player is trying to achieve some goal (the aircraft mission) and interacts with the game via screens and inceptors that provide sufficient but limited information to allow this to happen (the aircraft sensor feed, displays and controllers). Thus, it is plausible to investigate whether video game experience and skills can be of particular benefit to UAS supervision. Indeed, video game players (VGP) who have no piloting experience may well be better suited to the role of UAS supervisor as these individuals will tend not to base aviation decisions on proprioceptive cues available to pilots of manned aircraft (McKinley & McIntire, 2009). Plus, VGPs are argued to be able to track more targets (Castel, Pratt, & Drummond, 2005), have improved psychomotor skills (Griffith, Voloschin, Gibb, & Bailey, 1983), quicker reaction times (Yuji, 1996) and enhanced spatial skills (Dorval & Pepin, 1986). Importantly, many studies have found the skills, abilities and other characteristics (SAOCs) of VGPs transfer to other cognitive tasks (Gopher, Weil, & Bareket, 1994; Green & Bavelier, 2007) but may not have the tactical and/or operational awareness.

McKinley and McIntire (2009) compared VGPs, professional pilots and a control group that had no gaming or pilot experience on UAS cognitive tasks. It was found that VGPs and professional pilots did not significantly differ but both were superior when compared to the control group in aircraft control and landing skills. These findings suggest that VGPs possess skills that have direct application to UAS supervision, and that VGPs and professional pilots possess some skills relevant to the supervision of UASs. However, more research is needed to consolidate these outcomes and across other measures. For example, it is important that self-confidence in decisions across decision-risk categories is associated with accurate responses made. This paper aims to assess these measures across a range of potential UAS supervisors.

In addition, and in order to identify suitable recruits for supervisory roles, it is beneficial to look at various typologies of potential agents. As noted, there has been some research which investigates the relationship different groups of potential supervisors have with the autonomous system operating the UAS with regard to levels of trust, what affects trust (Lee & See, 2004; Ruff et al., 2004) and abilities to effectively supervise a UAS (McKinley & McIntire, 2009). However, an extensive literature search has found a vacuum of research, focused on the supervisor, which is concerned with own supervisory levels of confidence and accuracy across potential UAS groups relative to decisions made. With this in mind, the present research focuses on four different groups of potential UAS supervisors' confidence and accuracy across risk decision, including some comparison to broad psychological constructs. As Riley (1994) suggests, the four different groups distinguished by their skill

levels in aviation can have an impact on a supervisory confidence and accuracy. The four groups examined by this research are, (a) *a control group*; individuals with no gaming or pilot experience, (b) *VGP*; such individuals have been shown to possess cognitive abilities necessary for supervising a UAS, (c) *private pilot*; individuals who hold a private pilot's licence and (d) *professional pilot*; individuals who are either instructors, commercial airline or military pilots.

The work here raises the notion that supervisor confidence is conceptually different to that work conducted on trust in this context; confidence here is a qualifier, which is associated with a particular decision. An individual who makes a decision associates this with a level of certainty (decision confidence) which arises as a result of specific knowledge with the decision built on reasoning; it is not synonymous with trust which is largely based on intuitions (Muir, 1994; Shaw, 1997). This means an individual makes an evaluation of decisions and reports a level of confidence in those decisions that, ideally, correlate with correct performance (i.e. accuracy). Such a correlation is known as the within-subject confidence and accuracy (W-S C-A) relationship; measure of metacognitive sensitivity that enables the expression of individual confidence in correct or incorrect responses (Wheatcroft & Ellison, 2012; Wheatcroft & Woods, 2010; Yeung & Summerfield, 2012). The measure assists the gauging of individual decision-making across a course of actions and so is very important where complexity may exist. Moreover, confidence has been related to decision success (i.e. increased accuracy; Bingi & Kasper, 1995), and overconfidence in wrong decisions can result in inappropriate, perhaps fatal, action. Adidam and Bingi (2000) note that if an individual has more confidence in their decisions they tend to allocate more resources (i.e. cognitive ability) into implementing the decision; though this work cannot necessarily be generalised to aerial settings. Nevertheless, pilots may have greater metacognitive sensitivity (W-S C-A) to supervise a UAS than non-pilots because skill levels have been shown to positively affect confidence and accuracy (Riley, 1994). However, it has also been suggested that simulation training (i.e. playing video games) can increase confidence in decision-making (Atinaja-Faller, Quigley, Banichoo, Tsveybel, & Quigley, 2010) implying VGPs may also exhibit high W-S C-A. This research will determine which potential UAS supervisory group is most metacognitively confidence-accuracy sensitive, and moreover, across decision risk categories.

On the note of decision danger/risk, research has suggested the difficulty of a decision task can influence confidence and accuracy; the easier a task, the greater the concordance between confidence and accuracy, and vice versa (Kebbell, Wagstaff, & Covey, 1996; Wheatcroft, Wagstaff, & Manarin, 2015). Decision difficulty, in the context of UAS supervision, can be induced by varying the potential danger/risk of the decision needed to be made—relevant given that decisions carrying dangerous implication can be more difficult to make (Riley, 1994). Thus, decision danger/risk may reduce individual confidence and affect decision-making variably across the potential UAS groups; whilst overconfidence can lead to risky (Krueger & Dickson, 1994); or inaccurate decisions (Wheatcroft, Wagstaff, & Kebbell, 2004). Potential thereby exists for groups to be highly confident and wrong.

Finally, it is also useful to explore personality typology of individuals across the potential groups as certain groups may respond in particular ways due to stable deep-seated predispositions (Chidester, Helmreich, Gregorich, & Geis, 1991). For example, those who have higher levels of ambiguity tolerance are more decisive and display greater confidence in choices (Ghosh & Ray, 1997; Maddux & Volkmann, 2010). The NEO-PIR is a general measure of five factors of personality (i.e. neuroticism, extraversion, openness to experiences, agreeableness and conscientiousness; Costa & McCrae, 1992; McCrae, & Costa, 1992). Measures of conscientiousness, openness to experience and agreeableness from the NEO-PIR have been shown to be positively related to pilot performance of manned aircraft (Barrick & Mount, 1991; Burke, 1995; Chidester, Kanki, Foushee, Dickinson, & Bowles, 1990; Siem & Murray, 1994) but as yet no work has considered these factors in a UAS context, as described, nor across group type. Research suggests that the five factors of the NEO-PIR have implication for aviation these may have importance to UAS decision confidence and accuracy.



## 2. Rationale and aims

Potential UAS group supervisor's confidence, accuracy and W-S C-A in decisions related to personality constructs, together with the examination of the impact of decision danger is of critical importance in this context; particularly as concordant confidence and accuracy is relevant to high-performance successful decisions and vice versa.

The current study therefore explores accuracy, confidence and W-S C-A relationship across groups (control, VGP, private, and professional pilots) to identify potential UAS supervisor's levels on these measures and in relation to decision style. In addition, the decisions made will likely vary by the decision danger/risk to reveal the impact that increased danger has on decision confidence and accuracy. The varied decision danger is designed to induce decision difficulty, which previous research has found to negatively impact confidence and accuracy (Kebbell et al., 1996; Wheatcroft et al., 2015). The present study will also assess whether groups differ in decision confidence applied to manual decisions in comparison to automated decisions. Further, group personality traits will be assessed via standardised tools related to confidence, accuracy and W-S C-A. Insight into associations with measures will help to determine traits useful for UAS supervision.

In the light of the exploratory nature of some of the above, two key non-directional hypotheses were formulated: (i) levels of confidence, accuracy and W-S C-A will differ across groups and levels of decision danger; and (ii) psychometric tests will reveal personality characteristics that differ across groups and in relation to confidence, accuracy and the W-S C-A relationship.

Importantly, two key directional hypotheses were also expressed: (iii) as decision danger increases, there will be a significant decrease in confidence, accuracy and W-S C-A across the groups; and (iv) decision confidence in automated and manual decisions will differ across groups and be negatively impacted by increased decision danger.

## 3. Method

### 3.1. Participants

The sample consisted of four different groups (control, video game players (VGP), private, and professional pilots) each made up of 15 participants (i.e. 60 participants in total; 51 male, 9 female). The minimum age of any participant was 17 as this is the minimum legal age a person can hold a pilot's licence in the UK. There was no maximum age because holding a pilot license is determined by the ability to pass an appropriate medical (Civil Aviation Authority, 2010). The Control group consisted of participants with no gaming or pilot experience, recruited via the University of Liverpool ( $M = 39.4$ ,  $SD = 18.8$ ). The VGP group was recruited via the University of Liverpool ( $M = 21.7$ ,  $SD = 2.9$ ). The private pilot licence group was recruited from flying clubs in North West England ( $M = 45.1$ ,  $SD = 16.1$ ). The professional pilot group was identified in one of three ways by; (1) an airline transport pilots licence (ATPL), (2) an instructor rating or (3) a military pilot recruited from various established flying institutions around North West England ( $M = 46$ ,  $SD = 13.4$ ). Opportunity sampling was employed. Any potential for representation bias and motivated responses was reduced by targeting a defined population and sample frame matched as keenly as possible. Participants responses were kept confidential and were only identified by a number on their consent form and answer sheets.

### 3.2. Design

The independent variables are represented by the UAS interact groups and the level of potential danger of the decision (i.e. decision danger). The research is independently structured via a 4 (UAS Group: Control/VGP/Private Pilot/Professional Pilot)  $\times$  3 (Level of Decision: Low/Medium/High) design. The dependent variables are measured as confidence, accuracy, within-subjects confidence-accuracy (W-S C-A), correlation scores and psychometric scores.

### 3.3. Materials

A set of demographic questions asked participant sex and age. A Tolerance of Ambiguity questionnaire was used to assess individual tolerance of ambiguity (Budner, 1961) and Decision Style was used (Need for Closure; Roets & Van Hiel, 2007). The NEO-PIR is a five-factor model of personality and consists of assessments of the major factors (i.e. Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness). The NEO-PIR has strong support for reliability, construct and discriminant validity (see Costa & McCrae, 1992; Piedmont & Weinstein, 1993).

To provide the participants with as realistic a scenario as possible to allow them to make decisions, a series of pre-recorded video vignettes of typical scenarios that might be encountered during a typical flight were recorded. To ensure ecological validity, this environment was modelled such that it contained information and displays that are typical of current UAS supervisory environments, with additional display elements to convey decisions being made by the Executive Agent (see Webster, Cameron, Fisher, & Jump, 2014). The supervisor's station information was displayed on a visual screen with four individual display units that showed:

- (1) *An external view of the simulated (virtual) outside world.* This emulated the view from a forward-looking camera mounted on the UAS in a good visual environment.
- (2) *A moving map display.* This showed a real-time indication of the UASs world plan-position and the current route planned by the Executive Agent.
- (3) *The "Basic Six" flight instruments.* These instruments provide pilots of manned aircraft with the essential information required to conduct a flight. They are arranged in a standard configuration of 2 rows  $\times$  3 columns. The first row, moving left-to-right comprises: air speed indicator (ASI); attitude direction indicator (ADI) and altimeter. The second row moving left-to-right comprises: turn and slip indicator; horizontal situation indicator (HSI) and vertical speed indicator (VSI).
- (4) *The Aircraft Information Panel.* Removing the pilot from the aircraft deprives the aircraft of a useful sensory system but also deprives the pilot of a number of valuable sensory cues that can be used to make decisions (engine noise, vibration etc.). This panel provided some limited information on the state of the aircraft (control surface positions, fuel remaining etc.) plus information concerning the communication status between the aircraft's Executive Agent and the relevant air traffic control (ATC) station. A number of these information messages were colour-coded to indicate the urgency with which they should be attended to (red = immediate action; orange = prepare to take action; green = no action required).

Figure 1 shows the standard set of screens used to create the vignette videos.

The aircraft flight dynamics model was created using the multi-body dynamics software FLIGHTLAB [3] and was configured to be representative of a small general-aviation trainer aircraft. A piece of code was written to make the outputs of this model drive the visualisation of the outside world using the Microsoft FSX gaming software as the display engine. The other three displays were generated using Presagis' industry-standard VAPS (Virtual Avionics Prototyping Software) display creation software, with an additional specific piece of interface code being written to make the displays respond in the appropriate manner to the aircraft flight dynamic model's outputs. A Flight Briefing was drawn up to explain to participants what a UAS is, the overall mission of the flight, the goals of safety and performance, the capabilities and constraints of a UAS, and the standard operational procedures in aviation. An event and decision log was developed and the strongest answer for all 21 decisions was identified and verified by two UAS pilot experts (Cronbach's Alpha = 1). The flight was separated into seven phases: taxi (the air vehicle manoeuvres on the ground to reach the runway threshold); take off (the air vehicle manoeuvres on the ground to line-up with the runway centre-line, accelerates to a particular speed, rotates and becomes airborne); climb out (the vehicle achieves the desired climb attitude, heading and speed and continues in this mode until the desired cruise altitude is reached);

**Figure 1.** Standard set of screens used to create the vignette videos.



cruise (the vehicle achieves the straight and level flight attitude and speed and follows heading autopilot commands to follow the pre-planned route); descent (in the vicinity of the destination airport, the vehicle begins a descent from the cruise altitude to achieve an altitude and inertial position that is suitable to begin the approach to the runway); approach (the air vehicle achieves the desired approach speed and configuration and lines up with the extended runway centre-line. It flies a heading such that it remains in this alignment and descends at an appropriate rate such that it is at approximately 50ft above ground level as it crosses the runway threshold); and landing (from the so-called 50ft “screen height”, the air vehicle rate of descent is reduced such that it comes into contact with the runway surface at an acceptably low rate of descent). Each phase included three events that could occur across the flight which varied in potential danger/risk (low, medium, high) of the decision. As such, there were 21 events that required a decision to be made to each. Every event had three options to choose from. There always existed the option to (a) allow the autonomous system to control the UAS and (b) to intervene and manually fly the vehicle. This increased the ecological validity of the experiment because, as with field operators, those supervisors taking part needed to balance the competing goals of safety and performance using automatic or manual control. The event and decision log provided a baseline measurement tool against which the research measurements (confidence, accuracy and W-S C-A) could be scored. Such methods and measures have been used successfully in previous research (Wheatcroft & Ellison, 2012; Wheatcroft & Woods, 2010). A Likert scale (ranging from 1–not at all confident at all to 10–extremely confident) was used to record participant confidence in each of the decisions made.

### 3.4. Procedure

The investigation was approved by the University of Liverpool ethics committee. Participants were assigned to the appropriate group as defined by the criteria. The experiment was conducted with mixed groups of no more than 20 participants being tested at any one time. Participants were seated separately from each other in front of a projector and instructed not to confer. Participants were then given the information sheet to read, asked if they had any questions and once satisfied they had an understanding of the experiment signed the consent form to proceed. Participants first completed the Demographic, Tolerance of Ambiguity, Decision Style and NEO-PIR questionnaires. All participants were given the opportunity to practise a short flight to familiarise them to the requirements of the study. For the main experiment, participants were given the Flight Briefing to read and instructed to supervise a UAS on a civilian cargo flight from A to B (Liverpool John Lennon International Airport to Blackpool International Airport) that was shown using the 33-min long vignette. This route was chosen as it was short enough to conduct the experiment several times whilst exposing the participants to, for example, several different kinds of airspace classification and scenarios that might be representative of a longer duration flight. During the flight procedure, 21 events required the participant to make a key decision. As each event arose, the flight was paused and the participants had 45 s to choose from one of the three options presented to them. They had to select the one which they believed to be the best decision that met the terms of the briefing and then rate how confident they were in that decision. The simulation attempted as far as possible to mirror the context of decision-making required in this setting. Once the decision-making time had elapsed, the flight sequence was re-started from the paused condition until the next event occurred. This process was repeated until the flight had come to an end. Once completed participants were debriefed fully to ensure they knew what they had taken part in and given the opportunity to ask questions and reminded of their right to withdraw data anytime. Ethical protocols were followed at all stages during the study.

## 4. Results

Participants' psychometric test scores, overall accuracy (i.e. the number of decisions made correctly) and confidence scores for the event and decision logs were calculated.

### 4.1. Psychometric data

First, one-way ANOVAs were conducted on the psychometric data to assess and compare each group on the psychometric measures (see Table 1).

**Table 1. Means and standard deviations of psychometric measures across UAS group**

Psychometric	UAS group			
	Control	VGP	Private pilot	Professional pilot
Ambiguity tolerance A	50.13 (5.22)	49.13 (6.69)	48.60 (9.36)	48.13 (6.65)
Decision style	56.93 (11.42)	54.60 (10.49)	55.40 (7.82)	48.93 (8.62)
Ambiguity tolerance B	35.67 (6.76)	33.87 (5.72)	35.27 (4.65)	31.73 (5.66)
Decisiveness	21.27 (5.39)	20.73 (5.61)	20.13 (4.16)	17.20 (3.71)
Neuroticism	23.20* (8.69)	19.67 (6.47)	20.87 (8.13)	15.40* (4.84)
Extraversion	29.53 (8.23)	33.00 (6.61)	28.47 (5.73)	31.80 (6.34)
Open to experiences	32.80 (6.84)	29.60 (7.04)	28.80 (5.85)	28.73 (5.70)
Agreeableness	30.07 (5.41)	28.13* (4.93)	30.07 (5.01)	33.80* (5.62)
Conscientiousness	32.33 (6.60)	31.00 (4.75)	33.07 (7.05)	36.67 (4.75)

Note: Standard deviations are shown in parenthesis.

\* $p < .05$ .

There was a significant effect for neuroticism,  $F(3, 56) = 3.10, p = .034, p < .05$ , and agreeableness,  $F(3, 56) = 3.06, p = .035, p < .05$ . Neuroticism was lower for professional pilots ( $M = 15.40, SD = 4.84$ ) than for controls ( $M = 23.20, SD = 8.69$ ),  $p = .022, p < .05$ . No other comparisons were found to be significant for neuroticism ( $p > .05$ ). Agreeableness however was higher for professional pilots ( $M = 33.80, SD = 5.62$ ) compared to VGPs ( $M = 28.13, SD = 4.93$ ),  $p = .023, p < .05$ . No other comparisons for agreeableness were significant ( $p > .05$ ). No other effects were observed, for example, for tolerance of ambiguity A  $F(3, 56) = .22, p > .05$ ; decision style  $F(3, 56) = 1.95, p > .05$ ; tolerance of ambiguity B  $F(3, 56) = 1.43, p > .05$ ; decisiveness  $F(3, 56) = 2.16, p > .05$ ; extraversion  $F(3, 56) = 1.39, p > .05$ , openness to experiences  $F(3, 56) = 1.36, p > .05$  and conscientiousness  $F(3, 56) = 2.55, p > .05$ .

In order to establish if the psychometric scores were related to the accuracy, confidence and W-S C-A Pearson's correlations were performed. No significant relationships between the psychometric data and accuracy were found ( $p > .05$ ).

There was however a significant moderate negative relationship shown between neuroticism and confidence ( $r = -.415, p = .000, p < .001$ ); as neuroticism increases confidence decreases. A significant moderate positive relationship between conscientiousness and confidence ( $r = .374, p = .003, p < .01$ ) was also found; as conscientiousness increases, so does confidence. Further, a significant weak negative relationship between tolerance of ambiguity A and W-S C-A was found ( $r = -.300, p = .019, p < .02$ ); as tolerance of ambiguity A score decreases (i.e. greater tolerance to ambiguity) the W-S C-A relationship increases.

#### 4.2. Accuracy data

As above a  $4 \times 3$  repeated measures ANOVA was conducted on the data to analyse the effect of decision danger and UAS group on accuracy (see Table 2).

A 4 (UAS Group: control/VGP/private pilot/professional pilot)  $\times$  3 (Decision Danger: Low/Medium/High) repeated measures ANOVA was conducted to analyse the effect of decision danger and UAS group on accuracy.

A main effect of decision danger on accuracy was observed,  $F(2, 112) = 13.66, p = .000, p < .001$ ;  $\eta_p^2 = .20$ .  $F$  comparisons showed accuracy was higher for low ( $M = 5.10, SD = 1.41$ ) than for high ( $M = 4.32, SD = 1.59$ ) decision danger,  $p = .000, p < .001$ . Accuracy was also higher for medium ( $M = 5.33, SD = 1.00$ ) than for high ( $M = 4.32, SD = 1.59$ ) decision danger,  $p = .000, p < .001$ . However, no difference was observed between low and medium decision danger accuracy,  $p = .230, p > .05$ .

There was no main effect of UAS group on accuracy,  $F(3, 36) = 1.14, p = .240, p > .05$ , and no interaction of decision danger and group was observed,  $F(6, 112) = 1.13, p = .349, p > .05$ .

**Table 2. Means and standard deviations for overall accuracy and decision danger accuracy**

UAS group	Accuracy and decision danger (DD)			
	Overall accuracy	Low DD	Medium DD	High DD
Control	13.47 (3.83)	4.27 (1.83)	5.20 (1.15)	4.13 (1.69)
VGP	14.80 (2.15)	5.40 (.83)	5.20 (.862)	4.13 (1.55)
Private pilot	15.20 (3.41)	5.27 (1.44)	5.60 (1.24)	4.33 (1.80)
Professional pilot	15.67 (2.50)	5.47 (1.13)	5.33 (.72)	4.67 (1.40)
Total	14.78 (3.08)	5.10** (1.41)	5.33** (1.00)	4.32** (1.59)

Note: Standard deviations are shown in parenthesis.

\*\* $p < .001$ .



#### 4.3. Confidence data

As before, a  $4 \times 3$  repeated measures ANOVA was performed to examine the effect of decision danger and UAS group on confidence (see Table 3).

There was a main effect of decision danger on confidence  $F(2, 112) = 22.47, p = .000, p < .001; \eta_p^2 = .29$ .  $F$  comparisons showed confidence was higher for low ( $M = 56.55, SD = 9.55$ ) than for high ( $M = 52.37, SD = 9.88$ ) decision danger,  $p = .000, p < .001$ . Confidence was also higher for medium ( $M = 55.98, SD = 8.89$ ) as opposed to high ( $M = 52.37, SD = 9.88$ ) decision danger,  $p = .000, p < .001$ . No difference existed between low and medium decision danger confidence,  $p = .308, p > .05$ .

A main effect of UAS group was found for confidence  $F(3, 56) = 5.48, p = .018, p < .02; \eta_p^2 = .28$ .  $F$  comparisons showed confidence was higher for professional pilots ( $M = 180.47, SD = 18.24$ ) than for controls ( $M = 147.27, SD = 33.66$ ),  $p = .017, p < .02$ . Confidence was also higher for VGP's ( $M = 172.53, SD = 24.72$ ) than for controls ( $M = 147.27, SD = 33.66$ ),  $p = .031, p < .05$ . No other comparisons were found to be significant for overall confidence ( $p > .05$ ).

No interaction was observed between decision danger and interact UAS group on confidence  $F(6, 112) = 1.40, p = .237, p > .05$ .

#### 4.4. W-S C-A data

To establish if there were any significant effects of UAS group (control, VGP, private pilot, professional pilot) and level of decision danger (low, medium and high) for within-subjects confidence and accuracy (W-S C-A) correlations, it was first necessary to calculate each individual participant's W-S C-A score. The answer to each question was coded as correct or incorrect, and the confidence score for each question was recorded to generate a numerical relationship between confidence and accuracy for each participant (i.e. a point-biserial correlation). Table 4 illustrates.

**Table 3. Means and standard deviations for overall confidence and decision danger confidence**

UAS group	Confidence and decision danger (DD)			
	Overall confidence	Low DD	Medium DD	High DD
Control	147.27** (33.66)	48.93 (11.30)	51.33 (11.90)	47.00 (11.74)
VGP	172.53* (24.72)	59.53 (8.24)	58.60 (7.83)	54.40 (9.70)
Private pilot	159.33 (16.35)	54.93 (6.23)	53.60 (5.54)	50.80 (6.13)
Professional pilot	180.47** (18.24)	62.80 (5.80)	60.40 (6.50)	57.27 (8.85)
Total	164.90 (26.83)	56.55*** (9.55)	55.98*** (8.89)	52.37*** (9.88)

Note: Standard deviations are shown in parenthesis.

\* $p < .05$ .

\*\* $p < .02$ .

\*\*\* $p < .001$ .

**Table 4. Means and standard deviations for overall W-S C-A and decision danger W-S C-A**

UAS group	W-S C-A and decision danger (DD)			
	Overall WS-CA	Low DD	Medium DD	High DD
Control	.23 (.20)	.34 (.38)	.19 (.53)	-.10 (.41)
VGP	.30 (.31)	.21 (.53)	.22 (.46)	.27 (.47)
Private pilot	.34 (.18)	.48 (.40)	.24 (.39)	.09 (.54)
Professional pilot	.26 (.27)	.57 (.40)	.27 (.45)	.04 (.38)
Total	.29 (.24)	.40** (.44)	.23* (.45)	.07** (.46)

Note: Standard deviations are shown in parenthesis.

\* $p < .05$ .

\*\* $p < .001$ .

A further  $4 \times 3$  repeated measures ANOVA was conducted to analyse the effect of decision danger and UAS group on W-S C-A.

There was a main effect of decision danger on W-S C-A,  $F(2, 112) = 7.96, p = .016, p < .02; \eta_p^2 = .12$ .  $F$  comparisons showed W-S C-A was highest for low ( $M = .40, SD = .44$ ) than for high ( $M = .07, SD = .46$ ) decision danger,  $p = .000, p < .001$  and W-S C-A was also higher for medium ( $M = .23, SD = .45$ ) than for high ( $M = .07, SD = .46$ ) decision danger,  $p = .040, p < .05$ . However, no difference was found between medium and low decision danger,  $p = .068, p > .05$ .

There was no main effect of UAS group on W-S C-A,  $F(3, 56) = .99, p = .405, p > .05$ , nor was there any interaction of decision danger and UAS group observed,  $F(6, 112) = 1.33, p = .249, p > .05$ .

#### 4.5. Between-subjects confidence-accuracy (B-S C-A)

In order to establish if confidence scores related to accuracy scores, a between-subjects Pearson's correlation was also conducted. A significant weak positive correlation was observed between confidence and accuracy ( $r = .272, p = .035, p < .05$ ).

#### 4.6. Decision type: Manual vs. automated data

One-way ANOVAs were conducted on decision confidence in manual and automated decisions for both high and low decision danger levels to assess differences between the groups (control, VGP, private pilots, professional pilots; see Table 5).

##### 4.6.1. Manual choice-low decision danger

Significant differences were found between the groups in reported decision confidence in manual decisions made in conditions of low decision danger  $F(3, 56) = 11.385, p = .000, p < .001$ . Post hoc tests found that the VGP group ( $M = 18.5, SD = 2.1$ ) was significantly more confident in their manual decisions than the control ( $M = 14.3, SD = 3.2$ ),  $p = .000, p < .001$ . The professional pilot group ( $M = 19.2, SD = 1.1$ ) was also significantly more confident in their decisions than the control ( $M = 14.3, SD = 3.2$ ),  $p = .000, p < .001$ , and private pilots ( $M = 16.8, SD = 2.9$ ),  $p = .035, p < .05$ . No other comparisons were significant.

##### 4.6.2. Manual choice-high decision danger

Significant differences were found between groups in conditions of high decision danger and decision confidence in manual decisions,  $F(3, 56) = 5.841, p = .002, p < .01$ .  $F$  tests showed that professional pilots ( $M = 17.4, SD = 2.8$ ) were significantly more confident than the control ( $M = 12.1, SD = 4.8$ ),  $p = .000, p < .001$ , and as compared with private pilots ( $M = 14.4, SD = 2.3$ ),  $p = .016, p < .02$ . No other comparisons were significant.

##### 4.6.3. Automated choice-low decision danger

Similarly to manual decision there were significant differences found between groups in their reported decision confidence in automated decision choices in low decision danger conditions  $F(3, 56) = 5.086, p = .004, p < .01$ . Post analysis found that professional pilots ( $M = 43.6, SD = 5.0$ ) were more confident in automated decisions in low decision danger than both control ( $M = 34.6, SD = 8.5$ ),  $p = .009, p < .01$  and private pilots ( $M = 38.1, SD = 4.4$ ),  $p = .019, p < .02$ . No other comparisons were significant.

**Table 5. Means and standard deviations for decision confidence of decision danger and decision choice**

Decision danger (DD)	Decision choice (DC)	Control	VGP	Private	Professional	Total
Low	Manual	14.33 (3.24)	18.47 (2.10)	16.80 (2.88)	19.20 (1.15)	17.20 (3.06)
High	Manual	12.07 (4.82)	15.60 (3.91)	14.40 (2.26)	17.40 (2.79)	14.87 (4.00)
Low	Automatic	34.60 (8.54)	40.40 (7.19)	38.13 (4.42)	43.60 (5.03)	39.18 (7.15)
High	Automatic	34.93 (8.47)	38.80 (7.13)	36.40 (5.12)	39.87 (8.00)	37.50 (7.37)

Note: Standard deviations are shown in parenthesis.

**Table 6. Means and standard deviations for decision confidence in decision danger and decision choice**

Decision danger (DD)	Decision choice (DC)	Mean (SD)	Decision danger (DD)	Decision choice (DC)	Mean (SD)
High	Manual	14.87 (4.00)	High	Automatic	37.50 (7.37)
Low	Manual	17.20 (3.06)	Low	Automatic	39.18 (7.15)
Total manual		16.34 (3.54)	Total automatic		38.34 (6.74)

Note: Standard deviations are shown in parenthesis.

#### 4.6.4. Automated choice-high decision danger

No effects were observed for confidence in automated decision choices in conditions of high decision danger,  $F(3, 56) = 1.417, p = .247, p > .05$ .

To investigate confidence across manual and automated decisions overall (regardless of group), a series of paired *t*-tests were conducted (see Table 6).

First, data were analysed to examine if decision confidence differed in high and low decision danger in automated and manual decisions. A Bonferroni correction was applied,  $p < .01$ .

For automated decisions, participants were more confident in low decision danger ( $M = 39.18, SD = 7.15$ ) than high decision danger ( $M = 37.50, SD = 7.37$ ),  $t(59) = 2.402, p = .019, p < .02$ . However, this was not significant once the correction had been applied. For manual decisions however participants were more confident, despite the correction, in those decisions classed as low decision danger ( $M = 17.20, SD = 3.06$ ) than high danger ( $M = 14.87, SD = 4.00$ ),  $t(59) = 5.633, p = .000, p < .001$ .

In conditions of low decision danger participants applied higher confidence ratings in the decisions that involved use of the autonomous system ( $M = 39.18, SD = 7.15$ ) than manual decisions to intervene ( $M = 17.20, SD = 3.06$ ),  $t(59) = -31.401, p = .000, p < .001$ . Similarly in high decision danger conditions participants were more confident in their decision when agreeing with the autonomous system ( $M = 37.50, SD = 7.37$ ) compared to manual choice ( $M = 14.87, SD = 4.00$ ),  $t(59) = -26.699, p = .000, p < .001$ .

When collapsing the level of decision danger the results showed that participants were, overall, significantly more confident in autonomous system decisions ( $M = 38.34, SD = 6.74$ ) as compared with manual override decisions ( $M = 16.34, SD = 3.54$ ),  $t(59) = -35.358, p = .000, p < .001$ .

## 5. Discussion

This study investigated accuracy, confidence and within-subjects confidence-accuracy (W-S C-A) relationships across UAS groups (control, VGPs, private and professional pilots) to identify potential UAS supervisor level on factors relevant to decisions made where that danger/risk was manipulated and options to manually intervene or allow the autonomous systems to control the UAS were provided. Personality constructs were also considered.

### 5.1. Confidence

As predicted the groups differed significantly in decision confidence. It was found that professional pilots and VGPs showed significantly higher confidence compared to the control group. One reason for the greater confidence of these two groups may well be explained by the impact of prior experience, training and familiarity that both professional pilots and VGPs have with mission-based tasks which involve split-second reactive decisions and that have implication in the real or virtual world. Indeed, it has been shown that experience and training, including simulation, such as playing video games, can result in increased decision confidence (Atinaja-Faller et al., 2010; Chung & Monroe,

2000; Payne et al., 2002) and that familiarity can also result in increased decision confidence by enabling the illusion that individuals are accurately remembering important detail (Chandler, 1994).

Although not significant, the direction of the findings support the possibility that professional pilots show greater levels of confidence because they are used to the possibility of reality danger ( $M = 180.47$ ), whereas VGPs are operating within virtual danger ( $M = 172.53$ ). As professional pilots frequently make aviation decisions that have potential for major implication across all aspects of professional, private and the lives of others they thereby exhibit more confidence in such decisions. Conversely, private pilots, for example, do not fly as a career and are not used to the added stress of these types of issues which may be expressed in confidence levels. The study thus lends some support to previous work (Chung & Monroe, 2000; Keibell et al., 1996) which shows, for example, as difficulty increases confidence would be expected to decrease. In turn, this may suggest that high danger levels can lead to problematic decision-making.

Furthermore, it is an interesting observation that professional pilots do exhibit more confidence across the decision dangers than do private pilots again supporting the idea that experience, training and familiarity in dealing with high impact decisions will effectively increase individual confidence. However, confidence without experience, training and relevant knowledge can be regarded as overconfidence. In fact, research suggests that when complex tasks are unfamiliar such overconfidence relative to accurate decision-making can occur (Wheatcroft et al., 2004). Professionals not only have experience but are highly regarded and relied upon for their expertise; though it is necessary that the professional label is coupled with necessary training, particularly as it has been suggested that confidence can positively affect risk taking behaviour (Krueger & Dickson, 1994). One can then be more confident that a decision made with high confidence will more likely be accurate.

### **5.2. Confidence in decision type: Manual vs. automated**

A UAS supervisory role involves the allocation of functions between automated and manual which impact on how the system performs (Lee & Moray, 1992) hence, supervisors need to be able to apply the correct amount of confidence to decisions to both automated and those which require intervention. In this study, participants were always provided with the decision option to allow the autonomous system to control the UAS or to intervene and manually fly the vehicle and confidence ratings from both these decision types were obtained. Data were analysed to assess whether groups would differ in how much confidence was applied to manual decisions in comparison to automated decisions and vice versa as a function of different levels of decision danger.

The results demonstrate support for the hypothesis that confidence in decisions to automate and manually intervene will differ between groups. Professional pilots were more confident in manual decisions in both high and low decision danger and automated decisions in low decision danger levels. Similarly to overall confidence, experience may be able to explain this finding, as professional pilots are more likely to have experience making both manual decisions and also using autonomous systems compared to other groups, therefore displaying higher levels of confidence in those decisions. Research has shown that pilots tend to rely more on automation in comparison to student populations (Riley et al., 1993), which would explain elevated confidence in automated decisions. Thus, it seems that experience may increase confidence in both automated and manual decisions.

However, what is interesting to note is that no significant difference was found between groups in confidence of automated decisions in conditions of high decision danger. This would demonstrate that all groups displayed similar levels of confidence in those decisions in these conditions, providing further support to the idea that increases in decision danger does impact on confidence and is most evident in decisions to allow autonomous control the UAS.

Further analysis examined the data collapsed across all groups. Supervisors were significantly more confident in decisions in low danger than high decision danger and this occurred in both automated and manual decisions. Conditions of high decision danger were characterised as

encompassing more risk, consequently increasing decision complexity. In these conditions, it was found that decision confidence in choice to use automation was reduced in comparison to decisions made in low decision danger. Hence, individuals felt less confident in decisions made by the UAS when in conditions of increased risk and complexity. Such a finding lends support to previous research which showed in conditions of high risk automation is relied on less (Lyons & Stokes, 2011), as supervisors tend to display less confidence in the decision. Whilst in this study for manual decisions, danger increased confidence.

All groups tended to place a higher degree of confidence in a decision when they chose to let the autonomous system control the UAS. This was observed to occur in both high and low decision danger and regardless of decision danger level in the collapsed data. As mentioned, confidence scores do not necessarily relate to accuracy therefore these findings could demonstrate a tendency for supervisors to be overconfident in decisions made by autonomous systems, providing some support for automation bias (Mosier & Skitka, 1996). The idea that supervisors believe decisions made by automation have superior knowledge and consequently more confidence is placed in that decision is not new (Dijkstra et al., 1998). The concept is further supported by the reduced confidence shown in manual decisions in this study; contrasting previous research which argues manual decisions are preferred (de Vries, Midden, & Bouwhuis, 2003; Lee & Moray, 1992).

Although it can be argued that confidence in automation is beneficial in that it reduces workload (Parasuraman, Cosenzo, & de Visser, 2009) overconfidence can also cause operators not to attend to conflicting data (Cummings, 2004) and ignore erroneous decisions (Mosier, Palmer, & Degani, 1992). Hence, it is imperative that supervisors are able to correctly discriminate between incorrect and correct autonomous decisions and place the appropriate confidence in the actual decisions taken.

### 5.3. W-S C-A (and B-S C-A)

Confidence, while very important, represents overconfidence if it is not correlated in the appropriate direction with accuracy decisions. It is imperative then, when a supervisor exhibits a high level of confidence that a positive relationship exists between their assessment and the accuracy of response. That is to say: the greater confidence a supervisor expresses in their decision the more accurate those answers should be. However, whilst correlation is necessary it is not a sufficient condition for causality. The supervisor may have more confidence in their decisions because they have learned through past experience that their decisions have high accuracy under similar conditions.

There was no group effect for W-S C-A. This suggests the metacognitive measure, as one might expect, requires a different skill set not necessarily afforded by past experience. Thus group membership does not significantly improve individual awareness of the accuracy or inaccuracy of judgements made (DePaulo & Pfeifer, 2006). However, there was a main effect of decision danger; largely, when decision danger was high W-S C-A significantly decreased. This finding validates the categorisation of decision danger and reinforces that decisions which carry dangerous implication are harder to assess, judge and therefore make. A closer examination of the means showed that, although no interaction was observed, VGPs W-S C-A remained relatively constant across the three decision danger categories and were able to produce the highest W-S C-A in the most difficult category. In comparison, other groups achieved roughly no correlation or one which was negative (i.e. control) for the same category. To be able to maintain positive W-S C-A levels for high decision dangers is crucial as these decisions are considered critical junctures at which things could go wrong. The observation that VGPs produce the highest correlation between confidence and accuracy for risk suggests that VGPs may be a good resource for UAS supervision, as, according to the findings here, this group are able to show the best awareness of the accuracy or inaccuracy of their decisions, particularly those characterised by high danger and avoidance of overconfident ratings. However, a person's accuracy for risk assessment and making the correct decisions is a key indicator of suitability for UAS operations—as confidence could improve with experience. There was a significant positive relationship between confidence and accuracy (B-S C-A).



#### 5.4. Accuracy

The groups did not differ in the accuracy of decisions. It is however plausible this outcome could have been observed given the standardisation of the study. Groups however were separated by their experiences and the finding that groups made reasonably equivalent decisions in terms of accuracy lends support to the Chung and Monroe (2000) finding that experience has no effect on accuracy. Indeed, Boot, Kramer, Simons, Fabiani, and Gratton (2008) suggest that any improved performance seen in video game players compared to non-players may either be due to practice in honing cognitive skills through the act of playing video games, or could be a result of self-selection to pursue video game playing given pre-existing skills. Video game players may have the capacity to develop relevant UAS skills through practice or may inherently possess the skills which drew them to game playing. However, the findings here do not support that view. Accuracy can be seen as a measure of not only whether the group members accurately identified the optimal response but also whether a supervisor allows for automatic control or takes control manually. The accuracy score can provide some insight into whether supervisor trust is appropriate to UAS capability. It could be said that as decision danger increases accuracy decreases and that supervisor trust is negatively affected and thereby inappropriate. For example, either misuse (i.e. using automation for more than it should be used for—the danger is too high so individual allows automation to make the decision/carry out action), or disuse (i.e. using manual control unnecessarily—the danger is considered too great to trust the automation so individual is more confident in own ability) can occur more readily under such conditions. Of course, this suggestion would require further investigation. What can be said is that given the best information and learning experience the role of UAS supervisor is within the scope of non-pilot trained individuals; while the role requires new skill sets and aviation experience may provide some slight advantage it was certainly not found to be a significant factor here.

Broadly speaking, for decision danger the hypotheses were largely supported; accuracy, confidence and W-S C-A relationships reduced significantly as decision danger increased. The simple exception to this was that no differences were observed between low and medium decision danger for accuracy, confidence or W-S C-A. Therefore, a constant feature was that high decision danger impacted negatively on and across the accuracy, confidence and W-S C-A measures. It also had a role to play in the confidence expressed when choosing to intervene or rely on automation.

#### 5.5. Personality constructs

It was predicted that the measured characteristics would differ across the groups and in part this was supported. Professional pilots scored lower for neuroticism than the control group which suggests the professional pilot group are more likely to address problems in an emotionally stable, calm, even-tempered and relaxed manner, and would be better equipped to cope with the stresses involved. Given the crew-context professional pilots work in they are much more likely to express these characteristics in an altruistic fashion where successful task completion takes priority. Moreover, the professional pilot group scored significantly higher on agreeableness than VGPs; indicating the latter would be more prone to competitiveness rather than helpfulness.

Of note is that, overall, neuroticism was negatively related to confidence; thus, those individuals who score highly on this construct would be less able to control impulses and be susceptible to irrational thoughts and/or behaviour—which may well increase in intensity under stress. Neuroticism construct screening for professional pilots is thereby very informative as inability to cope with stress can inhibit problem-solving, increase anxiety and make for less confident decisions (Michie, 2002). Further, the finding that professional pilots express significantly higher levels of confidence perhaps supports this idea; though the authors remain mindful that this higher level of confidence was not always reflected in accurate responses, particularly for decisions classified as high danger (see W-S C-A). Nonetheless, these findings suggest an advantage in selection of individuals for training who score low on neuroticism as it can be considered vital for confidence in critical decision-making.

Correlational analysis, across all participant groups, showed that conscientiousness was positively related to confidence. The more planning, organisation and task focus an individual has means they

are likely to express increased confidence; in this case in decisions made. It suggests that conscientiousness is a desirable trait that UAS supervisors should hold to exhibit increased decision confidence. Researchers have considered aptitude tests (Carretta, Rose, & Barron, 2015) and the utility of personality (Chappelle, McDonald, Heaton, Thompson, & Haynes, 2012) in the US Air Force, with the suggestion that other measures to supplement the current arrangements would be helpful.

What is as important is whether a relationship exists between confidence and accurate decisions. Here, intolerance of ambiguity (A) was negatively related to W-S C-A. In this case, lower levels of tolerance of ambiguity (A) means one expresses greater tolerance to ambiguous contexts and tasks. Aviation is characterised by the need to make critical and potentially irreversible decisions during flight without the benefit of discussion and timed reflection. Therefore, greater tolerance of these kinds of situations is psychologically advantageous in that individuals can make confident and accurate decisions without the undue negative effects dissonance can bring. It follows that intolerant individuals would feel uncomfortable and be motivated to reduce this by making a decision inhibited by lowered confidence and inability to accurately judge correctness. As such, individuals who have a greater tolerance to ambiguity will show increased W-S C-A relationships and is thereby an important resource in the role of UAS supervision illustrating the importance of greater sensitivity in the assessment of efficacious decision-making. The W-S C-A correlation affords a start point for meta-cognitive measurement in this context.

Of course, the participants knew the decisions had little real-life implication and thus the outcomes are generalisable only in this context. Despite this limitation the study attempted to maintain high verisimilitude as the simulation equipment was modelled on a UAS supervisory environment which incorporated the known requirements of supervisory control. Further the W-S C-A measure has been applied successfully in different contexts (Wheatcroft & Ellison, 2012; Wheatcroft & Woods, 2010). There are however traits other than those studied here that may be relevant to accuracy (Szalma & Taylor, 2011), confidence and decision-making in this important human machine interface.

It would therefore be beneficial to conduct further research into the impact of conditions containing even greater ecological validity. One way might be to instruct one group that they are indeed monitoring a UAS. Further, while this study was able to verify response accuracy against UAS supervisor expert decision logs there is scope to systematically increase the difficulty and complexity of events and to measure multiple decisions and/or sources of information. Indeed, the complexity of the factors and effects involved suggest that for any selection tools to be effective the optimal profiles would need to be developed separately for each level, type and so on (Szalma & Taylor, 2011). Latency to decisions could also be important to measure across groups and environments. Moreover, other groups (i.e. air traffic controllers who have experience of supervising multiples of aircraft) and group age may also be assessed.

The study adds to current literature which has the goal of developing ways of identifying and selecting appropriate personnel for specific tasks in this context. One might argue that focusing on SAOCs rather than specific VGP experience would result in a larger pool but this is yet to be established.

## 6. Conclusion

The Civil Aviation Authority (2012) recognises there is currently no approved training course for UAS supervisory role for vehicles above 20 kg MTOM. It does however express the view that a UAS supervisor need not have manned aircraft pilot experience in the recognition that a supervisor may require alternative skill sets. The findings here give ground to the idea that VGPs could be considered as a resource; indeed, VGPs display constant W-S C-A across decision danger categorisations. VGPs exhibit some skills that may be required in successful UAS supervision, particularly as they are least likely to exhibit overconfidence in decision judgements. All groups displayed an increase in decision confidence in automated decisions which can be problematic when unmatched against decision accuracy. Personality constructs measured suggest operators be selected for low neuroticism, high

conscientiousness and tolerance to ambiguous contexts. It is important to note that for supervisors to appropriately increase decision confidence, the experience gained, in training, familiarity (simulated or real), in mission-based tasks and time-limited decisions involving criticality will most likely be required to be updated as part of continuous personal development.

#### Funding

The authors received no direct funding for this research.

#### Author details

Jacqueline M. Wheatcroft<sup>1</sup>

E-mail: [jacmw@liverpool.ac.uk](mailto:jacmw@liverpool.ac.uk)

ORCID ID: <http://orcid.org/0000-0001-7212-1598>

Mike Jump<sup>2</sup>

E-mail: [mjump1@liverpool.ac.uk](mailto:mjump1@liverpool.ac.uk)

Amy L. Breckell<sup>3</sup>

E-mail: [a.breckell@liverpool.ac.uk](mailto:a.breckell@liverpool.ac.uk)

Jade Adams-White<sup>1</sup>

E-mail: [ps0u91c6@liverpool.ac.uk](mailto:ps0u91c6@liverpool.ac.uk)

<sup>1</sup> Department of Psychological Sciences, Eleanor Rathbone Building, University of Liverpool, Liverpool L69 7ZA, UK.

<sup>2</sup> Department of Engineering, University of Liverpool, Liverpool L69 7ZA, UK.

<sup>3</sup> School of Psychology, University of Liverpool, Liverpool L69 7ZA, UK.

#### Citation information

Cite this article as: Unmanned aerial systems (UAS) operators' accuracy and confidence of decisions: Professional pilots or video game players?, Jacqueline M. Wheatcroft, Mike Jump, Amy L. Breckell & Jade Adams-White, *Cogent Psychology* (2017), 4: 1327628.

#### Cover image

The image was created by the researchers using multi-body dynamics software FLIGHTLAB and Presagis' industry-standard display creation Virtual Avionics Prototyping Software.

#### References

- Adidam, P. T., & Bingi, R. P. (2000). The importance of decision confidence to strategy outcomes. *The Journal of Applied Business Research*, 16, 35–50.
- Atinaja-Faller, J., Quigley, L., Banichoo, S., Tsveybel, V., & Quigley, M. (2010, November). *Simulation training effect on confidence levels of senior clinical decision-making nursing students in managing an acute event*. Paper presented to International Conference of Education, Research and Innovation, Madrid.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26. <https://doi.org/10.1111/peps.1991.44.issue-1>
- Bingi, R. P., Kasper, G. (1995, March). *A model of the effect of decision aids on decision confidence and decision success*. Paper presented to SouthWest DSI (SWFAD), Houston.
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, 129, 387–398. <https://doi.org/10.1016/j.actpsy.2008.09.005>
- Budner, S. (1961). Intolerance of ambiguity as a personality variable. *Journal of Personality*, 30, 29–50.
- Bureau d'Enquêtes et d'Analyses (BEA). (2012). *Final report: On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris*. Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, Ministère de l'Écologie, du Développement durable, des Transports et du Logement, 27th July. Retrieved from <https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf>
- Burke, E. (1995). Pilot selection II: Where do we go from here? In R. S. Jensen (Ed.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 1347–1353). Columbus, OH: The Ohio State University.
- Carretta, T. R. (2013). Predictive validity of pilot selection instruments for remotely piloted aircraft training outcome. *Aviation, Space, and Environmental Medicine*, 84, 47–53. <https://doi.org/10.3357/ASEM.3441.2013>
- Carretta, T. R., Rose, M. R., & Barron, L. G. (2015). Predictive validity of UAS/RPA sensor operator training qualification measures. *The International Journal of Aviation Psychology*, 25, 3–13. <https://doi.org/10.1080/10508414.2015.981487>
- Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica*, 119, 217–230. <https://doi.org/10.1016/j.actpsy.2005.02.004>
- Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition*, 22, 273–280. <https://doi.org/10.3758/BF03200854>
- Chappelle, W., McDonald, K., Heaton, J.N., Thompson, W., & Haynes, J. (2012). *Neuropsychological and personality attributes distinguishing high vs. low training performance of MQ-1B pilot trainees*. Paper presented to Aerospace Medicine Association Annual Conference, Atlanta, GA.
- Chidester, T., Helmreich, R., Gregorich, S., & Geis, C. (1991). Pilot personality and crew coordination: Implications for training and selection. *The International Journal of Aviation Psychology*, 1, 25–44. [https://doi.org/10.1207/s15327108ijap0101\\_3](https://doi.org/10.1207/s15327108ijap0101_3)
- Chidester, T., Kanki, B., Foushee, H., Dickinson, C., & Bowles, S. (1990). *Personality factors in flight operations (Vol. 1): Leader characteristics and crew performance in a full-mission air transport simulation* (NASA Technical Memorandum-102259). Moffett Field: CAL NASA Ames Research Centre.
- Chung, J., & Monroe, G. S. (2000). The effects of experience and task difficulty on accuracy and confidence assessments of auditors. *Accounting and Finance*, 40, 135–151. <https://doi.org/10.1111/acfi.2000.40.issue-2>
- Civil Aviation Authority. (2010). *Lasors 2010*. Norwich: The Stationary Office.
- Civil Aviation Authority. (2012). *CAP722 unmanned aircraft system operations in UK airspace – guidance*. Norwich: The Stationary Office.
- Civil Aviation Authority. (2013). *Monitoring matters: Guidance on the development of pilot monitoring skills* (CAA Paper 2013/02). West Sussex: Author. Retrieved from <https://publicapps.caa.co.uk/docs/33/9323-CAA-Monitoring%20Matters%202nd%20Edition%20April%202013.pdf>
- Civil Aviation Authority. (2016). *CAP1377 ATM automation: Guidance on human-technology integration*. West Sussex: Author.
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13, 653–665. [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I)
- Crang, E. A., & Lenfestey, A. G. (2009). *Architecting human operator trust in automation to improve system effectiveness in multiple Unmanned Aerial Vehicle (UAV) control (AFIT/GSE/ENV/09-M06)*. Wright-Patterson AFB, OH: Air Force Institute of Technology.
- Cummings, M. L. (2004). *Automation bias in intelligent time critical decision support systems*. Paper presented to AIAA 1st Intelligent Systems Technical Conference, Chicago, IL.

- Department of Defense. (2007). *Unmanned systems roadmap 2007–2032* (see AD No. ADA475002). Washington, DC: Defense Technical Information Center.
- DePaulo, M. B., & Pfeifer, L. R. (2006). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16, 249–267.
- de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*, 58, 719–735.  
[https://doi.org/10.1016/S1071-5819\(03\)00039-9](https://doi.org/10.1016/S1071-5819(03)00039-9)
- Dijkstra, J. J., Liebrand, W. B., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*, 17, 155–163.  
<https://doi.org/10.1080/014492998119526>
- Dorval, M., & Pepin, M. (1986). Effect of playing a video game on a measure of spatial visualization. *Perceptual Motor Skills*, 62, 159–162.  
<https://doi.org/10.2466/pms.1986.62.1.159>
- Drury, J. L., & Scott, S. D. (2008). Awareness in unmanned aerial vehicle operations. *The International C2 Journal*, 2, 1–10.
- Dzindolet, M. T., Beck, H. P., Pierce, L. G., & Dawe, L. A. (2001). *A framework of automation use* (Report No. ARL-TR-2412 (Aberdeen Proving Ground)). Adelphi, MD: Army Research Laboratory.  
<https://doi.org/10.1037/e447142006-001>
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York, NY: McGraw-Hill.
- Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight deck automation issues. *The International Journal of Aviation Psychology*, 9, 109–123.  
[https://doi.org/10.1207/s15327108ijap0902\\_2](https://doi.org/10.1207/s15327108ijap0902_2)
- Ghosh, D., & Ray, M. R. (1997). Risk, ambiguity, and decision choice: Some additional evidence. *Decision Sciences*, 28, 81–104. <https://doi.org/10.1111/dec.1997.28.issue-1>
- Glussich, D., & Histon, J.M. (2010). *Human/automation interaction accidents: Implications for UAS operations*. Paper presented to Digital Avionics Systems Conference (DASC), Salt Lake City, UT.
- Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors*, 36, 387–405.
- Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, 18, 88–94.  
<https://doi.org/10.1111/j.1467-9280.2007.01853.x>
- Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills*, 57, 155–158.  
<https://doi.org/10.2466/pms.1983.57.1.155>
- Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology*, 87, 653–662.  
<https://doi.org/10.1111/bjop.1996.87.issue-4>
- Krueger, N., & Dickson, P. R. (1994). How believing in ourselves increases risk taking: Perceived self-efficacy and opportunity recognition. *Decision Sciences*, 25, 385–400.  
<https://doi.org/10.1111/dec.1994.25.issue-3>
- Latane, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York, NY: Appleton-Century-Crofts.
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35, 1243–1270.  
<https://doi.org/10.1080/00140139208967392>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46, 50–80.  
<https://doi.org/10.1518/hfes.46.1.50.30392>
- Lyons, J. B., & Stokes, C. K. (2011). Human-human reliance in the context of automation. *Human Factors*, 54, 112–121.  
<https://doi.org/10.1177/0018720811427034>
- Maddux, J. E., & Volkman, J. (2010). Self-efficacy. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation*. Oxford: Wiley-Blackwell.
- Madsen, M., & Gregor, S. (2000). *Measuring human-computer trust*. Gladstone: Central Queensland University.
- Martins, E., & Soares, M. (2012). Automation under suspicion: Case flight AF-477 Air France. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41, 222–224.
- McCrae, R. R., & Costa, P. T. (1992). Discriminant validity of NEO-PIR facet scales. *Educational and Psychological Measurement*, 52, 229–237.  
<https://doi.org/10.1177/001316449205200128>
- McKinley, A. R., & McIntire, L. K. (2009). *Operator selection for unmanned aerial vehicle operators: A comparison of video game players and manned aircraft pilots* (AFRL-RH-WP-TR-2010-0057). Air Force Research Laboratory. Retrieved from <http://www.dtic.mil>
- Michie, S. (2002). Causes and management of stress at work. *Occupational and Environmental Medicine*, 59, 67–72.  
<https://doi.org/10.1136/oem.59.1.67>
- Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application* (pp. 201–220). Mahwah, NJ: Erlbaum.
- Mosier, K. L., Palmer, E. A., & Degani, A. (1992). *Electronic checklists: Implications for decision-making*. Paper presented to Human Factors Society 36th Annual Meeting, Human Factors Society, Atlanta, GA.
- Muir, B. M. (1994). Trust in automation: Part II. Experimental studies of trust and human intervention in a process control stimulation. *Ergonomics*, 39, 429–460.
- Murnaw, R. J., Sarter, N. B., & Wickens, C. D. (2001). *Analysis of pilots' monitoring and performance on an automated flight deck, 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.
- National Transportation Safety Board. (1973). *Eastern Airlines L-1011: Miami, Florida, 20 December 1971* (NTSB Report-ARR-73-14). Washington, DC: Author.
- Parasuraman, R., Cosenzo, K. A., & de Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, 21, 270–297.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52, 381–410.  
<https://doi.org/10.1177/0018720810376055>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39, 230–253.  
<https://doi.org/10.1518/00187209778543886>
- Parasuraman, R., Molloy, R., & Singh, I. L. (2009). Performance consequences of automation-induced 'complacency'. *International Journal of Aviation Psychology*, 3(1), 1–23.
- Payne, F., Harvey, K., Jessopp, L., Plummer, S., Tylee, A., & Gournay, K. (2002). Knowledge, confidence and attitudes towards mental health of nurses working in NHS Direct and the effects of training. *Journal of Advanced Nursing*, 40, 549–559.  
<https://doi.org/10.1046/j.1365-2648.2002.02413.x>
- Perkins, L., Miller, J., Hashemi, A., & Burns, G. (2010). Designing for human centred systems: Situational risk as a factor of trust in automation. Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society, 27 September–1 October. San Francisco, CA.
- Piedmont, R. L., & Weinstein, H. P. (1993). A psychometric evaluation of the new NEO-PIR facet scales for



- agreeableness and conscientiousness. *Journal of Personality Assessment*, 60, 302–318.  
[https://doi.org/10.1207/s15327752jpa6002\\_8](https://doi.org/10.1207/s15327752jpa6002_8)
- Prinzel, L. J., DeVries, H., Freeman, F. G., & Mikulka, P. (2001). *Examination of automation-induced complacency and individual differences variates* (see technology memorandum No. 211413). Hampton, VA: NASA Langley Research Center.
- Riley, V. (1994). *Human use of automation* (Unpublished doctoral dissertation). University of Minnesota, USA.
- Riley, V., Lyall, E., & Weiner, E. (1993). *Analytic methods for flight-deck automation design and evaluation, phase two report: Pilot use of automation* (Federal Aviation Administrator Contractor Report). Minneapolis, MN: Honeywell Technology Centre.
- Roets, A., & Van Hiel, A. (2007). Separating ability from need: Clarifying the dimensional structure of the need for closure scale. *Personality and Social Psychology Bulletin*, 33, 266–280.  
<https://doi.org/10.1177/0146167206294744>
- Rose, M. R., Barron, L. G., Carretta, T. R., Arnold, R. D., & Howse, W. R. (2014). Early identification of unmanned aircraft pilots using measures of personality and aptitude. *International Journal of Aviation Psychology*, 24, 35–52.
- Ross, J. M., Szalma, J. L., Hancock, P. A., Barnett, J. S., & Taylor, G. (2008). *The effect of automation reliability on user automation trust and reliance in a search-and-rescue scenario* (pp. 1340–1344). Paper presented to Humans Factors and Ergonomics Society 52nd Annual Meeting, New York, NY.
- Ruff, H. A., Calhoun, G. L., Draper, M. H., Fontejon, J. V., & Guilfoos, B. J. (2004). Exploring automation issues in supervisory control of multiple UAVs. *Proceedings of the Human Performance, Situation Awareness and Automation Technology Conference*, 218–222.
- Sanchez, J., Rogers, W. A., Fisk, A. D., & Rovira, E. (2014). Understanding reliance on automation: Effects of error type, error distribution, age and experience. *Theoretical Issues in Ergonomics Science*, 15, 134–160.  
<https://doi.org/10.1080/1463922X.2011.611269>
- Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology*, 4(1), 1–28.  
[https://doi.org/10.1207/s15327108ijap0401\\_1](https://doi.org/10.1207/s15327108ijap0401_1)
- Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37, 5–19.  
<https://doi.org/10.1518/001872095779049516>
- Shaw, R. B. (1997). *Trust is the balance: Building successful organizations on results, integrity, and concern*. San Francisco, CA: Jossey-Bass.
- Siem, F. M., & Murray, B. S. (1994). Personality factors affecting pilot combat performance: A preliminary investigation. *Aviation, Space, and Environmental Medicine*, 65, A45–A48.
- Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, 52, 701–717.  
<https://doi.org/10.1006/ijhc.1999.0349>
- Szalma, J. L., & Taylor, G. S. (2011). Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied*, 17, 71–96.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed.). Chicago, IL: Nelson Hall.
- Toma, C., & Butera, F. (2009). Hidden profiles and concealed information: Strategic information sharing and use in group decision making. *Personality and Social Psychology Bulletin*, 35, 793–806.  
<https://doi.org/10.1177/0146167209333176>
- Tvaryanas, A. P., Thompson, B. T., & Constable, S. H. (2005). *U.S. military unmanned aerial vehicle mishaps: Assessment of the role of human factors using HFACS* (Technical Report). Brooks City-Base, TX: USAF Human Performance Directorate.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Walliser, J. C. (2011). *Trust in automated systems: The effect of automation level on trust calibration*. Naval Postgraduate School. Retrieved July 2011, from [http://edocs.nps.edu/npspubs/scholarly/theses/2011/June/11Jun\\_Walliser.pdf](http://edocs.nps.edu/npspubs/scholarly/theses/2011/June/11Jun_Walliser.pdf)
- Webster, M., Cameron, N., Fisher, M., & Jump, M. (2014). Generating certification evidence for autonomous unmanned aircraft using model checking and simulation. *Journal of Aerospace Information Systems*, 11, 258–279.  
<https://doi.org/10.2514/1.1010096>
- Wheatcroft, J. M., & Ellison, L. E. (2012). Evidence in court: Witness preparation and cross-examination style effects on adult witness accuracy. *Behavioral Sciences & the Law*, 30, 821–840. <https://doi.org/10.1002/bsl.2031>
- Wheatcroft, J. M., & Woods, S. (2010). Effectiveness of witness preparation and cross-examination non-directive and directive styles on witness accuracy and confidence. *International Journal of Evidence and Proof*, 14, 189–209.
- Wheatcroft, J. M., Wagstaff, G. F., & Kebbell, M. R. (2004). The influence of courtroom questioning style on actual and perceived eyewitness confidence and accuracy. *Legal & Criminological Psychology*, 9, 83–101.  
<https://doi.org/10.1348/135532504322776870>
- Wheatcroft, J. M., Wagstaff, G. F., & Manarin, B. (2015, April 3). The influence of delay and item difficulty on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research*, 1, 1–9
- Will, R. P. (1991). True and false dependence on technology: Evaluation with an expert system. *Computers in Human Behavior*, 7, 171.  
[https://doi.org/10.1016/0747-5632\(91\)90006-M](https://doi.org/10.1016/0747-5632(91)90006-M)
- Woods, D. D., & Sarter, N. B. (2000). Learning from automation surprises and going sour accidents. In N. Sarter & R. Amalberti (Eds.), *Cognitive engineering in the aviation domain* (pp. 327–354). Hillsdale, NJ: Erlbaum.
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1310–1321.  
<https://doi.org/10.1098/rstb.2011.0416>
- Yuji, H. (1996). Computer games and information-processing skills. *Perceptual and Motor Skills*, 83, 643–647.  
<https://doi.org/10.2466/pms.1996.83.2.643>





© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

**Cogent Psychology (ISSN: 2331-1908) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

