INDIVIDUAL DIFFERENCES IN CUE ACQUISITION AND STAGES OF SKILL DEVELOPMENT

Nadya Christye Yuris

PhD in Psychology

Department of Psychology

Faculty of Human Sciences

Macquarie University, Sydney, Australia

Thesis submitted on 13th November 2019

TABLE OF CONTENTS

DECLARATION

This thesis has not previously been submitted for a degree or diploma in any university or institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself. The research reported in this thesis has received ethical approval from the Macquarie University Ethics Committee on 09th December 2016 (5201600842), 14th March 2017 (5201700150), and 14th March 2018 (5201800115).

Nadya Christye Yuris Date: 13/11/2019

SUMMARY

In many operational domains, the utilisation of cues enables operators to recognise and respond rapidly and successfully to changes in the system state. This process is initiated by matching features in the environment with events or objects in memory in the form of cues. Although cue utilisation represents a useful and reliable measure of concurrent performance, it is a construct reliant on associations that have been acquired through previous experience. Typically, this restricts measures of cue utilisation to operational personnel who have acquired a level of experience.

For those environments where cue utilisation is necessary, and where the selection of personnel occurs at the initial stages of skill acquisition, an alternative approach is required that evaluates the capacity of the candidate to identify, acquire, and retain cue-based associations in memory. Therefore, the aim of this programme of research was to construct and evaluate a measure of domain-independent cue acquisition and test the validity of the measure in predicting cue-based performance across a range of domains.

In Study 1, a series of domain-independent tasks were developed that were designed to examine the predictive validity of different assessments of cue acquisition. Of the five tasks constructed, the Timed-Search Task (TST), incorporating a time constraint and implicit associations, was the only measure that positively predicted subsequent performance on a novel rail control task, in which improved performance is depended upon successful cue acquisition.

Based on the outcomes of Study 1, Study 2 was conducted to establish the construct validity of the TST by evaluating its relationship to cue utilisation in the context of general aviation. Individual differences in the TST were related to performance on two of the five tasks that comprised a composite measure of cue utilisation. The results suggested that individual differences in cue acquisition may be associated with the more rudimentary aspects of cue utilisation.

Study 3 was designed to evaluate the association between cue acquisition and cue utilisation in a dynamic task that was more complex than rail control. Within the context of motor vehicle driving, higher cue utilisation was associated with fewer driving errors, fixations and saccades, accounting for cue acquisition. Cue acquisition was not associated with overall driving performance. The results suggest that, where cue utilisation significantly impacts performance, the influence of cue acquisition on performance may be restricted to the initial stages of skill acquisition.

Study 4 was designed to examine the predictive validity of cue acquisition in a ‘real-world’ context. Consistent with Study 2, general aviation comprised the testing domain, with pilots completing a flight training exercise and their performance assessed using flight control measures and instructor ratings, accounting for hours of flight experience. The results failed to provide support for an association between the TST and flight performance, although difficulties in the level of experimental control precluded a conclusive assessment.

Study 5 was undertaken to test the outcomes of Study 4 using a flight simulation context that offered a degree of experimental control. Licensed pilots flew a simulated flight, during which an engine failure occurred within proximity to an alternate destination. Consistent with the preceding results, cue acquisition was not associated with flight control measures nor was it predictive of landing success, accounting for flight experience.

In combination, the results across the five studies suggest that the impact of individual differences in cue acquisition, as measured by the TST, appeared to be limited to the initial learning outcomes associated with novel operations. After the initial acquisition of relevant cues, it was apparent that individual differences in cue acquisition had minimal contribution to domain-specific performance.

CHAPTER ONE

Overview

On February 04th 2015, TransAsia Airways Flight GN235, an ATR-72, departed Taipei Songshan Airport bound for Kinmen, Republic of China. During the initial climb, the aircraft experienced a loss of control and crashed into the Keelung River, killing 43 and injuring 17 passengers and crew. The subsequent investigation report highlighted a number of probable causes, one of which was the pilot's competence in assessing the situation as it presented (Aviation Safety Council, 2016).

Shortly after take-off from Taipei, one of the two engines on GN232 'flamed-out' signifying a loss of power. The co-pilot identified, and verbally expressed to the captain, the malfunction in the right engine. The captain, however, responded by shutting off the functioning left engine despite the co-pilot's verbalisation that it was the right engine that had failed. This action was also contrary to the written advice displayed on the Engine Warning Display that the engine should be identified clearly before shutting it down. The captain also disregarded the subsequent stall warnings, including the auditory stall warning and the activation of the 'stick shakers', and had not recognised that the airspeed of the aircraft was approaching a stall. Two minutes after the left engine was shut down, the aircraft stalled, failed to recover, and crashed into the Keelung river. In this case, it appeared that the captain failed to appropriately assess and integrate the relevant information available in sufficient time to avert a collision with terrain (Aviation Safety Council, 2016).

Like Flight GN235, Air France Flight 447, an Airbus A340, crashed into the Atlantic Ocean after failing to recover from a stall following a series of erroneous assessments and responses (Bureau of Enquiry and Analysis for Civil Aviation Safety, 2012). On June 1st, 2009, AF447 departed Rio de Janeiro, Brazil for Paris, France. The aircraft was flown into cloud layers in the Intertropical Convergence Zone, where icing occurred that caused a malfunction in the speed sensors. This led to the automatic disconnection of the autopilot, and the two junior pilots, who were in command at the time, were required to take manual control of the aircraft.

Following the transfer of control from the automated system to the pilots, the Pilot Flying abruptly increased the pitch of the aircraft to a steep climb, which immediately triggered the auditory stall warning. Throughout the period leading up to the descent, neither the Pilot Flying nor the Pilot Not Flying addressed the stall warning. The Pilot Flying continued to draw back on the side stick, culminating in an extreme nose-up attitude, a loss of horizontal speed, and an increase in downward vertical speed. One minute after the autopilot was disengaged, the aircraft stalled, and three minutes later, the failure to recover from the stall resulted in the fatal crash.

In effect, the Pilot Flying had formed an inaccurate assessment of the situation that confronted Flight AF447. He appeared not to have recognised nor integrated the associations between his actions and the stall warning. This apparent lack of comprehension of the causal associations contributed to the confusion and difficulties for both pilots in establishing a working understanding of the situation.

Situation Assessment

In the case of both GN235 and AF447, there was a failure on the part of the pilots to construct accurate assessments of the emerging situations. Situation assessment is a cognitive process that requires the extraction of information from the environment and its integration with existing information in memory (Horrey & Wickens, 2001). It occurs during the initial stages of decision making and establishes the basis for the subsequent formation or derivation of decision plans and behavioural responses (Lipshitz, Klein, Orasanu, & Salas, 2001). Importantly, situation assessment is a process that allows operators to determine the appropriate courses of actions, including whether further information is needed, which response is most appropriate, and/or whether the situation requires an intervention.

Given that action plans are derived from situation assessments, the accuracy of these assessments will, in part, influence the effectiveness of the action plans selected. In aviation, situation assessment for pilots who do not hold instrument ratings is based primarily on visual information in the environment, including visibility, cloud ceiling, the location of other aircraft, and reference to the horizon. For example, in a simulated flight study, Wiegmann, Goh, and O'Hare (2002) observed that pilots who chose to continue to fly a planned route into deteriorating weather conditions were less accurate in their assessment of visibility and cloud ceiling compared to those pilots who elected to divert to an alternate destination. This observation suggests that the pilots' responses to the situation were associated with the accuracy with which they interpreted the weather conditions that they confronted during the flight.

According to Goh and Wiegmann (2001), pilots who chose to continue the flight as planned also expressed greater confidence in their piloting abilities and judged hazards as less severe than pilots who chose to divert, despite both groups having acquired similar levels of flight experience and training. In combination, these results suggest that the pilots' decisions to continue a flight in the face of deteriorating weather conditions may result from inaccurate assessments of the severity of the situation to which they are exposed.

Consistent with Wiegmann et al. (2002), Engström, Gregersen, Hernetkoski, Keskinen, & Nyberg (2003) noted that poorer driving performance in operating a four-wheel motor vehicle is associated with a tendency to make less precise assessments of driving conditions, especially in identifying hazards. Mueller and Trick (2012) attributed these characteristics to novice drivers in particular, and suggested that novice drivers are less accurate in their assessments of hazards compared to more experienced drivers, and that this lower accuracy contributes to their generally higher rates of accidents and/or incidents involving collisions, slower brake responses, and poor compensation for speed when driving in foggy conditions. In combination, these observations suggest that the inaccurate assessments of hazards may contribute to ineffective responses in high-risk, dynamic environments, where operators either failed to respond to a change in the system state or responded inappropriately.

Assessments Under Time Constraints

In high-risk domains, such as aviation and firefighting, assessments of situations are often necessary under time constraints. The failure to formulate accurate assessments in sufficient time can result in delayed and/or inappropriate responses, culminating in accidents and/or incidents with significant consequences for operators, the public, and/or infrastructure.

Time constraint creates a perceived reduction in time available to process information, which in turn, results in increases in anxiety and stress (Maule, Hockey, & Bdzola, 2000). Under time pressure, stress consumes working memory resources resulting in fewer residual cognitive resources available to manage attention and perceptual information (Stokes & Raby, 1989; Wickens, Stokes, Barnett, & Hyman, 1988). Consequently, time constraints impose limitations on the capacity to search for, identify, and acquire information from the environment, thereby impeding the capacity to acquire and process perceptual information to form appropriate assessments.

Accordingly, under time constraint, successful performance requires that operators employ cognitive strategies that enable efficiencies in the acquisition, integration, and application of task-related information. These cognitive strategies normally comprise pre-existing frameworks in memory that generate expectancies and direct the allocation of attention towards critical sources of information (Gigerenzer & Gaissmaier, 2011).

Pre-existing frameworks in memory are referred to as schemas or mental structures consisting of information acquired from and organised based on previous experiences (Ghosh & Gilboa, 2014). Schemas act as cognitive templates against which perceived features in the environment can be matched to derive rapid situation assessments (Kaempf, Klein, Thordsen,

& Wolf, 1996). When a match is identified, the situation is perceived by the operator as 'familiar' and an assessment is constructed based on the information in the matched schema (Klein, 2008).

An illustration of the importance of schemas can be drawn from comparisons between chess masters and novices when they are presented with complex chess configurations for a brief period (Chase & Simon, 1973). When the configurations represent meaningful patterns of chess pieces that could emerge during the course of a game chess, grandmasters are able to reconstruct the configurations with greater accuracy than novice chess players. However, when the configurations of chess pieces are presented in random configurations, chess masters and novices do not significantly differ in the accuracy of their reconstructions. Rather than simply a superior memory, the outcomes suggest that chess grandmasters possess a repertoire of precise, detailed configurations of chess pieces in the form of schemas in memory, the application of which reduces cognitive demands, thereby enabling superior performance. This advantage that grandmasters have over novices disappears when the configurations are random and do not match existing schemas.

Situation Assessment Beyond Aviation

In corporate organisational contexts, the processes engaged during situation assessment are typically described as 'sensemaking'. Although there are substantial variations in the conceptualisations of sensemaking in the literature, a number of the underlying components remain consistent (Klein, Moon, & Hoffman, 2006). In particular, there is a fundamental emphasis on the comprehension and the prediction of situations by comparing perceived stimuli against schemas or mental templates such as organisational policies, constraints and expectations (Gephart, 1993; Maitlis & Sonenshein, 2010; Weick, Sutcliffe, & Obstfeld, 2005).

Weick (1995) first referred to sensemaking as the process of understanding situations that appear to violate expectations and/or are novel or uncertain to the perceiver. For example, when organisations are subjected to restructuring or change, middle managers must engage in sensemaking to develop an understanding of the nature of the change imposed by senior managers before they can implement appropriate action plans and communicate with their employees effectively (Balogun & Johnson, 2004; Rouleau & Balogun, 2011). Consistent with situation assessment, sensemaking requires the acquisition and integration of information from the environment and memory to form functional interpretations that can be acted upon.

According to Maitlis and Christainson (2014), sensemaking comprises four components, including dynamic processing, a reliance on cues, social-based operations, and human-environment interaction. As a dynamic process, making sense of situations is updated constantly, where new information is integrated with a pre-existing understanding of similar situations (Mills, Thurlow, & Mills, 2010). Therefore, the efficiency and accuracy of predictions, and ultimately, the selection of appropriate actions in response to a change in the system state, is likely to comprise interpretations drawn from previous perceptions or experiences.

Previous experience enables the development, testing, and reinforcement of schemas and cues. These become central to the process of sensemaking as they govern both the initiation of sensemaking, and the construction of the assessments (Whiteman & Cooper, 2011). When operators perceive situations in the environment, they rely upon cues to position the situations perceived into existing schemas in memory. Any discrepancy in the ‘fit’ between perceived cues and schemas triggers a need to construct interpretations that account for violations or novelty (Maitlis & Christianson, 2014).

Sensemaking typically occurs within sociocultural contexts (Gephart, 1993; Rouleau & Balogun, 2011). Therefore, both individual characteristics and external social factors influence the trajectory and outcome of the sensemaking process. Individual differences in experience, cognitive processing, and emotional states during sensemaking constitute individual characteristics that contribute to the accuracy of the outcomes (Weick et al., 2005). However, these individual characteristics influence sensemaking due to their interaction with external stakeholders (Mills et al., 2010). In this case, sensemaking changes depending upon the characteristics and behaviours of the individuals, who, collectively, may be involved in the sensemaking process (Balogun & Johnson, 2005).

Ultimately, sensemaking yields interpretations of information much like the process of situation assessment (Klein et al., 2006; Weick et al., 2005). As active participants, operators form interpretations of situations in the context of plausible actions with which they can engage (Weick et al., 2005). Therefore, the process is action-oriented such that it prepares operators with the information necessary to initiate a response as accurately and efficiently under the circumstances. It is also part of an ongoing interpretation of a situation that is constructed through the interplay between the operators and the environment, and which is guided by task-related cues and where necessary, social interactions.

Consistent with sensemaking, situation assessment is a process that is inextricably associated with real-world contexts and is characterised by dynamic interactions with environmental features (Klein, 2008). The interpretation of these features draws on schema-based strategies as a means of simplifying and reducing the demands on cognitive resources (Hodgkinson, Bown, & Maule, 1999). In the medical context, these schema-based strategies are known as illness scripts, and are adopted by medical clinicians to assist with patient diagnosis (Charlin, Tardif, & Boshuizen, 2000).

In effect, illness scripts are mental frameworks that incorporate the signs, symptoms, and developmental contexts of different diseases (Kok, Bruin, Robben, & Merriënboer, 2012). During clinical diagnostic reasoning, medical clinicians acquire observable information from patients to form clinical diagnoses of conditions (Bowen, 2006). This involves matching the features observed, including verbal reports, physical manifestations symptoms, and medical history, against the medical clinicians' illness scripts in memory (Coderre, Mandin, Harasym, & Fick, 2003).

Illness scripts provide the working templates to interpret information and/or select responses or medical treatments. If the signs, symptoms, and the developmental contexts of patients match the representations of features within an existing illness script, clinicians are prompted to the information in memory pertaining to the matched illness script, including possible conditions and relevant treatment plans (Charlin et al., 2000; Croskerry, 2009). This reliance on illness scripts is most prevalent amongst experienced clinicians who, in comparison to novice or student clinicians, tend to be more accurate in their diagnoses (Coderre et al., 2003).

Beyond operational domains such as aviation, corporate organisation, and medical practice, situation assessment is also prevalent in daily human activities. For instance, the assessment of products by consumers, and consequently, their purchases, function through a process of matching the features of products against representations or schemas in memory (Henson & Northen, 2000; Roininen, Arvola, & Lähteenmäki, 2006). Differences in quality, reliability, and capability are inferred through the interpretation of features such as brand names, and their correspondence to pre-existing representations (Grunert, 2005; Macdonald & Sharp, 2000).

In summary, situation assessment is a process that likely involves matching environmental features against representations in memory. The outcomes provide operators with workable templates from which to select or construct appropriate action plans. Referred to as sensemaking or diagnostic reasoning in some environments, the fundamental cognitive components and processes appear consistent (Figure 1).

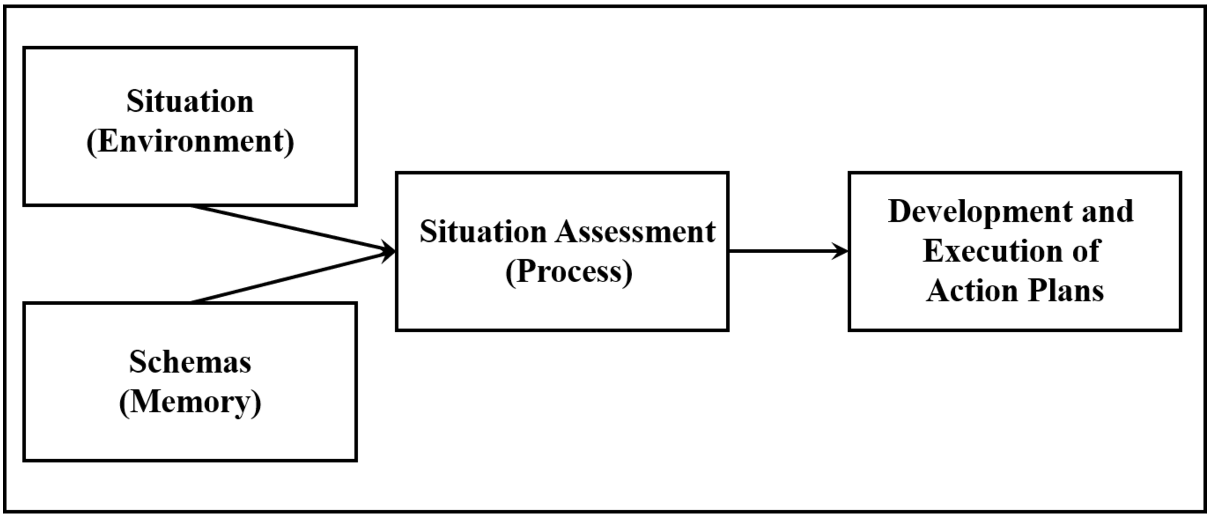

Figure 1. Theoretical model of situation assessment based on Klein (2008). The process of situation assessment relies on the perceived situation in the environment that are matched against schemas in memory to form assessments of the presenting situations on which action plans are developed.

References

Aviation Safety Council. (2016). *Aviation Occurrence Report: 4 February, 2015, TransAsia Airways Flight GE235 ATR72-212A, Loss of Control and Crashed into Keelung River Three Nautical Miles East of Songshan Airport* (ASC-AOR-16-06-001). Retrieved from http://www.asc.gov.tw

Bureau of Enquiry and Analysis for Civil Aviation Safety. (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*. Retrieved from https://www.bea.aero/

Balogun, J., & Johnson, G. (2004). Organizational restructuring and middle manager sensemaking. *The Academy of Management Journal, 46*(4), 523-549. doi:10.5465/20159600

Balogun, J., & Johnson, G. (2005). From intended strategies to unintended outcomes: The impact of change recipient sensemaking. *Organization Studies, 26*(11), 1573-1601. doi:10.1177/0170840605054624

Bowen, J. L. (2006). Educational strategies to promote clinical diagnostic reasoning. *The New England Journal of Medicine, 355*(21), 2217-2225. doi:10.1056/NEJMra054782

Charlin, B., Tardif, J., & Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge. *Academic Medicine, 75*(2), 182-190. doi:10.1097/00001888-200002000-00020

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81. doi:10.1016/0010-0285(73)90004-2

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37*(8), 695-703. doi:10.1046/j.1365-2923.2003.01577.x

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*(8), 1022-1028. doi: 10.1097/ACM.0b013e3181ace703

Engström, I., Gregersen, N. P., Hernetkoski, K., Keskinen, E., & Nyberg, A. (2003). *Young novice drivers, driver education and training: Literature review*. Linköping, Sweden: Swedish National Road and Transport Research Institute.

Gephart, R. P. (1993). The textual approach: Risk and blame in disaster sensemaking. *Academy of management Journal, 36*(6), 1465-1514. doi:10.5465/256819

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104-114. doi:10.1016/j.neuropsychologia.2013.11.010

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision making. *Annual Review of Psychology, 62*(1), 451-482. doi:10.1146/annurev-psych-120709-145346

Goh, J., & Wiegmann, D. A. (2001). Visual flight rules flight into instrument meteorological conditions: An empirical investigation of the possible causes. *The International Journal of Aviation Psychology, 11*(4), 359-379. doi:10.1207/s15327108ijap1104_3

Grunert, K. G. (2005). Food quality and safety: consumer perception and demand. *European Review of Agricultural Economics, 32*(3), 369-391. doi:10.1093/eurrag/jbi011

Henson, S., & Northen, J. (2000). Consumer assessment of the safety of beef at the point of purchase: a pan-European study. *Journal of Agricultural Economics, 51*(1), 90-105. doi:10.1111/j.1477-9552.2000.tb01211.x

Hodgkinson, G. P., Bown, N. J., & Maule, A. J. (1999). Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal, 20*(10), 977-985. doi:10.1002/(SICI)1097-0266(199910)20:10<977::AID-SMJ58>3.0.CO;2-X

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. doi:10.1177/154193120104500411

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Klein, G. A. (2008). Naturalistic decision making. *Human factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G. A., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70-73. doi:10.1109/mis.2006.75

Kok, E. M., Bruin, A. B. H., Robben, S. G. F., & Merriënboer, J. J. G. (2012). Looking in the same manner but seeing it differently: Bottom-up and expertise effects in radiology. *Applied Cognitive Psychology, 26*(6), 854-862. doi:10.1002/acp.2886

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331-352. doi:10.1002/bdm.381

Macdonald, E. K., & Sharp, B. M. (2000). Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. *Journal of Business Research, 48*(1), 5-15. doi:10.1016/s0148-2963(98)00070-8

Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. *The Academy of Management Annals, 8*(1), 57-125. doi:10.1080/19416520.2014.873177

Maitlis, S., & Sonenshein, S. (2010). Sensemaking in crisis and change: Inspiration and insights from Weick (1988). *Journal of Management Studies, 47*(3), 551-580. doi:10.1111/j.1467-6486.2010.00908.x

Maule, A. J., Hockey, G. R., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica, 104*(3), 283-301. doi: 10.1016/S0001-6918(00)00033-0

Mills, J., Thurlow, A., & Mills, A. J. (2010). Making sense of sensemaking: The critical sensemaking approach. *Qualitative Research in Organizations and Management: An International Journal, 5*(2), 182-195. doi:10.1108/17465641011068857

Mueller, A. S., & Trick, L. M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis & Prevention, 48*, 472-479. doi:10.1016/j.aap.2012.03.003

Roininen, K., Arvola, A., & Lähteenmäki, L. (2006). Exploring consumers' perceptions of local food with two different qualitative techniques: Laddering and word association. *Food Quality and Preference, 17*(1-2), 20-30. doi:10.1016/j.foodqual.2005.04.012

Rouleau, L., & Balogun, J. (2011). Middle managers, strategic sensemaking, and discursive competence. *Journal of Management Studies, 48*(5), 953-983. doi:10.1111/j.1467-6486.2010.00941.x

Stokes, A. F., & Raby, M. (1989). Stress and cognitive performance in trainee pilots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 33*(14), 883-887. doi:10.1177/154193128903301404

Weick, K. E. (1995). *Sensemaking in organizations.* Thousand Oaks, CA: Sage.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science, 16*(4), 409-421. doi:10.1287/orsc.1050.0133

Whiteman, G., & Cooper, W. H. (2011). Ecological sensemaking. *The Academy of Management Journal, 54*(5), 889-911. doi:10.5465/amj.2008.0843

Wickens, C. D., Stokes, A. F., Barnett, B., & Hyman, F. (1988). Stress and pilot judgment: an empirical study using midis, a microcomputer-based simulation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 32*(2), 173-177. doi:10.1177/154193128803200238

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 189-197. doi:10.1518/0018720024497871

CHAPTER TWO

Experience and Expertise

In many domains, operators are regarded as experts if they perform consistently to a superior standard on tasks that are representative of the operators' roles within that domain (Ericsson & Charness, 1994). Expertise in sports, for instance, is determined based on clearly defined measures of performance, which typically constitute a single or a small number of observable outcomes taken under standardised conditions (Ericsson, Krampe, & Tesch-Römer, 1993). As cases in point, expert marathon runners record relatively shorter times to complete a race, while expert baseball batters score relatively more home runs during a season (Ericsson & Charness, 1994).

Unlike sporting domains, superior performance in work settings, such as medicine and nursing, tends to be more complex and difficult to capture. This is due largely to the range and complexity of domain-specific tasks that are undertaken, and the variability in definitions of successful performance (Ericsson, 2007; Ericsson, Whyte, & Ward, 2007). For example, successful performance for a nurse might involve ensuring that a patient is as comfortable as possible during palliative care, while successful performance for a train driver might involve adhering to signals during a challenging route where there may be multiple distractions.

In those instances where successful performance is ill-defined, standardised assessments can be used that target a limited number of observable measures that are derived from the tasks undertaken. These are intended to account for differences in the contextual factors during which tasks might be performed (Ericsson et al., 2007). For instance, in the case of expert physicians, the accuracy of their initial diagnoses has been employed as an indicative representation of performance (Ericsson, 2004). Similarly, skilled performance in aeronautical decision making has been determined based, not only on total flight hours, but also the recency of flying experience and the diversity of flight, including the type of aircraft, weather conditions, and terrain (Jensen, 1997).

In the operational context, exposure to, or experience in a domain is often equated to superior performance and capability (Ericsson & Lehmann, 1996; Ericsson et al., 2007). This assumption is not unreasonable since, in comparison to non-experts, experts tend to spend a greater amount of time engaging in deliberate, task-related practice (Baker, Cote, & Abernethy, 2003). Ericsson and Lehmann (1996) argue that domain-specific practice constitutes exposure or experience that, together with active engagement in the performance of a task, facilitates the acquisition of the physical and/or cognitive strategies necessary for superior performance. Consequently, a reasonably strong relationship is inevitable between task-related exposure and performance.

There is little doubt that extensive experience in a domain is a necessary prerequisite for expertise (Butterworth & Reppert, 1960; Ericsson, 2004; Nodine et al., 1999). However, it is also clear that the value of experience lies in the extent to which it corresponds to the desired task. For example, amongst physicians, practitioners with greater domain-related experience in cardiology are significantly more accurate in detecting abnormalities in cardiac rhythms than physicians with less domain-related experience (Butterworth & Reppert, 1960). Similarly, practising mammographers with more than five years of dedicated mammography experience are faster and more accurate in detecting abnormalities in digitised mammograms in comparison to radiology technicians with no mammography experience or radiology residents with little experience of mammograms (Nodine et al., 1999).

The value of experience within a domain lies in the opportunity to acquire detailed and highly refined schemas in memory (Randel, Pugh, & Reed, 1996). This process occurs through repeated and consistent exposure to the co-occurrence of features and/or events or objects across different situations within domain-specific environment. This exposure enables the acquisition of associations in memory that provide the foundation for schemas. It also allows a continuous process of evaluation and testing so that relatively weaker associations can be identified and discarded while stronger associations can be retained and further refined (Kim, Seitz, Feenstra, & Shams, 2009).

The advantage associated with highly refined associations in memory lies in their capacity to enable accurate performance while minimising the demands on cognitive resources (Jitendra & Hoff, 1996; Laxmisan et al., 2007). In the absence of highly developed schemas, operators are required to formulate assessments of situations based on a rudimentary understanding of the context, imposing significant demands on cognitive processing, potentially causing delays, and increasing the likelihood of error and system failures.

Dual-Processing Model of Human Cognition

According to the dual-processing model of human cognition (Evans, 2003, 2008), operators rely on two distinct but complementary cognitive systems to form appropriate assessments of, and respond effectively to everyday situations. Although the two systems have variously been described as Implicit and Explicit, Experiential and Rational, or Intuitive and Analytic, the present programme of research adopts the neutral terminology of System 1 and System 2.

System 1 processing is generally regarded as a more primitive form of processing, characterised by rapid, automatic, and nonconscious processing that requires few cognitive resources (Evans, 2003). The application of schemas during situation assessment constitutes System 1 processing. The matching of perceived stimuli against existing schemas obviates the need for effortful processing, and therefore, situation assessment and response selection can be executed swiftly (Glöckner & Witteman, 2010).

If appropriate schemas are absent, operators typically rely on System 2 processing to form an assessment of a situation. System 2 processing involves the deliberate and analytical processing of information, a process that is effortful and constrained by the limited capacity of working memory (Evans, 2003). Operators usually revert to System 2 processing when they are unfamiliar with the features and/or events or objects within the presenting environment.

At the early stages of learning, operators must inevitably rely on System 2 processing to actively and deliberately identify relevant features to form working assessments of a situation (Ericsson et al., 1993). However, as operators accumulate experience, they are exposed to a range of situations. Through a process of associative learning, perceived features and events become paired in memory to form feature-event/object associations (Mitchell, Houwer, & Lovibond, 2009). Eventually, these associations are assimilated or accommodated into existing domain-specific schemas which form the functional mechanisms of System 1 processing.

Heuristics and Biases

During the course of information processing, System 1 and System 2 are applied successively depending upon the familiarity of the situation (Evans, 2008; Klein, 2008). For instance, in clinical diagnosis, physicians may rely initially on System 1 processing to diagnose patient's condition rapidly by matching visual, auditory, and tactile symptoms to existing schemas in memory (Croskerry, 2009). However, if no match is identified, System 2 processing is engaged to consider the pattern of symptoms presented, and generate and test working hypotheses of the patient's condition.

One of the chief criticisms associated with System 1 processing is that it obviates the requirement for conscious processing (Sloman, 1996). This lack of conscious engagement was strongly associated with errors in reasoning and the opportunity for bias (Kahneman, 2003), leading to an extensive programme of research that was broadly intended to identify strategies that would encourage System 2 processing, even in seemingly familiar contexts (Baron, 2012; Critchlow, 1987). The interest in System 2 processing is captured in normative models of decision-making, including Expected-Utility Theory (Edwards, 1954; Friedman & Savage, 1952)

Expected-Utility Theory is based on the proposition that optimal decision-making is a compensatory choice with the least risk and greatest outcome (Friedman & Savage, 1952). The theory is based on the proposition that identifying the optimal choice involves evaluating the utility and probability of all of the possible outcomes and selecting the behaviour or option that has the highest expected utility. This process is necessarily based on the assumption that operators will apply System 2 processing since, in the given situation, all of the possible options must be examined. However, this exhaustive process of information acquisition and processing can impose significant demands on limited cognitive resources.

Normative models are often derived from philosophical or mathematical arguments which do not necessarily reflect the demands of real-world decision-making (Fishburn, 1988). In real-world situations, the normative risk assessment may not evaluate a low-probability yet disastrous event as having the highest expected utility, and therefore, rendering the outcome of System 2 processing as irrational in the context of the event that occurs. Consequently, any training schemes and support systems that are designed based on normative models can lack practical relevance or validity since they simplify decision problems, and fail to account for individual differences, and the dynamic interactions between stakeholders and systems that characterise real-world operational settings (Simon, 1979).

In response to the perceived limitations of normative models as both instructions and explanatory tools, alternative models have been considered based primarily on a descriptive-based approach to decision-making (Meso, Troutt, & Rudnicka, 2002; Schwarz, Bless, Strack, & Klumpp, 1991). Tversky and Kahneman (1974) introduced a descriptive-based approach to human decision-making, which modelled decision-making as a function of naturally occurring heuristics and biases.

Heuristics are cognitive strategies in the form of rules or IF-THEN statements that simplify information processing by disregarding aspects of perceived information to derive faster decision outcomes (Gigerenzer & Gaissmaier, 2011; Keren & Teigen, 2004). Since they function as a trade-off between the speed of processing and the accuracy of an outcome, there are situations in which the process yields less accurate outcomes, particularly where relevant information is masked or misinterpreted (Gigerenzer & Gaissmaier, 2011). When the trade-off between accuracy and effort results in substandard accuracy, bias is said to have occurred.

In controlled laboratory studies, Tversky and Kahneman (1973) observed that participants tended to rely on heuristics during decision-making. For instance, when presented with the names of famous and non-famous individuals, participants were more likely to recall names of famous individuals than non-famous individuals. Further, when asked to judge whether there were more male or female names on a presented list, participants were likely to perceive a greater frequency of males where the male names were famous, despite there being fewer male names on the list. A similar effect was evident in the case of female names.

In the case of judgements of names on a list, participants appeared to rely on the availability heuristic, where events are rated more frequently depending upon the ease with which instances can be recalled from memory (Schwarz et al., 1991; Tversky & Kahneman, 1973). Although this may be a reliable assumption in some circumstances, the application of the availability heuristic to form frequency judgments can, on occasion, result in biased outcomes, reflected in inaccurate judgements.

According to Gigerenzer (2008), heuristics constitute the broad associations in memory that form the foundations of schemas. In familiar situations, they provide the templates to formulate assessments simply and swiftly (Svenson, 1979). When perceptual information is consistent with heuristics in memory, operators are able to form rapid judgments based on the rules that are encapsulated in the matched heuristics.

In one sense, heuristics can be considered a cognitive strategy that may involve the application of domain-specific schemas and other broad IF-THEN rules. Where operators possess in memory, relevant schemas established through domain-specific experience, the application of heuristics through System 1 processing is likely to result in effective decision outcomes and task performance (Dane, Rockmann, & Pratt, 2012). In contrast, the application of heuristics in the absence or limited availability of relevant schemas is likely to result in biases and ineffective performance.

In real-life operational contexts, the swift and accurate performance of expert operators perhaps reflects the application of existing schemas in memory through the application of intuitive, System 1 processing. In this regard, the descriptive-based approach of Naturalistic Decision-Making (NDM) represents a relevant explanation of this operational decision-making by means of System 1 processing. The NDM recounts the decision-making strategies engaged by expert operators in specific operational settings, and the interplay between System 1 and System 2 during situation assessment in real-life operational contexts.

Naturalistic Decision-Making

The NDM framework first emerged at a conference sponsored by the Army Research Institute in response to the lack of functional models of decision-making that can be applied in the context of real-world operational settings (Lipshitz, Klein, Orasanu, & Salas, 2001). These real-world settings involve time pressure, uncertain goals, risk, and dynamic conditions, where there are significant consequences for poor performance (Lipshitz & Strauss, 1997). NDM is a framework, the intention of which is to describe the cognitive strategies adopted by expert operators to achieve superior and effective decision performance in response to challenging situations (Klein, 2008; Lipshitz, Klein, & Carroll, 2006).

By targeting expert performance, there is an assumption that strategies can be identified to improve the performance of less experienced operators. A case in point is the observation that experts rarely generate and compare all of the possible assessments and responses prior to initiating a response (Kaempf, Klein, Thordsen, & Wolf, 1996). Instead, the approach is rapid, often nonconscious, and non-compensatory, suggesting that experts are drawing on a repertoire of highly specialised schemas in assessing situations.

Like the descriptive approach to decision-making, the goal of the NDM framework is to identify and describe the procedural structures that are engaged during the decision-making process (Lipshitz, 1994). No predictions are made as to which option might be selected in a given instance since the schemas on which decisions are based are idiosyncratic and matched to specific situations. Consequently, the principal implication of the NDM framework is the importance ascribed to previous domain-specific experiences in the context within which a decision might be made.

While the NDM framework incorporates various models that are intended to capture real-world experience, few of these models explain the process of situation assessment, despite the fact that it constitutes a critical precursor to the activation of appropriate schemas and the subsequent resolution of a problem. An exception is the Recognition-Primed Decision Model (RPD) which incorporates situation assessment as a necessary stage of problem identification, prior to resolution (Klein, 2008).

Recognition-Primed Decision Model

Consistent with the NDM framework, the RPD model is a descriptive account of expert decision-making in complex, dynamic and time-constrained settings (Klein et al., 1986; Klein, Calderwood, & Clinton-Cirocco, 2010). The initial model was developed from cognitive task analyses of experienced fire ground commanders who were formulating decisions under uncertainty and time pressure (Klein et al., 1986). Twenty-six firefighters with an average experience of 23.2 years were interviewed to derive decision responses in 156 highly challenging situations.

In most of these demanding incidents, experienced fire ground commanders relied on their experience to assess the situation, and derive the most typical course of action. For previous experience to support situation assessment in a specific context, the RPD model presupposes that information derived from previous experience is stored in memory in the form of schemas.

These schemas are characterised by multiple links between units of information (Ghosh & Gilboa, 2014). These units of information include representations of features that comprise the environment, and events or objects that occur within domain-specific situations. For instance, urban fire grounds are typically characterised by features such as fire and smoke, and events or objects such as trapped victims and unstable infrastructure (Okoli, Watt, & Weller, 2017).

During situation assessment, features in the environment are perceived and matched against representations of those features in memory (Kaempf et al., 1996; Klein, 2008). When a match occurs, the RPD model would predict that events or objects relevant to the matched feature are retrieved as a template consisting of several elements, including expectations, plausible goals, and previously successful responses (Klein, 2008). When perceived features result in the retrieval of templates, they are known as cues.

In real-world settings, and particularly high-risk operations, a successful outcome is determined by both the accuracy and the speed of situation assessment. When information is interpreted accurately, operators can swiftly initiate responses to changes in the system state. Should the initial responses be ineffective, operators then reassess the situation, adapting their responses as part of a continuous, cyclical process (Lipshitz et al., 2001).

In the case of fire fighters, features related to smoke on a fire ground comprise critical information that must be interpreted accurately as part of the process of situation assessment (Okoli et al., 2017). For example, the colour, movement, and/or location of the smoke (features) are associated with the severity, progression, and/or volatility of the fire (events). The specificity and accuracy of these feature-event/object associations in memory assure both the safety of fire fighters and the most efficient and effective means of attacking the fire.

Given that experience is a necessary condition for the acquisition of feature-event/object associations, active practice within the domain-specific environment should be associated with the accuracy and specificity of feature-event/object associations retained in memory. In turn, a greater repertoire of accurate, precise feature-event/object associations that comprise domain-specific schemas should be associated with accurate and timely task performance (Chase & Simon, 1973; Mann, Williams, Ward, & Janelle, 2007; McPherson, 1999).

Evidence to support the association between experience and task performance can be drawn from a number of domains including chess (Chase & Simon, 1973), driving (Underwood, Crundall, & Chapman, 2002), and medicine (Nodine et al., 1999). However, the relationship between experience and performance is non-linear (Crane et al., 2018; Loveday, Wiggins, Searle, Festa, & Schell, 2012; Todd & Thomas, 2012), suggesting that there are individual differences in both the rates of exposure to high quality and meaningful experiences, and the capacity to derive feature-event/object associations necessary to form schemas in memory. In many domains, including aviation (Todd & Thomas, 2012) and driving (Duncan, Willia, & Brown, 1991), the duration over which experience is acquired is insufficient to differentiate operator performance. Therefore, mere exposure to the problem environment may be a necessary but insufficient requirement for the acquisition and retention of feature-event/object associations. Operators need to possess an intrinsic capability to identify critical features and associated events or objects, form associations between features and events or objects to form cues, retain these cue-based associations in memory, and identify opportunities for the utilisation of these cues in practice (Wiggins, 2015).

Importance of Cues in Performance Outcomes

Behaviour indicative of the utilisation of cues has been associated with improved task performance in a number of operational domains, including electrical system power diagnosis (Loveday, Wiggins, Harris, O'Hare, & Smith, 2012), pre-flight and inflight decision-making (Wiggins, Azar, Hawken, Loveday, & Newman, 2014), and paediatric intensive care (McCormack, Wiggins, Loveday, & Festa, 2014). In fast-paced games like squash and racquet ball, the availability and utilisation of cues provide players with a distinct advantage in anticipating the location of an opponent's shot (Abernethy, 1990; Abernethy, Gill, Parks, & Packer, 2001; Mann et al., 2007). Similarly, in the context of power transmission, operators who demonstrate behaviours associated with the utilisation of cues are more accurate in diagnosing the causes of a fault (Loveday, Wiggins, Harris, et al., 2012).

The difference in performance associated with the utilisation of cues can be ascribed to an associated reduction in cognitive load (Brouwers, Wiggins, Griffin, Helton, & O'Hare, 2017; Brouwers, Wiggins, Helton, O'Hare, & Griffin, 2016). For example, Brouwers et al. (2017) demonstrated a relationship between the utilisation of cues and sustained attention, a task that consumes cognitive resources. They categorised participants into two groups based on behaviours that were associated with greater or lesser utilisation of cues, and noted that the degredation of performance amongst participants who utilised cues to a relatively lesser extent was greater than the rate at which performance degraded amongst participants who utilised cues to a relatively greater extent.

The relatively slower degradation in task performance amongst individuals who utilised cues to a relatively greater extent suggests that cues may enable more rapid processing since they obviate the requirement to process large amounts of information during situation assessment (Rieskamp & Hoffrage, 2008). Since less information is processed within a specified period, the rate at which performance degrades is reduced. As might be predicted, this suggests that one of the advantages afforded by greater utilisation of cues is a greater efficiency in information processing and a consequent reduction in the demands for cognitive resources.

Cue Utilisation

The utility of cue utilisation in generating superior performance outcomes highlights the importance of accurately identifying and assessing individual variations in the extent to which cues are utilised during situation assessment. In doing so, more accurate predictions can be made in discriminating potential candidates or operators. Cue utilisation has been the subject of considerable interest in the literature, with a range of approaches having been adopted to examine the underlying cognitive and perceptual mechanisms, including prescriptive, descriptive, and behavioural approaches (Balatsoukas et al., 2013; Hursch, Hammond, & Hursch, 1964; Wiggins, 2012).

The prescriptive approach is laboratory-based, and involves the acquisition and application of probabilistic relations between artificial features and events (Hursch et al., 1964). Although the assessment is conducted within laboratories, the incorporation of probability is intended to simulate the learned associations that occur within real-world contexts. This includes uncertainty and complex interactions so that the relationships between features and events or objects are not absolute across any situation (Bernieri, Gillis, Davis, & Grahe, 1996; Brunswik, 1955). As a consequence, performance in this case is likely to reflect individual differences in integrating and identifying the relevance of multiple cues to anticipate the situations presented.

The Multiple-Cue Probability Learning (MCPL) framework is a paradigm that has been widely applied to examine the basic structure of probabilistic learning in judgment and decision-making (Chasseigne, Mullet, & Stewart, 1997; Orquin, 2014; Rolison, Evans, Dennis, & Walsh, 2012; York, Doherty, & Kamouri, 1987). MCPL studies are based on the theoretical framework of Brunswik's Lens Model, which describes the underlying processes governing the operators' interaction with the environment during decision-making (Brunswik, 1955). Figure 1 illustrates the linear relationship between perceptual features or cues and the actual state of situations.

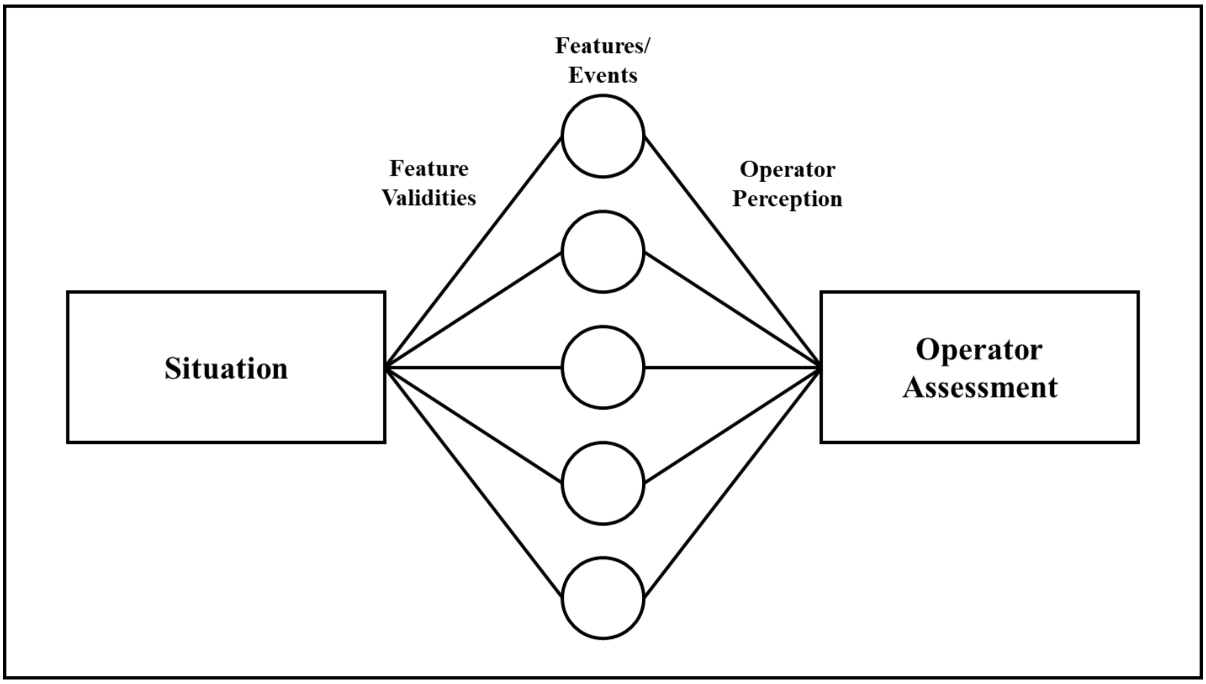

Figure 1. A schematic representation of Brunswik's Lens Model illustrating the linear relationship between the actual state of situations (left) and operators' assessments derived from perceptual information (right).

The actual state, represented on the left side of the model, is related to the operators' perception, represented on the right side of the model, via the lens in the centre, which consists of features that comprise the environment. The associations between the features and the actual state of the environment constitute the validities of the actual feature-event/object associations, which comprise the contributory effects of each feature to the actual state.

During situation assessment, operators assess the features in the environment against feature-event/object associations in memory to form an assessment of the actual state. Where a feature in the environment triggers the construction of an assessment based on a feature-event/object association in memory, it functions as a cue for the operator. The validity of these acquired associations in relation to the actual associations determines the accuracy of the operators' assessment, with a higher correlation generally equating to greater accuracy (Brehmer, 1972; Hursch et al., 1964).

The interactions between operators and their environments is limited by the operators' perceptual constraints such that the actual state of the situations must be inferred from perceptual cues (Doherty & Kurz, 1996). This need for inference gives rise to uncertainty in the perceived relationship between the cues and the actual state. Consequently, the effectiveness of assessments is dependent upon the intrinsic predictability of the situations based on the features, and the extent to which the operators' subjective perception of the cues matches the relationships actually reflected between features and situations (Summers & Hammond, 1966).

The standard MCPL task requires participants to formulate predictions of outcomes based on a set of, typically, four cues (Knowlton, Squire, & Gluck, 1994). Outcomes are generally binary, and the strength of associations varies between each cue and each of the two outcomes. Cues can be utilised in combination with other cues so that the predictive probability of the pattern of cues is based on the combined probabilities of all of the cues present in the pattern.

On each trial, participants are presented with a pattern of cues, and provided with feedback, having made their predictions. Over a series of trials, participants demonstrated a gradual improvement in performance, reflecting a refinement in the match between the probabilities of the patterns of features to the actual outcomes and the participants' subjective perceptions of those cues (Castellan Jr., 1973; Rolison et al., 2012).

As a measure of cue utilisation, the prescriptive approach, in some respects, lacks ecological validity. Despite efforts to simulate real-world cues, the rigid representations of cue validities as static associations overlooks the importance of dynamic and changing situations, together with the adaptability of individual perception and cognition (Speekenbrink & Shanks, 2010). Therefore, a descriptive approach to cue utilisation may serve as a more accurate representation of cue utilisation within ecologically valid conditions.

The descriptive approach to the evaluation of cue utilisation captures differences in behavioural measures that correspond directly to the acquisition and application of cues within natural settings. Observable differences in cue utilisation have been captured predominantly in visual search behaviours, and to some extent, subjective reports during task performance (Balatsoukas et al., 2013; Beanland, Lenné, Salmon, & Stanton, 2015; Gegenfurtner, Lehtinen, & Säljö, 2011; Stein & Brennan, 2004). Under the presumption that better performers and/or more experienced operators utilise cues to a greater extent, these responses to domain-specific tasks are evaluated against performance outcomes and/or experience as a means of describing retrospectively, the rate at which cues may have been acquired and utilised during task performance.

Subjective reports, in the form of think-aloud protocols, is a common assessment method to describe the cognitive processes that underpin cue utilisation (Cotton & Gresty, 2006). In a standard think-aloud assessment, participants undertake domain-specific tasks within the natural work environment while verbally stating aloud their thoughts as they undertake the tasks (Fonteyn, Kuipers, & Grobe, 1993). However, verbally expressing cognitive processes appears to influence behavioural performance insofar as task performance under think-aloud assessments differs from performance under standard observations or retrospective verbal reports (Aitken, Marshall, Elliott, & McKinley, 2011; Kuusela & Paul, 2000).

Subjective reports are also limited to information available consciously to the operators. Since the utilisation of cues is reliant on System 1 processing which functions on nonconscious cognitive and perceptual processes, behavioural measures that rely on verbal reports may not capture the processes that occur in the absence of conscious awareness.

Observations of subliminal behaviours that relate to the appraisal of situations is a descriptive approach, which obviates the need to rely on verbal reports. In many domains, including medicine, sports, automotive, and aviation, eye-tracking is widely used as a surrogate measure for the visual-cognitive aspects governing cue utilisation during task performance (Kato & Fukuda, 2002; Reina, Moreno, & Sanz, 2007; Schriver, Morrow, Wickens, & Talleur, 2008; Tien et al., 2014). Visual search measures, including fixation, saccades, dwell time, and scan path, reflect varying characteristics of visual attention control, and provides insights into the features that the operators perceive as important during task performance (Al-Moteri, Symmons, Plummer, & Cooper, 2017).

Amongst baseball players, for example, visual fixations of more experienced players tend to cluster around the shoulder and torso of the pitcher, while less experienced players tend to exhibit patterns of visual fixations that are dispersed across the entire body of the pitcher (Kato & Fukuda, 2002). This observation suggests that more experienced players are more targeted in their visual search in comparison to less experienced players. Further, less experienced players lack direction in their visual search behaviours, suggesting that they are unable to identify relevant features.

Differences in visual search behaviours between more and less experienced operators are also apparent in aviation (Doane, Sohn, & Jodlowski, 2004). For example, more experienced pilots spent more time fixating on more relevant cues pertaining to the diagnosis of aircraft complications compared to less experienced pilots (Schriver et al., 2008). Given that more experienced pilots exhibit more accurate decision-making outcomes than less experienced pilots, it might be argued that the improved outcomes are associated with improved attentional processes in the form of more accurate cue selection.

The manner in which operators adapt their visual processing to various task complexities also differs between more or less experienced operators (Bruder, Eißfeldt, Maschke, & Hasse, 2013). For instance, in a more complex and demanding situation, such as driving on a dual-carriageway, more experienced drivers tend to exhibit more extensive visual search behaviours compared to less experienced drivers (Crundall & Underwood, 1998; Falkmer & Gregersen, 2001). This observation persists even when the demands associated with the physical control of the vehicle is removed, suggesting that less experienced drivers may have not acquired sufficient schemas in memory to identify relevant cues to monitor a demanding situation (Underwood, Chapman, Bowden, & Crundall, 2002).

Given that many operational tasks are performed on the basis of visual information, visual search behaviours provide useful observations of information acquisition during decision-making. However, the descriptive approach only presents behavioural observations as to which inferences of visual cue processing can be made. While these behaviours may appear to be relevant, they are not necessarily reflective of cue utilisation as inferences must be formulated post-hoc.

Considering the limitations of the prescriptive and descriptive approaches, the behavioural approach to assessing cue utilisation is based on the assumption that a psychological construct can be inferred based on the behavioural representations relevant to the underlying cognitive functions (Wiggins, 2012). In effect, it posits that differences in cue utilisation can be established by assessing the utilisation of cues in a given context and accounting for differences in experience or exposure. It bridges the gap between the prescriptive and descriptive approaches by instituting theoretically-based structures to the study of cue utilisation without compromising the ecological validity of responses.

There are five behavioural measures that are thought to be relevant to the assessment of cue utilisation, including the capacity to rapidly identify key features, the capability to accurately recognise key features under time-constraint, the ability to rapidly and precisely differentiate features, the capability to discriminate relevant from less relevant features, and the capacity to prioritise the acquisition of key features during problem orientation (Wiggins, 2012). Behaviours in response to these tasks are presumed to reflect a range of processes attributable to the utilisation of cues, and are consistent with the key characteristics of situation assessment evident in the Recognition-Primed Decision-Model.

Firstly, when presented with a problem situation, operators with higher cue utilisation tend to be faster in responding to problem scenarios in comparison to operators with lower cue utilisation (Loveday, Wiggins, Festa, Schell, & Twigg, 2013). Similarly, operators with higher cue utilisation display greater attentional control and direct their visual attention towards more relevant cues in comparison to operators with lower cue utilisation (Schriver et al., 2008).

Secondly, under time pressure, operators with higher cue utilisation demonstrate a greater capacity to sustain a similar rate of response without compromising response accuracy (Lorains, Ball, & MacMahon, 2013; Schriver et al., 2008). Differences in performance under time pressure elucidate the superior processing efficiency afforded by higher cue utilisation. Given that cue utilisation relies on System 1 processing to construct assessments, operators with higher cue utilisation appear less susceptible to the time-performance trade-offs. By contrast, operators with lower cue utilisation are forced to compromise performance, given the time constraint, to undertake the more effortful, time-consuming System 2 processing.

Thirdly, differences in processing efficiency between operators with higher and lower cue utilisation are evident in the precision and methods of processing features in the environment. For example, operators with higher cue utilisation tend to be more precise in distinguishing more relevant pairs of features and events or objects from less relevant pairs during situation assessment (Morrison, Wiggins, Bond, & Tyler, 2013). By contrast, operators with lower cue utilisation are less precise in their differentiation of more from less relevant associations. This capacity to more precisely discern the relevance of feature-event/object associations supports the retrieval of more appropriate assessments in response to features in the environment.

Fourthly, operators with higher cue utilisation are more sensitive to the differences between relevant and irrelevant features in comparison to operators with lower cue utilisation. According to the information-reduction hypothesis (Haider & Frensch, 1999), superior task performance is influenced by the capacity to identify and process task-relevant features and set aside task-irrelevant features. As a result, operators can allocate attentional resources to more relevant features, thereby conserving resources during decision-making.

When exposed to similar features, operators with higher cue utilisation are more accurate than operators with lower cue utilisation in differentiating relevant from irrelevant features (Pauley, O'Hare, & Wiggins, 2009; Weiss & Shanteau, 2003). Arguably, operators with higher cue utilisation rely, to a greater extent, on existing feature-event/object associations in memory to recognise relevant features from irrelevant features.

Finally, there is evidence to suggest that operators with higher cue utilisation tend to adopt a strategy of information acquisition that is more efficient and directed compared to operators with lower cue utilisation (Wiggins & O'Hare, 1995). This observation suggests that higher cue utilisation is associated with a more effective method of prioritising the acquisition of information during problem-solving.

The behavioural approach to the assessment of cue utilisation is dependent upon the evaluation of the operators' performance on all five behavioural measures. The standard approach to assessing cue utilisation is by establishing cue utilisation typologies (Loveday, Wiggins, & Searle, 2013). These typologies are derived from the combination of performance on all five tasks to produce distinct patterns of performance.

Higher cue utilisation is generally characterised by relatively higher performance on all five behavioural measures (Brouwers, Wiggins, & Griffin, 2018; Watkinson, Bristow, Auton, McMahon, & Wiggins, 2018; Wiggins, Griffin, & Brouwers, 2019). This is demonstrated in rapid responses in identifying key features, greater accuracy in recognising key features, greater precision in differentiating between associations more rapidly, greater discrimination between relevant and irrelevant features, and a more efficient acquisition of key features. The distinguishing pattern of performance of higher and lower cue utilisation typologies has been validated against performance outcomes in various domains including aviation (Wiggins et al., 2014), water safety (Wiggins et al., 2019), and clinical audiology (Watkinson et al., 2018).

Differences in cue utilisation during situation assessment are, however, only partly related to their previous experience (Loveday, Wiggins, Searle, et al., 2012). For example, controlling for experience, operators who demonstrated behaviours consistent with higher cue utilisation in the context of power system control, outperformed operators with lower cue utilisation (Loveday, Wiggins, Harris, et al., 2012; Loveday, Wiggins, Searle, et al., 2012). The observation that individual differences in the utilisation of cues influences performance outcomes, independent of experience, suggests that the cue utilisation is not a passive by-product of the accumulation of experience and feature-event/object associations in memory. Rather, it may be governed by a latent trait that regulates the acquisition of cues. Therefore, the acquisition of cues may constitute an independent construct that varies across individuals. This individual difference in cue acquisition may contribute to differences in the accuracy of and rate at which feature-event associations are acquired and integrated into existing schemas to support cue utilisation during situation assessment.

Ultimately, cue utilisation may be the end result of a long process of cue acquisition. It reflects the presence of schemas that comprise relevant feature-event/object associations in memory, and which support the pattern matching process necessary for timely and accurate situation assessment. However, what remains unclear are the cognitive mechanisms that enable the acquisition of cues, and whether these differ for different operators depending upon individual capabilities and the quality and quantity of exposure to the operational environment within which feature-event/object relationships are expected to be applied.

References

Abernethy, B. (1990). Anticipation in squash: differences in advance cue utilization between expert and novice players. *Journal of Sports Science, 8*(1), 17-34. doi:10.1080/02640419008732128

Abernethy, B., Gill, D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. *Perception, 30*(2), 233-252. doi:10.1068/p2872

Aitken, L. M., Marshall, A., Elliott, R., & McKinley, S. (2011). Comparison of 'think aloud' and observation as data collection methods in the study of decision making regarding sedation in intensive care patients. *International Journal of Nursing Studies, 48*(3), 318-325. doi:10.1016/j.ijnurstu.2010.07.014

Al-Moteri, M., Symmons, M., Plummer, V., & Cooper, S. (2017). Eye tracking to investigate cue processing in medical decision-making: A scoping review. *Computers in Human Behavior, 66*, 52-66. doi:10.1016/j.chb.2016.09.022

Baker, J., Cote, J., & Abernethy, B. (2003). Sport-specific practice and the development of expert decision-making in team ball sports. *Journal of Applied Sport Psychology, 15*(1), 12-25. doi: 10.1080/10413200305400

Balatsoukas, P., Ainsworth, J., Williams, R., Carruthers, E., Davies, C., McGrath, J., . . . Buchan, I. (2013). Verbal protocols for assessing the usability of clinical decision support: the retrospective sense making protocol. *Stud Health Technol Inform, 192*, 283-287. doi: 10.3233/978-1-61499-289-9-283

Baron, J. (2012). The point of normative models in judgment and decision making. *Frontiers in Psychology, 3*, 577. doi:10.3389/fpsyg.2012.00577

Beanland, V., Lenné, M. G., Salmon, P. M., & Stanton, N. A. (2015). Variability in decision-making and critical cue use by different road users at rail level crossings. *Ergonomics, 59*(6), 1-13. doi:10.1080/00140139.2015.1095356

Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology, 71*(1), 110-129.

Brehmer, B. (1972). Cue utilization and cue consistency in multiple-cue probability learning. *Organizational Behavior and Human Performance, 8*(2), 286-296. doi:10.1016/0030-5073(72)90051-7

Brouwers, S., Wiggins, M., & Griffin, B. (2018). Operators who readily acquire patterns and cues, risk being miscued in routinized settings. *Journal of Experimental Psychology: Applied, 24*(2), 261-274. doi:10.1037/xap0000151

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics, 60*(11), 1500-1515. doi:10.1080/00140139.2017.1330494

Brouwers, S., Wiggins, M. W., Helton, W. S., O’Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. *Frontiers in Psychology, 7*, 435. doi:10.3389/fpsyg.2016.00435

Bruder, C., Eißfeldt, H., Maschke, P., & Hasse, C. (2013). Differences in monitoring between experts and novices. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 295-298. doi:10.1177/1541931213571065

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*(3), 193. doi:10.1037/h0047470

Castellan Jr., N. J. (1973). Multiple-cue probability learning with irrelevant cues. *Organizational behavior and human performance, 9*(1), 16-29. doi: 10.1016/0030-5073(73)90033-0

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81. doi:10.1016/0010-0285(73)90004-2

Chasseigne, G., Mullet, E., & Stewart, T. R. (1997). Aging and multiple cue probability learning: The case of inverse relationships. *Acta Psychologica, 97*(3), 235-252. doi:10.1016/s0001-6918(97)00034-6

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology, 37*(1), 45-54. doi:10.1111/j.1467-8535.2005.00521.x

Crane, M. F., Brouwers, S., Wiggins, M. W., Loveday, T., Forrest, K., Tan, S. G. M., & Cyna, A. M. (2018). “Experience isn't everything”: How emotion affects the relationship between experience and cue utilization. *Human factors, 60*(5), 685–698. doi:10.1177/0018720818765800

Critchlow, B. (1987). A utility analysis of drinking. *Addictive Behaviors, 12*(3), 269-273. doi:10.1016/0306-4603(87)90038-4

Croskerry, P. (2009). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in Health Sciences Education, 14*(Suppl 1), 27-35. doi:10.1007/s10459-009-9182-2

Crundall, D. E., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics, 41*(4), 448-458. doi:10.1080/001401398186937

Dane, E., Rockmann, K. W., & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes, 119*(2), 187-194. doi:10.1016/j.obhdp.2012.07.009

Doane, S. M., Sohn, Y., & Jodlowski, M. T. (2004). Pilot ability to anticipate the consequences of flight actions as a function of expertise. *Human factors, 46*(1), 92-103. doi: 10.1518/hfes.46.1.92.30386

Doherty, M. E., & Kurz, E. M. (1996). Social judgement theory. *Thinking & Reasoning, 2*(2-3), 109-140. doi:10.1080/135467896394474

Duncan, J., Willia, Y., & Brown, I. (1991). Components of driving skill: Experience does not mean expertise. *Ergonomics, 34*(7), 919-937. doi:10.1080/00140139108964835

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*(4), 380. doi:10.1037/h0053870

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic medicine : journal of the Association of American Medical Colleges, 79*(10 Suppl), 70-81. doi:10.1097/00001888-200410001-00022

Ericsson, K. A. (2007). An expert-performance perspective of research on medical expertise: the study of clinical performance. *Medical Education, 41*(12), 1124-1130. doi:10.1111/j.1365-2923.2007.02946.x

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747. doi:10.1037/0003-066X.49.8.725

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. doi: 10.1037/0033-295X.100.3.363

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305. doi:10.1146/annurev.psych.47.1.273

Ericsson, K. A., Whyte, J. I. V., & Ward, P. (2007). Expert performance in nursing. *Advances in Nursing Science, 30*(1), E58-E71. doi:10.1097/00012272-200701000-00014

Evans, J. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in cognitive sciences, 7*(10), 454-459. doi:10.1016/j.tics.2003.08.012

Evans, J. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*(1), 255-278. doi:10.1146/annurev.psych.59.103006.093629

Falkmer, T., & Gregersen, N. P. (2001). Fixation patterns of learner drivers with and without cerebral palsy (CP) when driving in real traffic environments. *Transportation Research Part F: Psychology and Behaviour, 4*(3), 171-185. doi:10.1016/S1369-8478(01)00021-3

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research, 3*(4), 430-441. doi:10.1177/104973239300300403

Friedman, M., & Savage, L. J. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy, 60*(6), 463-474.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*(4), 523-552. doi:10.1007/s10648-011-9174-7

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104-114. doi:10.1016/j.neuropsychologia.2013.11.010

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science, 3*(1), 20-29. doi:10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451-482. doi:10.1146/annurev-psych-120709-145346

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning, 16*(1), 1-25. doi:10.1080/13546780903395748

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 172. doi:10.1037/0278-7393.25.1.172

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review, 71*(1), 42-60. doi:10.1037/h0041729

Jensen, R. S. (1997). The boundaries of aviation psychology, human factors, aeronautical decision making, situation awareness, and crew resource management. *The International Journal of Aviation Psychology, 7*(4), 259-267. doi: 10.1207/s15327108ijap0704_1

Jitendra, A. K., & Hoff, K. (1996). The effects of schema-based instruction on the mathematical word-problem-solving performance of students with learning disabilities. *Journal of Learning Disability, 29*(4), 422-431. doi:10.1177/002221949602900410

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*(9), 697-720. doi:10.1037/0003-066X.58.9.697

Kato, T., & Fukuda, T. (2002). Visual search strategies of baseball batters: eye movements during the preparatory phase of batting. *Perceptual and Motor Skills, 94*(2), 380-386. doi:10.2466/pms.2002.94.2.380

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Klein, G. A. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society Annual Meeting, 30*(6), 576-580. doi:10.1177/154193128603000616

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making, 4*(3), 186-209. doi:10.1518/155534310x12844000801203

Klein, G. A., Pliske, R., Crandall, B., & Woods, D. D. (2005). Problem detection. *Cognition, Technology & Work, 7*(1), 14-28. doi:10.1007/s10111-004-0166-y

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106-120. doi:10.1101/lm.1.2.106

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *The American Journal of Psychology, 113*(3), 387. doi:10.2307/1423365

Laxmisan, A., Hakimzada, F., Sayan, O. R., Green, R. A., Zhang, J., & Patel, V. L. (2007). The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. *International Journal of Medical Informatics, 76*(11-12), 801-811. doi:10.1016/j.ijmedinf.2006.09.019

Lipshitz, R. (1994). Decision making in three modes. *Journal for the Theory of Social Behaviour, 24*(1), 47-65. doi:10.1111/j.1468-5914.1994.tb00246.x

Lipshitz, R., Klein, G., & Carroll, J. S. (2006). Introduction to the special issue. Naturalistic decision making and organizational decision making: exploring the intersections. *Organization Studies, 27*(7), 917-923. doi:10.1177/0170840606065711

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331-352. doi:10.1002/bdm.381

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes, 69*(2), 149-163. doi:10.1006/obhd.1997.2679

Lorains, M., Ball, K., & MacMahon, C. (2013). Expertise differences in a video decision-making task: Speed influences on performance. *Psychology of Sport and Exercise, 14*(2), 293-297. doi:10.1016/j.psychsport.2012.11.004

Loveday, T., Wiggins, M., & Searle, B. J. (2013). Cue utilization and broad indicators of workplace expertise. *Journal of Cognitive Engineering and Decision Making, 8*(1), 98-113. doi:10.1177/1555343413497019

Loveday, T., Wiggins, M. W., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In L. C. P., S. J., & F. A. (Eds.), *Pattern*

Recognition - Applications and Methods (Vol. 204, pp. 1-11): Springer, Berlin, Heidelberg.

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2012). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 90-107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2012). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 125-137. doi:10.1177/0018720812448475

Mann, D. T. Y., Williams, M. A., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology, 29*(4), 457-478. doi:10.1123/jsep.29.4.457

McCormack, C., Wiggins, M. W., Loveday, T., & Festa, M. (2014). Expert and competent non-expert visual cues during simulated diagnosis in intensive care. *Frontiers in Psychology, 5*, 949. doi:10.3389/fpsyg.2014.00949

McPherson, S. L. (1999). Expert-novice differences in performance skills and problem representations of youth and adults during tennis competition. *Research Quarterly for Exercise and Sport, 70*(3), 233-251. doi:10.1080/02701367.1999.10608043

Meso, P., Troutt, M. D., & Rudnicka, J. (2002). A review of naturalistic decision making research with some implications for knowledge management. *Journal of Knowledge Management, Volume 6*(Issue 1), 63-73. doi:10.1108/13673270210417709

Mitchell, C. J., Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*(02), 183. doi:10.1017/s0140525x09000855

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making, 7*(2), 211-226. doi:10.1177/1555343412459192

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., & Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology, 6*(10), 575-585. doi:10.1016/s1076-6332(99)80252-9

Okoli, J., Watt, J., & Weller, G. (2017). Towards the classification of fireground cues: A qualitative analysis of expert reports. *Journal of Contingencies and Crisis Management, 25*(4), 197-208. doi:10.1111/1468-5973.12129

Orquin, J. L. (2014). A Brunswik lens model of consumer health judgments of packaged foods. *Journal of Consumer Behaviour, 13*(4), 270-281. doi:10.1002/cb.1465

Pauley, K., O'Hare, D., & Wiggins, M. W. (2009). Measuring expertise in weather-related aeronautical risk perception: The validity of the cochran–weiss–shanteau (cws) index. *The International Journal of Aviation Psychology, 19*(3), 201-216. doi:10.1080/10508410902979993

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies, 45*(5), 579-597. doi:10.1006/ijhc.1996.0068

Reina, R., Moreno, F. J., & Sanz, D. (2007). Visual behavior and motor responses of novice and experienced wheelchair tennis players relative to the service return. *Adapted Physical Activity Quarterly, 24*(3), 254-271. doi: 10.1123/apaq.24.3.254

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica, 127*(2), 258-276. doi:10.1016/j.actpsy.2007.05.004

Rolison, J. J., Evans, J., Dennis, I., & Walsh, C. R. (2012). Dual-processes in learning and judgment: Evidence from the multiple cue probability learning paradigm. *Organizational Behavior and Human Decision Processes, 118*(2), 189-202. doi:10.1016/j.obhdp.2012.03.003

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors: The Journal of Human Factors and Ergonomics Society, 50*(6), 864-878. doi:10.1518/001872008x374974

Schwarz, N., Bless, H., Strack, F., & Klumpp, G. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195-202. doi:10.1037//0022-3514.61.2.195

Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review, 69*(4), 493-513.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22. doi:10.1037/0033-2909.119.1.3

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General, 139*(2), 266-298. doi:10.1037/a0018620

Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. In *the Proceedings of the 6th international conference on Multimodal interfaces,* 9-15. ACM.

Summers, D. A., & Hammond, K. R. (1966). Inference behavior in multiple-cue tasks involving both linear and nonlinear relations. *Journal of Experimental Psychology, 71*(5), 751-757. doi:10.1037/h0023122

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance, 23*(1), 86-112. doi:10.1016/0030-5073(79)90048-5

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2014). Eye tracking for skills assessment and training: a systematic review. *Journal of Surgical Research, 191*(1), 169-178. doi:10.1016/j.jss.2014.04.032

Todd, M. A., & Thomas, M. J. W. (2012). Flight hours and flight crew performance in commercial aviation. *Aviation, Space, and Environmental Medicine, 83*(8), 776. doi:10.3357/ASEM.3271.2012

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232. doi:10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. doi:10.1126/science.185.4157.1124

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(2), 87-97. doi:10.1016/S1369-8478(02)00008-6

Underwood, G., Crundall, D., & Chapman, P. (2002). Selective searching while driving: the role of experience in hazard detection and general surveillance. *Ergonomics, 45*(1), 1-12. doi:10.1080/00140130110110610

Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology*, 1-7. doi:10.1080/14992027.2018.1476782

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(1), 104-116. doi:10.1518/hfes.45.1.104.27233

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science, 65*, 118-124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., Griffin, B., & Brouwers, S. (2019). The potential role of context-related exposure in explaining differences in water safety cue utilization. *Human Factors*. doi:10.1177/0018720818814299

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied, 1*(4), 305. doi:10.1037/1076-898X.1.4.305

York, K. M., Doherty, M. E., & Kamouri, J. (1987). The influence of cue unreliability on judgment in a multiple cue probability learning task. *Organizational Behavior and Human Decision Processes, 39*(3), 303-317. doi:10.1016/0749-5978(87)90026-4

CHAPTER THREE

Cue Utilisation as a Selection Tool

Given that higher cue utilisation is generally associated with improved performance in the operational environment (Loveday, Wiggins, Harris, O'Hare, & Smith, 2012; Loveday, Wiggins, Searle, Festa, & Schell, 2012), it is a construct that is likely to represent a useful indicator of present and future performance in a range of occupational domains. An assessment of cue utilisation could be employed as part of an evaluation of performance to identify operators who may require further training, and recognise and advance those operators who are functioning at a relatively higher level.

More importantly, a measure of cue utilisation might be utilised during the selection process to identify candidates who are likely to demonstrate superior performance in the future. While current selection practices do not necessarily assess for cue utilisation explicitly, it might be argued that a commonly used selection tool, the Situation Judgement Test (SJT), incorporates some elements of cue utilisation.

SJTs are scenario-based assessment tools, where candidates are offered the opportunity to participate in, and respond to, a job-related scenario (McDaniel, Hartman, Whetzel, & Grubb III, 2007). For instance, in assessing socio-cultural competencies amongst military personnel, Reinerman-Jones, Matthews, Burke, and Scribner (2016) presented participants with a trust scenario, which detailed an event involving poor performance by a military personnel from a partner nation. Participants selected their response from a list of response options, one of which was assigned as the favoured reponse.

SJTs are administered as part of the selection process given their relatively strong predictive validity for later job performance (Christian, Edwards, & Bradley, 2010; Lievens & Sackett, 2012). Higher performance on an SJT is generally associated with higher job performance in various domains, such as the military (Hauenstein, Findlay, & McDonald, 2010; Reinerman-Jones et al., 2016) and corporate management (Motowidlo, Dunnette, & Carter, 1990). Further, the SJT is a versatile assessment tool that can be adapted to any domain or job-level by adjusting the complexity of scenarios to fit the target competencies (Lievens, Buyse, & Sackett, 2005).

While the SJT appears to constitute a valuable assessment tool, there have been concerns as to the underlying constructs that form the basis of the evaluation (McDaniel & Nguyen, 2001). SJTs were developed based on a practice-based approach to selection with the intention of predicting future performance (Whetzel & McDaniel, 2009). With a focus on functional outcomes, the underlying constructs measured by SJTs are largely ignored or derived retrospectively (Christian et al., 2010; Whetzel & McDaniel, 2009).

Sternberg and Wagner (1993) have suggested that performance on SJTs reflect a single construct such as practical intelligence or tacit knowledge, which refers to 'practical know-how'. They maintain that tacit knowledge is distinct from general intelligence measured using conventional intelligence tests, and constitutes the application of acquired knowledge, skills, and abilities. However, the application of tacit knowledge in this context is unsubstantiated (Gottfredson, 2003). Attempts to derive underlying constructs from SJT scores have yielded largely inconsistent, ambiguous and uninterpretable factors (McDaniel & Whetzel, 2005). An exception to the generally inconsistent findings is the association between performance scores on SJTs and measures of cognitive ability.

Higher performance scores on SJTs are generally associated with greater scores on measures of cognitive ability (Lievens et al., 2005; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). However, measures of cognitive ability do not account for all the variance in performance on SJTs (McDaniel et al., 2001). This suggests that there may be individual differences captured by SJTs that are relevant to job performance but which are not necessarily reflected in individual measures of cognitive ability.

Successful performance on SJTs might be attributable to a number of capabilities, including the timely and accurate assessment of the situation, and the generation and evaluation of responses drawn from memory. Due to their involvement in selection, scenarios that are incorporated within SJTs normally constitute low-fidelity accounts of situations that might occur within operational settings (Motowidlo et al., 1990). Consequently, any judgments in response to these hypothetical situations are likely to be a function of both the accurate assessment of a situation, and the opportunity and capability to select the appropriate response from the options available (Weekley & Ployhart, 2005).

Like performance in the operational context, superior performance on an SJT is likely to be dependent, at least in part, upon the accuracy of situation assessment. Since cue utilisation supports situation assessment (Wiegmann, Goh, & O'Hare, 2002), it might be inferred that performance on an SJT reflects individual differences in context-related cue utilisation. Higher cue utilisation should be associated with improved situation assessment through the application of relevant schemas, and in turn, more accurate responses on SJTs in comparison to participants with lower cue utilisation.

In responding to SJTs or performing within a specific operational domain, cue utilisation not only relates to features in the environment and schemas in memory, but also to the contextual or situational information that characterises the environment (Williams & Jackson, 2019). For example, expert and non-expert squash players participated in a simulated on-court game where, at varying points during the game, the scenarios were occluded before, at or after the ball had made contact with the opponents' rackets, and the participants were required to respond by performing a return stroke (Abernethy, Gill, Parks, & Packer, 2001).

Both expert and non-expert players demonstrated below 20% error rate when the occlusion occurred 20 milliseconds prior to contact, indicating that cue utilisation was evident in both groups. In particular, postural cues relevant to the contact were utilised to anticipate opponents' movements. However, non-expert players demonstrated an error rate above chance, while expert players demonstrated an error rate below chance when the occlusion occurred 580 milliseconds prior to contact. The longer interval meant that postural cues were not available at the time of assessment, suggesting that expert players were able to make appropriate anticipatory assessments based on contextual cues, such as action sequences (Murphy, Jackson & Williams, 2018) and the opponents' movement tendencies (Mann, Schaefers & Cañal-Bruland, 2014). Therefore, cue utilisation, when responding to SJTs or performing within a specific operational domain, is likely to rely on both environmental and contextual cues.

Given that cues can be derived from both environmental features and contextual attributes, it might be inferred that cues are, to an extent, idiosyncratic to individuals, and there may not be an ideal set of possible cues that underlie superior performance. Therefore, the accuracy-based response selection method adopted in SJTs may limit its capacity to determine levels of cue utilisation, but rather, may simply reflect the extent to which operators' responses align with the option favoured by a limited pool of subject-matter experts. The alternative is to evaluate cue utilisation indirectly using an assessment tool that measures behaviours that reflect the utilisation of cues (Watkinson, Bristow, Auton, McMahon, & Wiggins, 2018; Wiggins, 2012; Yee, Wiggins, & Searle, 2017).

Behaviour-based measures obviate the requirement to identify a priori, a proposed set of universal cues to which operators might attend. Behaviour that is associated with greater utilisation of cues is generally associated with higher performance in the operational environment (Loveday, Wiggins, Harris, et al., 2012; Loveday, Wiggins, Searle, et al., 2012). Since higher cue utilisation is associated with greater operational performance, establishing the propensity to acquire cues within an operational context is likely to constitute a useful opportunity for employee selection.

Employee Selection

Across many occupational domains, significant resources are invested in employee selection to ensure the identification of suitable candidates. Inadequate selection often results in increased costs to the organisation in engaging and training employees and, in some circumstances, results in the termination of employment and the requirement to recruit new employees (Golec & Kahya, 2007). Therefore, selection tools that contribute to distinguishing more from less suitable candidates are likely to be financially beneficial to an organisation.

In high-risk domains, such as aviation, the military, and firefighting, appropriate employee selection is critical for optimal performance and the resulting safety of the public and employees. Inadequate employee selection can result in an increased rate and/or severity of accidents and incidents. TransAsia Airlines, for instance, was dissolved in 2016 following two fatal plane crashes, one of which involved Flight GN235. The selection of pilots with a substandard performance record was one of the main issues that was cited as a significant factor associated with the crash (Aviation Safety Council, 2016).

Employee selection is often designed to identify the compatibility between the demands of the job and the capabilities of the person/employee (Sekiguchi, 2004). This compatibility between person and job demands, known as Person-Job fit (P-J), varies depending upon the match between the characteristics of the candidates and the specified job criteria (Edwards, 1991). Organisations often rely on P-J fit as a guiding framework to select potential candidates, as an increased P-J fit is associated with higher task performance, lower levels of stress (Chilton, Hagrave, & Rong, 2014), and improved job satisfaction (Warr & Inceoglu, 2012).

To ensure an increased 'fit' between employees and job roles, it is often necessary to undertake a job analysis to establish the essential tasks and the skills required to undertake relevant tasks successfully (F. Patterson, Ferguson, & Thomas, 2008; Schneider & Konz, 1989). By establishing the content of the job, organisations can evaluate different candidates systematically by matching standard measures pertaining to the relevant tasks against the candidates' performance on these measures. Greater correspondence between the standard measure and the candidates' performance should reflect an increased P-J fit.

The measurement tools used to assess task performance must demonstrate reasonable validity in differentiating more from less suitable candidates. Validity refers to the extent to which the measures derived from the measurement tools reflects the target abilities and skills (Elia & Stratton, 2011). Predictive validity, in particular, is necessary for measurement tools that are administered during selection to ensure that performance during selection will translate into effective performance on the job (Schmidt & Hunter, 1998).

For job roles involving tasks that require greater physical effort, measures of physical strength appear to show strong predictive ability with later job performance (Groeller, Fullagar, Sampson, Mott, & Taylor, 2015; Henderson, 2010; Henderson, Berry, & Matic, 2007). For example, amongst firefighters, candidates who achieve higher scores on measures of strength, including bench press and grip strength tests, tend to perform at a significantly higher level during job-specific performance, including roof ladder placement and an axe chopping exercise (Henderson et al., 2007).

For job roles involving tasks that are dependent upon cognitive skills such as situation assessment, there is a need to ensure that measurement tools incorporate measures of cue utilisation. However, the successful utilisation of cues requires that feature-event relationships lie resident in memory. While this might be appropriate for skilled operators, less experienced operators may have yet to be exposed to situations that enable the development of cues. In this case, it may be more appropriate to consider the propensity to acquire cue-based associations, rather than their utilisation.

Cue Acquisition as a Predictive Measure

Administering a measure of cue utilisation as an assessment tool during selection is limited to job roles where previous experiences have been required. For entry-level jobs, where candidates have little to no previous experience, a measure of cue utilisation is largely ineffective given that domain-specific experience is required to meaningfully respond to behavioural measures that reflect cue utilisation. This is predominantly because cue utilisation functions on domain-specific feature-event/object associations, which are acquired through domain-specific experience (Wiggins, 2012).

The domain-specificity of feature-event/object associations that constitute schemas also means that the influence of cue utilisation is limited to situation assessment within the given domain. For instance, feature-event/object associations that support cue utilisation in operating an aircraft are unlikely to support the utilisation of cues during medical diagnosis.

Given that domain-specific experience is required for cue utilisation, a measure of individual differences in the *acquisition* of domain-specific cues is more appropriate for assessing candidates for entry-level jobs. This measure should target a construct that embodies two characteristics. First, the construct should operate only in the absence of domain-specific experience or pre-existing knowledge. Second, the construct should relate to the ability to acquire relevant cues that support cue utilisation. These prerequisites were derived based on the assumption that a construct that bolsters cue utilisation is a strong predictor of future cue utilisation, and subsequently, job performance. In effect, this construct is equivalent to cue acquisition.

A significant component that enables cue utilisation is feature-event/object associations (Kaempf, Klein, Thordsen, & Wolf, 1996; Klein, 2008). At the initial stages of exposure to domain-specific environments, individuals interact with novel features and events or objects in the environment to form feature-event/object associations that are stored in memory (Wiggins, 2012). Arguably, this process of acquiring feature-event/object associations is independent of any domain-specific experience and constitutes the basis of cue utilisation.

The acquisition of relevant feature-event/object associations or cues is thought to occur through implicit learning. Implicit learning is a relatively primitive cognitive process that underlies the development of basic abilities, such as language, motor skills and causal reasoning (Reber, 1989; Reber, Walkenfeld, & Hernstadt, 1991). The acquisition of these fundamental abilities occurs in early infancy to form the foundation for the development of more complex cognitive functions or processes that rely on domain-specific knowledge (Kirkham, Slemmer, & Johnson, 2002).

Fundamentally, implicit learning operates on individuals' sensitivity to the regularities or patterns within complex environments (R. Patterson, Pierce, Bell, & Klein, 2010). In this regard, implicit learning is a constituent part of abstract reasoning, which constitutes the ability to identify associations between relatively discrete non-verbal mental representations, and the application of these associations to novel conditions (Kalbfleisch, van Meter, & Zeffiro, 2007; Wright, Matlen, Baym, Ferrer, & Bunge, 2008).

In a study of synthetic language, for instance, Reber (1969) presented participants with a study sheet comprising sentences constructed from artificial grammar. Following the study phase, participants were presented with a series of sentences that were not part of the study sheet, and were asked to indicate whether each sentence was grammatically consistent with the test array. Accuracy was provided to participants following each trial.

Reber (1969) observed that participants' accuracy improved over time. However, when questioned about the grammatical structure, participants were not able to verbally describe the structure. The results suggest that individuals possess an inherent capacity to derive relevant associations from complex and noisy stimuli without explicit knowledge of the process or the intention to derive any associations.

This capacity to derive associations from complex stimuli is not limited to the acquisition of synthetic language. For instance, Brouwers, Wiggins, Griffin, Helton, and O'Hare (2017) presented participants with a rail control task on which none of the participants had experience. Participants were asked to monitor the movement of the trains and redirect any trains that were running on incompatible tracks.

A pattern of train movement was incorporated into the task such that the derivation of the pattern would allow for more accurate and timely responses when redirecting trains. The outcomes indicated that, over the course of the 20-minutes testing phase and 170 trains monitored, participants committed an average of five errors, constituting an error rate of approximately 3.0%. Unlike Reber (1969), however, Brouwers et al. (2017) noted that some participants were able to explicitly identify and verbally describe the pattern.

The apparent ability to verbalise the underlying associations in Brouwers et al. (2017) does not necessarily indicate that the participants had acquired the associations through the process of explicit learning, which is presumed to occur with conscious awareness of the learning process and the target items to be learned (Ellis, 2009). Instead, participants in Brouwers et al. (2017) were simply instructed to perform the task, and were not informed of any patterns of train movement.

On the basis that participants were not made aware of any existing associations to be learned, it would, therefore, be erroneous to conclude that the verbalisation of the pattern of train movement observed in Brouwers et al. (2017) was the result of explicit learning. Rather, it might be argued that the verbalisation of the pattern was the observable outcome of implicit learning. The learning process itself is likely to have occurred without conscious awareness based on the assumption that there should be no intention to derive associations given that the instructions were performance-centred, rather than learning-centred.

The differences in the complexity of the embedded associations between Brouwers et al. (2017) and Reber (1969) may account for the apparent differences in the ability to verbalise the underlying associations. Arguably, the associations presented in the patterns of train movement in Brouwers et al. (2017) were more tangible for participants than the more abstract grammatical structures presented in Reber (1969). This theoretical proposition warrants further investigation, but is outside the scope of the present dissertation.

In Brouwers et al. (2017), performance accuracy and the verbalisation of the pattern of train movements can be taken as two outcome measures of implicit learning, which underpins the process of cue acquisition. While the behavioural performance measured by total accuracy was consistent across all participants, the verbalisation of associations was only observed amongst some of the participants. This incongruence between the two outcome measures suggests that only some participants achieved consistent accuracy through the acquisition and application of relevant feature-event/object associations. Consequently, it might be inferred that the inherent capacity for implicit learning varies across individuals such that there are individual differences in the rate at which individuals acquire relevant feature-event associations from novel environment.

Cue acquisition appears to be a construct that supports cue utilisation by facilitating the acquisition of relevant feature-event/object associations through exposure. In fact, there may be individual differences in cue acquisition, which contribute to the differences in domain-specific cue utilisation, and subsequently, task performance in any domain. The development and assessment of an instrument to measure individual differences in cue acquisition is the focus of the present dissertation.

Present Programme of Research

Measures of cue utilisation cannot be administered to derive meaningful predictions when selecting candidates for entry-level jobs. These candidates typically do not have previous experience to acquire relevant schemas on which cue utilisation functions to support effective situation assessment. Therefore, a measure of cue acquisition was proposed as an assessment tool to predict future domain-specific cue utilisation, and subsequently, job performance.

The aim of the present programme of research was to develop and evaluate a measure of cue acquisition to assess individual differences in the acquisition of relevant feature-event/object associations. In Study 1, a measure of cue acquisition was developed and evaluated against performance on a rail control task to establish construct validity. In Study 2, the measure of cue acquisition was evaluated against the measure of cue utilisation within the domain of general aviation to establish the association between cue acquisition and cue utilisation. In Study 3, the measure of cue acquisition was again evaluated against a measure of cue utilisation but in the context of driving a motor vehicle. The intention was to investigate the generalisability of the association between cue acquisition and cue utilisation across domains. In Studies 4 and 5, the measure of cue acquisition was evaluated against flight performance within an actual flight training and a flight simulator respectively to investigate the association between cue acquisition and task performance.

References

Abernethy, B., Gill, D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. *Perception, 30*(2), 233-252. doi:10.1068/p2872

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics, 60*(11), 1500-1515. doi:10.1080/00140139.2017.1330494

Chilton, M. A., Hagrave, B. C., & Rong, D. J. (2014). Person-job cognitive style fit for software developers: The effect on strain and performance. *Journal of Management Information Systems, 22*(2), 193-226. doi:10.1080/07421222.2005.11045849

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*(1), 83-117. doi:10.1111/j.1744-6570.2009.01163.x

Edwards, J. R. (1991). Person–job fit: A conceptual integration, literature review, and methodological critique. In Cooper C. (Ed.), *International Review of Industrial Andorganizational Psychology* (pp. 283–357). Chichester, UK: Wiley.

Elia, M., & Stratton, R. J. (2011). Considerations for screening tool selection and role of predictive and concurrent validity. *Current Opinion in Clinical Nutrition and Metabolic Care, 14*(5), 425-433. doi: 10.1097/MCO.0b013e328348ef51

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In D. Singleton (Ed.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching. Volume 42 of Second Language Acquisition* (pp. 3-26). Bristol, UK: Multilingual Matters.

Golec, A., & Kahya, E. (2007). A fuzzy model for competency-based employee evaluation and selection. *Computers & Industrial Engineering, 52*(1), 143-161. doi:10.1016/j.cie.2006.11.004

Gottfredson, L. S. (2003). Dissecting practical intelligence theory Its claims and evidence. *Intelligence, 31*(4), 343-397. doi:10.1016/s0160-2896(02)00085-5.

Groeller, H., Fullagar, H., Sampson, J. A., Mott, B. J., & Taylor, N. (2015). Employment standards for australian urban firefighters. *Journal of Occupational and Environmental Medicine, 57*(10), 1083-1091. doi:10.1097/JOM.0000000000000527

Hauenstein, N. M. A., Findlay, R. A., & McDonald, D. P. (2010). Using situational judgment tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology, 22*(3), 262-281. doi:10.1080/08995605.2010.492679

Henderson, N. D. (2010). Predicting long-term firefighter performance from cognitive and physical ability measures. *Personnel Psychology, 63*(4), 999-1039. doi:10.1111/j.1744-6570.2010.01196.x

Henderson, N. D., Berry, M. W., & Matic, T. (2007). Field measures of strength and fitness predict firefighter performance on physically demanding tasks. *Personnel Psychology, 60*(2), 431-473. doi:10.1111/j.1744-6570.2007.00079.x

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Kalbfleisch, M. L., van Meter, J. W., & Zeffiro, T. A. (2007). The influences of task difficulty and response correctness on neural systems supporting fluid reasoning. *Cognitive Neurodynamics, 1*, 71-84. doi: 10.1007/s11571-006-9007-4

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*(2). doi:10.1016/S0010-0277(02)00004-5

Klein, G. A. (2008). Naturalistic decision making. *Human factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*(3), 442-452. doi:10.1037/0021-9010.90.3.442

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*(2), 460. doi:10.1037/a0025741

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2012). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 90-107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2012). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 125-137. doi:10.1177/0018720812448475

Mann, D. L., Schaefers, T., & Cañal-Bruland, R. (2014). Action preferences and the anticipation of action outcomes. *Acta Psychologica, 152,* 1-9. doi: 10.1016/j.actpsy.2014.07.004

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*(1), 63-91. doi:10.1111/j.1744-6570.2007.00065.x

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730-740. doi:10.1037/0021-9010.86.4.730

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*(1&2), 103-113. doi:10.1111/1468-2389.00167

McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence, 33*(5), 515-525. doi:10.1016/j.intell.2005.02.001

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640-647. doi: 10.1037/0021-9010.75.6.640

Murphy, C. P., Jackson, R. C., & Williams, A. M. (2018). The role of contextual information during skilled anticipation. *The Quarterly Journal of Experimental Psychology*, 71, 2070–2087. doi: 10.1177/1747021817739201

Patterson, F., Ferguson, E., & Thomas, S. (2008). Using job analysis to identify core and specific competencies: implications for selection and recruitment. *Medical Education, 42*(12), 1195-1204. doi:10.1111/j.1365-2923.2008.03174.x

Patterson, R., Pierce, B. J., Bell, H. H., & Klein, G. (2010). Implicit learning, tacit knowledge, expertise development, and naturalistic decision making. *Journal of*

Cognitive Engineering and Decision Making, 4(4), 289-303. doi:10.1177/155534341000400403

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*(1), 115-119. doi:10.1037/h0027454

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-245. doi: 10.1037/0096-3445.118.3.219.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(5), 888-896.

Reinerman-Jones, L., Matthews, G., Burke, S., & Scribner, D. (2016). A situation judgment test for military multicultural decision-making: initial psychometric studies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*(1), 1482-1486. doi:10.1177/1541931213601340

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262. doi:10.1037/0033-2909.124.2.262

Schneider, B., & Konz, A. (1989). Strategic job analysis. *Human Resource Management, 28*(1), 51-63. doi:10.1002/hrm.3930280104

Sekiguchi, T. (2004). Person-organization fit and person-job fit in employee selection: A review of the literature. *Osaka keidai ronshu, 54*(6), 179-196.

Sternberg, R. J., & Wagner, R. K. (1993). The g-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science, 2*(1), 1-4. doi:10.1111/1467-8721.ep10770441

Warr, P., & Inceoglu, I. (2012). Job engagement, job satisfaction, and contrasting associations with person–job fit. *Journal of Occupational Health Psychology, 17*(2), 129. doi:10.1037/a0026859

Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology*, 1-7. doi:10.1080/14992027.2018.1476782

Weekley, J. A., & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance, 18*(1), 81-104. doi:10.1207/s15327043hup1801_4

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*(3), 188-202. doi:10.1016/j.hrmr.2009.03.007

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 189-197. doi:10.1518/0018720024497871

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Williams, A. M., & Jackson, R. C. (2019). Anticipation in sport: Fifty years on, what have we learned and what research still needs to be undertaken? *Psychology of Sport and Exercise, 42*(Research Quarterly for Exercise Sport 59 3 1988), 16-24. doi:10.1016/j.psychsport.2018.11.014

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience, 1*(8), 1-8. doi: 10.3389/neuro.09.008.2007

Yee, D. J., Wiggins, M. W., & Searle, B. J. (2017). The Role of Social Cue Utilization and Closing the Loop Communication in the Performance of Ad Hoc Dyads. *Human Factors, 59*(6)*,* 1009-1021. doi:10.1177/0018720817699512

CHAPTER FOUR

Timed-Search Task: An Instrument to Assess Individual Differences in Cue Acquisition[1]

[1] The author of the present thesis contributed approximately 80% to the preparation of the present paper in the design of the study, data collection, data analysis, and writing of the paper.

Abstract

While an association between cue utilisation and situation assessment has been reasonably well-established, the impact of individual differences on the acquisition of cues necessary for situation assessment remains unclear. To differentiate the acquisition of cue-based associations in memory, five prospective tasks were designed, that varied according to the process by which feature-event associations could be established. Performance on each of the five tasks was assessed against performance on a novel rail task that incorporated an implicit pattern of train movements. Eight-two undergraduate psychology students participated in the study. Of the five tasks, the Timed-Search Task, which incorporated restricted period of exposure and less salient characteristics, was the only measure that was associated with performance on the novel rail task. It was concluded that the Timed-Search Task could be used as a potential selection test to differentiate the rate at which cue-based associations are acquired during a novel task.

Keywords: cue acquisition, situation assessment, cues, pattern recognition

Timed-Search Task: An Instrument to Assess Individual Differences in Cue Acquisition

Situation assessment is a cognitive process involving the extraction, integration and evaluation of information in the environment to construct an understanding of the presenting situation (Fracker, 1988; Horrey & Wickens, 2001). During problem resolution and decision-making, the understanding derived from situation assessment provides the basis on which action plans are formed or selected (Klein, Calderwood, & Clinton-Cirocco, 2010). Therefore, an accurate situation assessment tends to result in the selection of more accurate action plans and response behaviours, which culminates in superior outcomes.

The accuracy of situation assessment varies between individuals (Horrey & Wickens, 2001). This individual variation arises from differences in the manner in which individuals interact with the environment. Specifically, prior experiences in similar situations appear to contribute to the acquisition of different sets of cues, which, when utilised, influence the accuracy of situation assessment (McCormack, Wiggins, Loveday, & Festa, 2014; Wiggins & O'Hare, 2003).

Cues constitute features in the environment that, when perceived, trigger the retrieval of related events or objects from memory (Kaempf, Klein, Thordsen, & Wolf, 1996). These feature-event/object associations facilitate the construction of assessments by providing operators with expectancies based on previous experiences (Klein, 2008). For example, repeated experience with cardiac patients might enable a health professional to draw an association between grey skin colour (feature) and an inadequate blood supply (event), prompting a working assessment of a possible cardiac concern. In the absence of experience necessary to draw the association, the response from the health professional is likely to be slow and/or inaccurate.

The acquisition and subsequent utilisation of cues during decision-making appears to be related to higher levels of operational performance (Abernethy, 2008; Loveday, Wiggins, & Searle, 2014; Mann, Williams, Ward, & Janelle, 2007; Wiggins, Azar, Hawken, Loveday, & Newman, 2014). For instance, chess players with a greater repertoire of chess configurations in memory show an improved capacity to reconstruct complex chess positions having been shown the pattern for brief periods (Chase & Simon, 1973). However, there is no superiority in performance when chess positions are configured randomly. These observations suggest that superior performance outcomes are facilitated by the utilisation of cues, since random configurations of chess pieces do not embody meaningful patterns of features.

Arguably, cue utilisation improves performance outcomes by affording decision-makers the capacity for rapid and accurate assessments of a situation (Simon & Chase, 1988). Rapid situation assessment reduces response latency in response to changes in a system state, while maintaining the accuracy of responses. This enables the conservation of cognitive resources necessary to respond to additional demands.

The conservation of cognitive resources is enabled through the utilisation of cues since the information retrieved from long-term memory exists as unified feature-event/object associations, thereby reducing the demands on working memory (Chung & Byrne, 2008). As a result, residual resources become available to assist with the management of additional tasks, should this become necessary.

Feature-event/object associations are developed through repeated exposure to the co-occurrence of features and events or objects in the environment (Wiggins, 2012). Therefore, feature-event/object associations are generally thought to be domain-specific. However, there is evidence to suggest that differences in cue utilisation within one domain are associated with performance outcomes in another, related domain (Brouwers, Wiggins, Griffin, Helton, & O’Hare, 2017; Brouwers, Wiggins, Helton, O’Hare, & Griffin, 2016; Renshaw & Wiggins, 2017). For example, Brouwers et al. (2017) reported a relationship between cue utilisation in the context of operating a motor vehicle and performance outcomes in a simulated rail control task. Given that the feature-event/object associations within the context of operating a motor vehicle are not directly relevant to rail control, the findings suggest that there may be an underlying trait or capability that governs the process of cue utilisation, independent of the domain-specific associations that actually direct behaviour.

In the case of Brouwers et al. (2017), participants had experience in operating motor vehicles but no experience in rail control. The absence of rail control associations in memory meant that any variation in performance on the rail control task could not be the product of prior cue utilisation. However, to provide a basis on which cues might be established, a pattern of train movements was incorporated into the rail control task, whereby only two of the four trains required re-routing at any one time. Therefore, any variations in performance on the rail task could only be attributed to differences in the ability to identify and learn the pattern of train movements, a process that might equate to *cue acquisition*.

According to Brouwers et al. (2017), participants who recorded greater levels of cue utilisation in operating a motor vehicle and higher performance on the rail control task were also up to 11 times more likely to report having identified the pattern of train movements during the rail control task. Since participants were not advised of the pattern of train movements, it suggests that there are individual differences in the capacity to identify, acquire, and retain feature-event/object associations in the form of cues.

The successful association between features and events or objects in memory initially requires the capacity to isolate features and events or objects from environmental noise. This process is likely to involve statistical learning, whereby features and events or objects that co-occur repeatedly, are presumed to co-exist, thereby forming the basis of cues (Kim, Seitz, Feenstra, & Shams, 2009). Importantly, this capability to recognise the co-occurrence of features and events or objects will determine the rate at which cues are acquired which, in turn, has an impact on performance (Bilalić, Langner, Erb, & Grodd, 2010).

The present study was designed to evaluate five goal-directed tasks that were developed as possible measures of individual differences in cue acquisition. All five tasks required participants to derive associations between features and events or objects through repeated exposure. However, the conditions under which these associations were expected to be derived, varied across the five tasks to distinguish the capabilities of participants.

For example, in some tasks, the period of exposure was limited to identify those participants who were better able to establish relationships quickly. In other tasks, there was a requirement to discern the characteristics of stimuli that might form the basis of associations. The characteristics relevant to the associations were made less salient to identify those participants who were more capable of identifying feature-event/object associations in the context of noisy stimuli.

The *period of exposure* constituted the time during which features and events were presented concurrently during trials. In tasks where exposure was time-restricted, participants with a greater capacity for cue acquisition were expected to perform unimpeded when establishing associations between features and events or objects, and thereby, maintained relatively higher levels of performance. Differences in performance between participants were expected to be less evident where the exposure to features and events or objects concurrently was unrestricted.

The *salience of characteristics* constituted the extent to which specific features and events or objects that might comprise the basis of relationships, were extracted from the visual scene for the participants. Where features and events or objects were more salient, attentional resources could be directed exclusively to establishing their co-occurrence (Reber, 1989). By contrast, where features and events or objects were less salient, extracting features and events or objects from the visual scene, prior to establishing their co-occurrence, must occur through selective attention in response to reduced attentional resources (Jiang & Chun, 2001). Therefore, differences in performance in this case, were likely to reflect differences in the efficiency with which attentional resources were allocated during the task.

The five cue acquisition tasks varied as to whether or not they incorporated a restricted period of exposure, and whether the features and events or objects that might form associations were more or less salient for the participants. The five tasks were referred to as: the (1) Search Task, (2) Timed-Search Task, (3) Explicit Association Learning Task, (4) Problem Solving Task, and (5) Categorisation Task (see Table 1). The combinations of period of exposure and salience of characteristics differed systematically across all tasks apart from the Problem-Solving Task and the Categorisation Task, both of which incorporated an unrestricted period of exposure and characteristics of associations that were salient. However, these tasks differed in the approach with which stimuli were presented with the Problem Solving Task incorporating a greater quantity of information within a less structured format than the Categorisation Task.

Table 1.
Variations of the Two Measurement Conditions across the Five Cue Acquisition Tasks

	Exposure Period	Salience of Characteristics
Search Task	Unrestricted	Less Salient
Timed-Search Task	Restricted	Less Salient
Explicit Association-Learning Task	Restricted	More Salient
Problem Solving Task	Unrestricted	More Salient
Categorisation Task	Unrestricted	More Salient

The construct validity of each of the five tasks was tested against performance on a simulated rail control task that incorporated a pattern of train movements, which when identified, would facilitate performance. Since the pattern of train movements was not made salient to participants and needed to be discerned through repeated exposure to transient stimuli (train movements), a positive relationship between performance on the cue acquisition tasks and performance on the rail control task would provide a degree of support for the construct validity of the cue acquisition task(s).

Methodology

Participants

Following ethical approval by the Human Research Ethics Committee (see Appendix A), 82 first and second year Psychology students were recruited from an Australian University via an online portal. Sixty-nine participants were female and 13 were male, ranging in age from 17 to 53 (M = 20.98, SE = 0.72).

None of the participants had previous experience in rail control operations to ensure that they were naïve to experimental task. Participants were excluded if they self-reported colour blindness and were awarded course credit for participation in the study.

Materials

The five mutually exclusive cue acquisition tasks were designed and administered via an online software platform. They included the: (1) Search Task, (2) Timed-Search Task, (3) Explicit Association Learning Task, (4) Problem Solving Task, and (5) Categorisation Task. All five tasks were designed within a fictional context, where the influence of previous experience was expected to be minimal. The features within the fictional context consisted of different characteristics of *fish* with no corresponding real-life counterparts.

Task 1: Search Task (Unrestricted, Less Salient)

The Search Task involved a simple search test, wherein participants were tasked to search for, and select a target fish from amongst a school of similar fish. The target fish was associated with a specified feature such that the acquisition of the feature-event/object association would result in a shorter response latency in identifying the target fish in subsequent trials.

Given that the participants had no existing associations with regards to the location of the target fish within the fictional context, it was assumed that the task would require participants to map the given target to the array of stimuli until a match was found. Across the series of trials, it was expected that participants would display individual differences in the rate at which they acquired and then utilised the feature-event/object associations. The successful identification and utilisation of the embedded association was expected to result in shorter response latencies across trials.

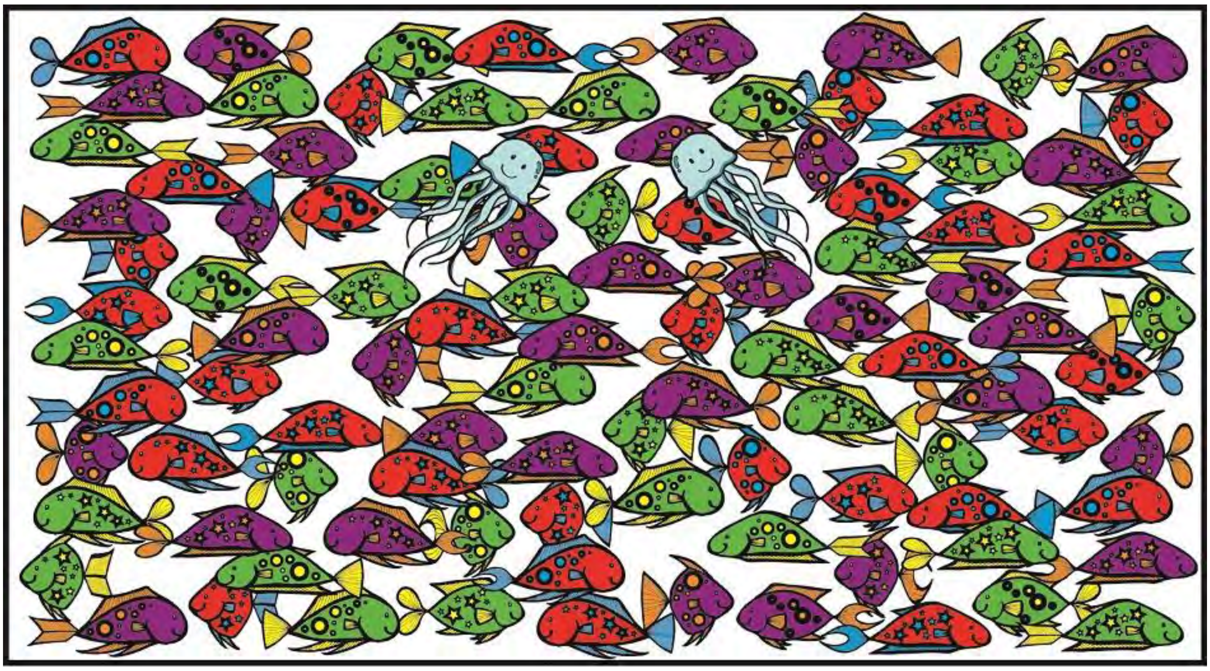

Figure 1. An example of one of the trials within the Search Task.

The task consisted of 12 trials. In each trial, participants were presented with an image displaying a variety of fictitious fish (see Figure 1). Participants were required to identify and select the target fish (a frowning face) as quickly as possible. There were two relevant associations that would assist performance. First, the target fish (event) was always located directly above the heads of the two jellyfish (feature). Second, the colour (feature) of the target fish's body was always green (see Figure 2). No feedback was provided in each trial, and no time limit was imposed for each trial.

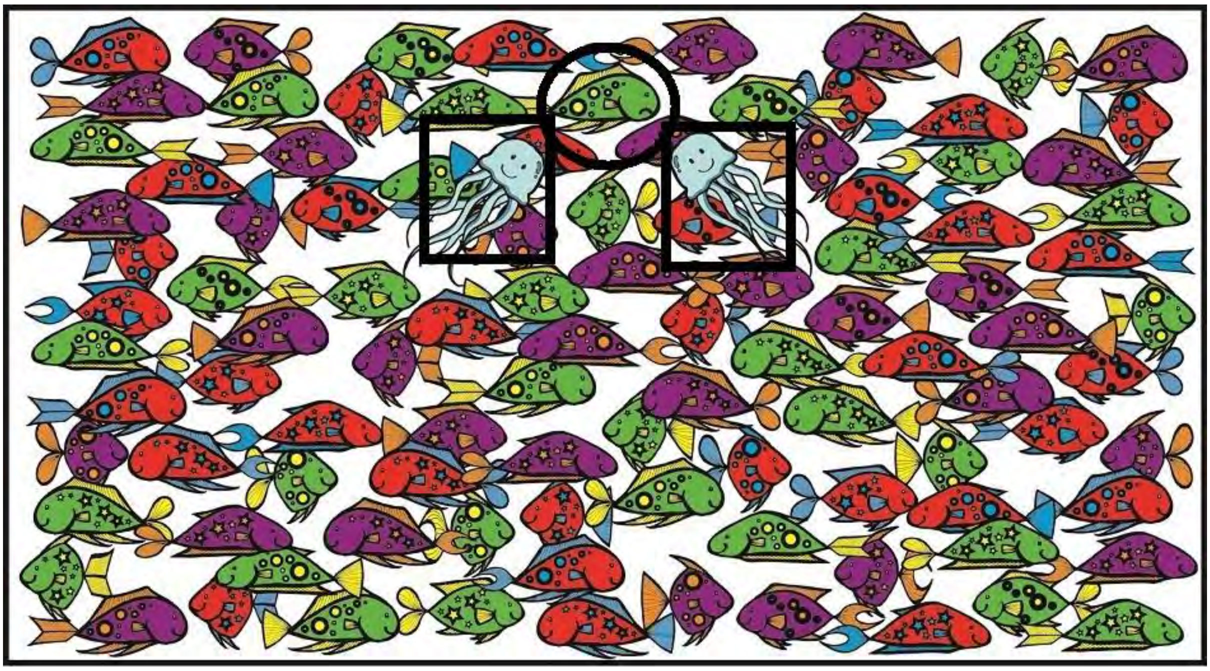

Figure 2. The target fish on all 12 trials was a frowning fish. In the figure above, the target fish is outlined with a circle. The rectangular boxes highlight the locations of the jellyfish, which were always swimming in the direction of the target fish. The body of the target fish was always green.

The response latencies of accurate responses were recorded. The cue acquisition score was computed by first dividing mean response latency by accuracy in the first and last four trials. Scores in the first four trials should reflect performance during the initial exposure to cues. Given the novel stimuli, scores were expected to be relatively higher as a result of greater response latency and lower accuracy.

Scores in the last four trials should reflect performance following the successful or unsuccessful acquisition of the relevant cues. Successful acquisition was expected to be reflected in lower scores as a result of lower response latency and higher accuracy, where unsuccessful acquisition was expected to be reflected in higher scores.

The difference between the scores for the first and last four trials was calculated to obtain the ‘cue acquisition score’. Lower cue acquisition scores were expected to reflect higher levels of cue acquisition.

Task 2: Timed-Search Task (Restricted, Less Salient)

The design of the Timed-Search Task was similar to the Search Task, but exposure time to the stimuli was fixed. This created a limiting factor on the mapping and matching process, such that the acquisition of the relevant associations was necessary for successful performance. The task consisted of 12 trials. In each trial, participants were presented with an image displaying eight schools of fish. Each school embodied a unique colour. In one of the eight schools, a fish with an angry face was present (see Figure 3). The goal was to identify the target, angry fish, and determine the shape on its body. The image was presented for a period of 10 seconds. Following the presentation, participants were asked to indicate, from a list of choices, the shape that they thought was displayed on the body of the target fish.

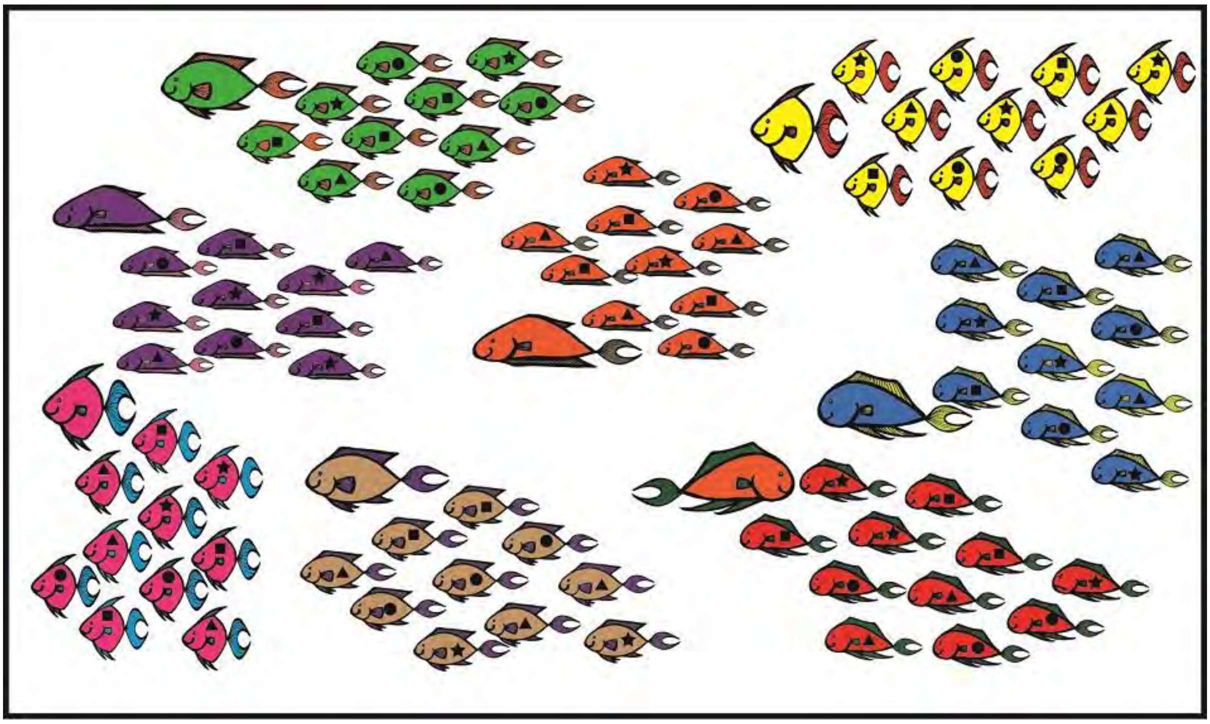

Figure 3. Example image presented on each trial of the Timed-Search Task. The configuration of the eight schools of fish differ for each image.

In this case, there were two relevant associations that could assist performance. First, the target fish was always in the school where the large, lead fish was facing the opposite direction. Second, the target fish was always located to the rear of the school (see Figure 4).

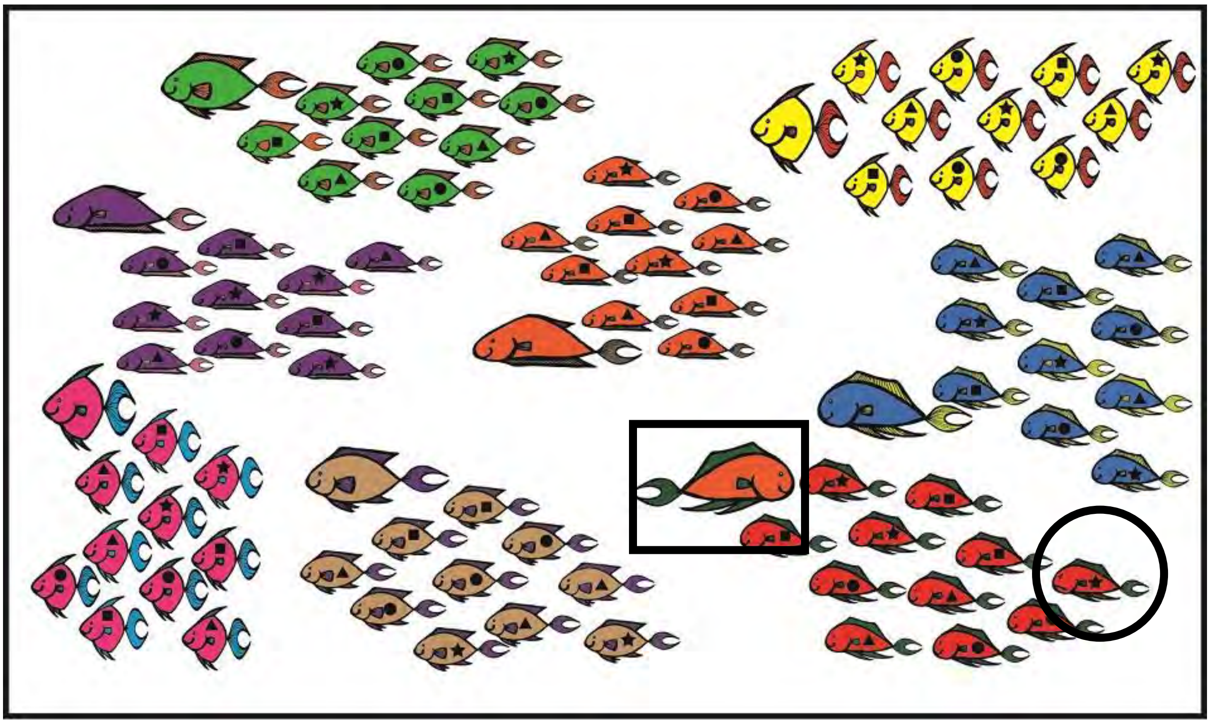

Figure 4. The target fish on all 12 trials was a fish with an angry face. In the above image, the target fish is highlighted with a circle. The rectangular box highlights the large lead fish, which was always facing the opposite direction. The target fish was always located at the back of the school.

For the Timed-Search Task, the accuracy of responses was recorded. The cue acquisition score was calculated by summing the learning score and the total accuracy score. The learning score was calculated by dividing the accuracy scores in the final four trials by the total accuracy score. The proportion of correct responses in the last four trials against the total correct responses was calculated to test whether performance improved across trials as a function of cue acquisition. Smaller scores were expected to reflect the rate of cue acquisition.

The addition of total accuracy scores to learning scores was intended to distinguish between smaller learning scores that reflect more rapid cue acquisition, which may occur in the first few trials, and smaller learning scores that reflect no learning across trials. The addition of greater total accuracy scores would reflect a greater rate of cue acquisition. Overall, greater cue acquisition scores were expected to reflect a higher level of cue acquisition.

Task 3: Explicit Association-Learning Task (Restricted, More Salient)

The Explicit Association-Learning Task was a two-part task involving an initial, untimed ‘study’ period followed by a ‘test’ period. Participants were informed that the test trials were based on the materials presented in the study sheet. Should the participants correctly derive the feature-event/object associations from the study sheet, performance on the tasks should be successful. Superior task performance was assumed to be reliant upon the acquisition of feature-event/object associations presented during the untimed study period prior to the commencement of the test task. Since the participants were informed that the task was based on the study sheet, measurement outcomes from the task represented explicit learning.

The study sheet presented during the study period displayed visual descriptions pertaining the diets and best friends of several fish (see Figure 5). During the study period, participants were instructed to examine the study sheet, and were specifically informed of the structure of the upcoming trials. Therefore, it was expected that participants would be aware that there were associations in the study sheet on which they would be tested.

The test task, following the study period, consisted of 12 trials. During the test task, participants were presented with an image of a fish not present in the study sheet, and either an image of a diet or best friend. Both images were presented simultaneously for a period of three seconds. The participants were then asked to rate, on a scale of one to seven, the extent to which the fish was related to the diet or best friend.

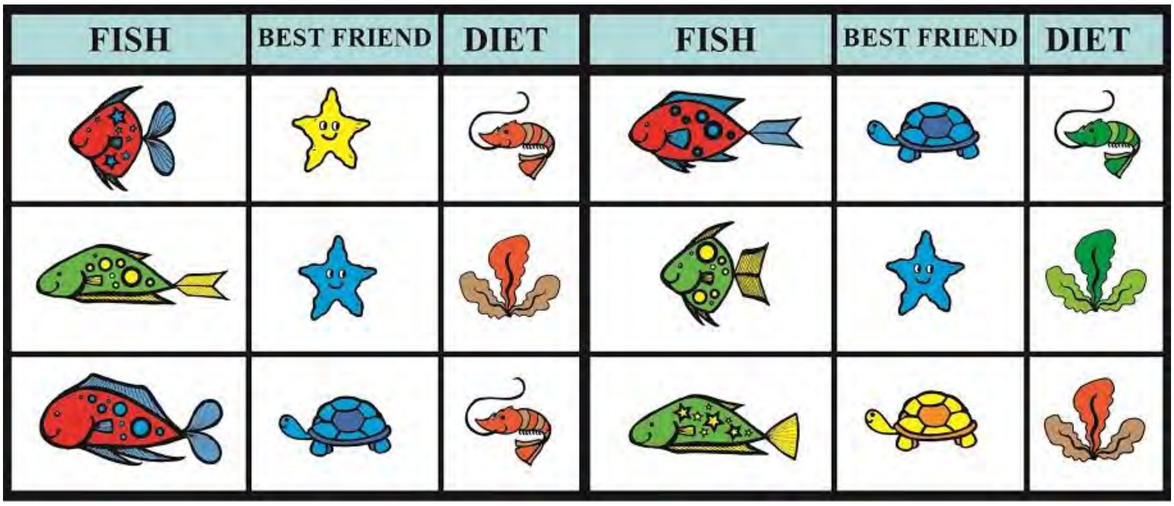

Figure 5. Study sheet presented in the Explicit Association-Learning Task

There were two independent associations that were included in the study sheet. First, all fish with red bodies had a diet of prawns of any colour, and all fish with green bodies had a diet of plants of any colour. Second, all fish with star patterns had yellow best friends regardless of the animal type, and all fish with circle patterns had blue best friends, regardless of the animal type (see Figure 6).

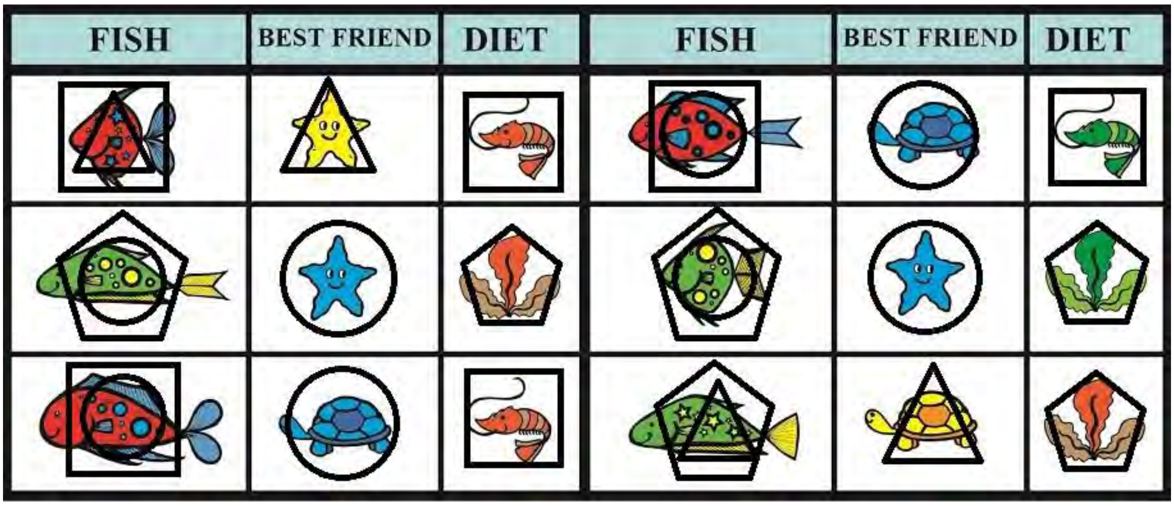

Figure 6. The associations in the study sheet are highlighted above. As indicated with triangles, fish with star shapes on their bodies had yellow best friends, and as indicated with circles, fish with round shapes on their bodies had blue best friends. As indicated with squares, fish with red body colour had diets of prawns, and as indicated with pentagons, fish with green body colour had diets of plants.

Rating scores were measured, and were used to derive accuracy scores. Rating scores ranged from one to seven with one being extremely unrelated and seven being extremely related. If participants rated an objectively unrelated item with one or two, and an objectively related item with six or seven, an accuracy score of one was assigned to the item. All other ratings that did not coincide with the accuracy scoring system were scored as zero. Scores on the test task were calculated by summing the accuracy scores. Greater scores were expected to reflect a higher level of cue acquisition.

Task 4: Problem Solving Task (Unrestricted, More Salient)

The Problem Solving Task consisted of six trials. In each trial, participants were presented with a short scenario, where they were required to select a fish, from a list of six fish, that they thought would match a fish that had lost a partner fish (see Figure 7). Following the decision, participants were asked to rate the importance of several features in arriving at their decision, including colour and shape of body. No time limit was imposed during both the evaluation and decision period.

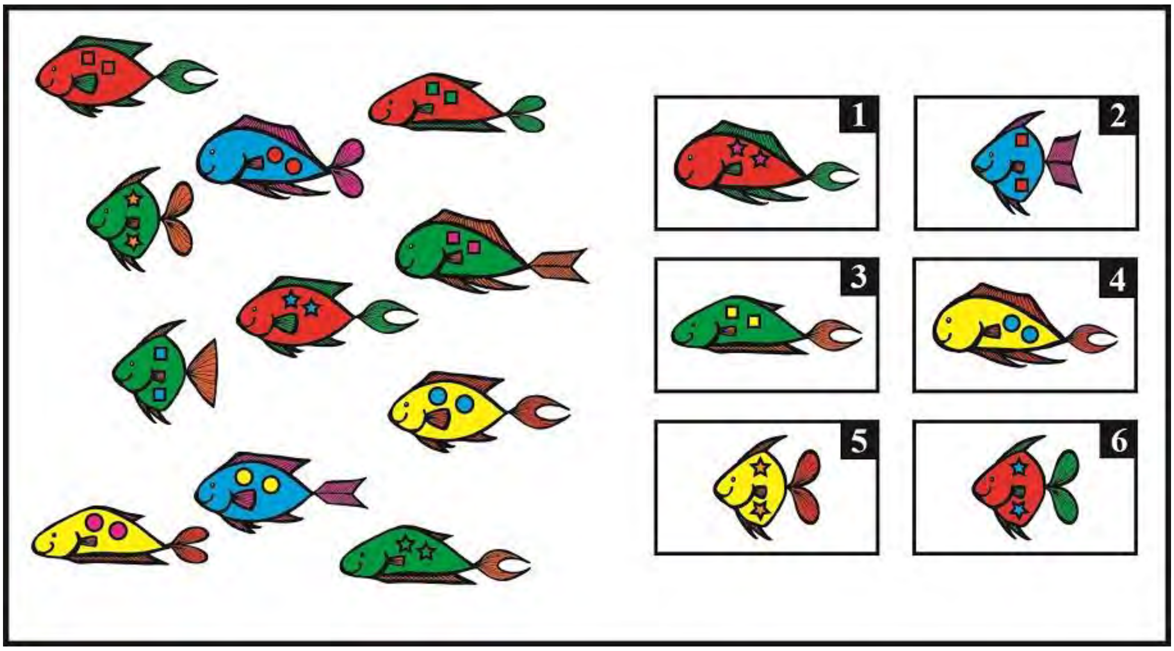

Figure 7. Example image presented on each trial of the Problem Solving Task.

There was one accurate feature-event/object association consisting of three characteristics: colour of body, pattern on body, and colour of pattern (see Figure 8). To select the correct response, participants were required to consider all three characteristics when evaluating the options.

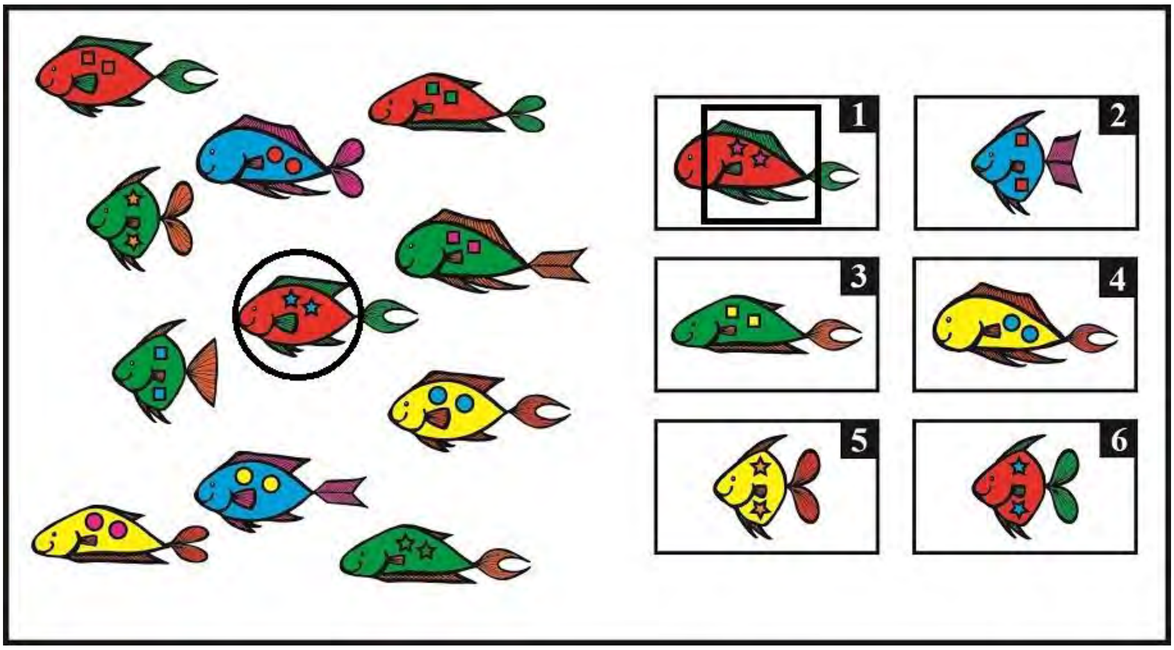

Figure 8. The fish to be matched must first be identified from the existing pool. In the figure above, the fish to be matched is highlighted with a circle. This fish can be identified by assessing the body colour and shape on the body of all the fish in the existing pool. The association governing this identification is that only fish with the same body colour and shape on body can match. The correct fish to be selected is highlighted with a square. This fish can be identified by assessing the body colour, shape on the body, and the colour of the shape of all the fish in the selection pool. The third association governing this identification is that fish with the same body colour and shape on body can only match if the shapes are of different colour.

Rating scores for all features were captured, the variance of all ratings aggregated for each trial, and the accuracy of responses determined. Scores in the first three trials were expected to represent initial performance following exposure to novel stimuli, while scores in the last three trials were expected to represent learned performance and the extent to which cues were acquired. Cue acquisition scores were computed by subtracting the mean variances of the first three trials from the last three trials, and then adding the total accuracy score. Greater cue acquisition scores were expected to reflect a higher level of cue acquisition.

Task 5: Classification Task (Unrestricted, More Salient)

The Classification Task was a two-part task involving an initial, untimed ‘study’ period followed by a ‘test’ period. Participants were informed that the test trials were based on the materials presented in the study sheet. Should the participants correctly derive the feature-event/object associations from the study sheet, performance on the tasks should be successful.

Prior to the commencement of the task, participants read a short scenario, and were presented with a study sheet, which was distinct from that presented in the Explicit Association-Learning Task (see Figure 9). The study sheet contained a 'personality' classification of a number of fish, which distinguished it from the associations incorporated into the other four tasks. No time limit was imposed during the study period.

Following the study period, participants proceeded to the test period, which consisted of six trials. In each trial, participants were asked to decide upon the personality classification of an unknown fish by accessing the details of the unknown fish from a list of tabs (see Figure 10). Each tab was labelled with an attribute (e.g., Body Shape), and would reveal the attribute details of the unknown fish when accessed. A time-limit of 60 seconds was imposed during the first two trials, 45 seconds in the subsequent two trials, and 30 seconds in the final two trials. Once the allocated time had elapsed or the participants were satisfied with the information that they retrieved, they selected a decision from a list of the six personality classifications, and then rated the importance of several features in arriving at that decision.

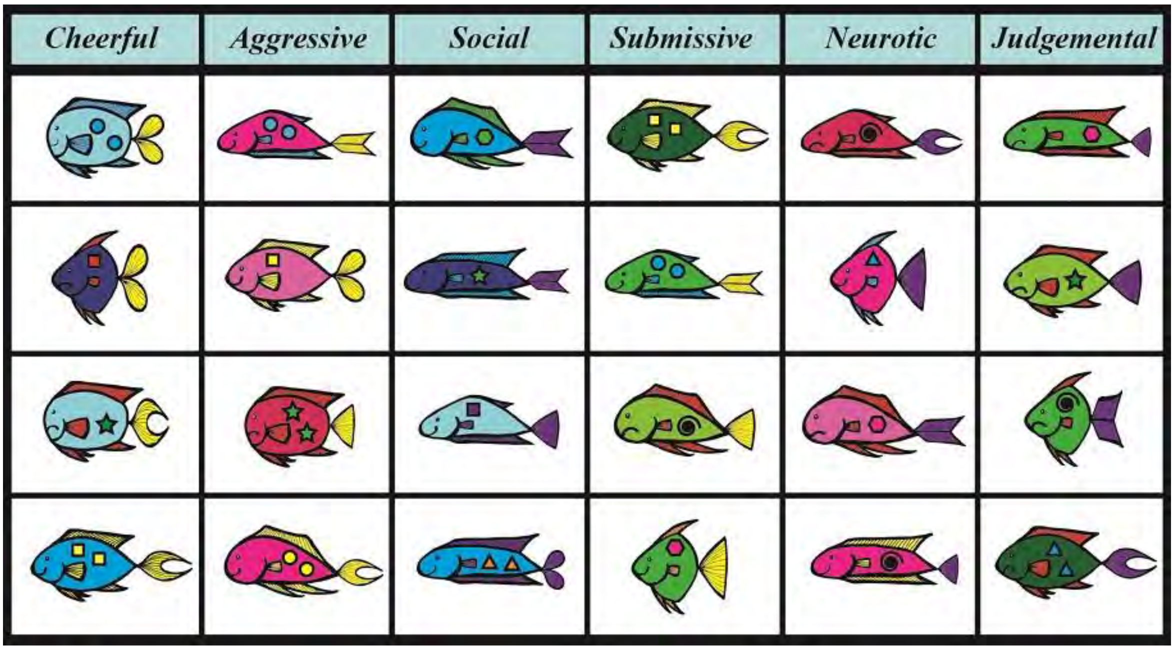

Figure 9. Study sheet presented in the Classification Task

In the Classification Task, one feature-event/object association was included, consisting of two characteristics: the colour of body, and the colour of tail and fins. Participants were required to access the tabs for both attributes to identify the correct classification for the unknown fish.

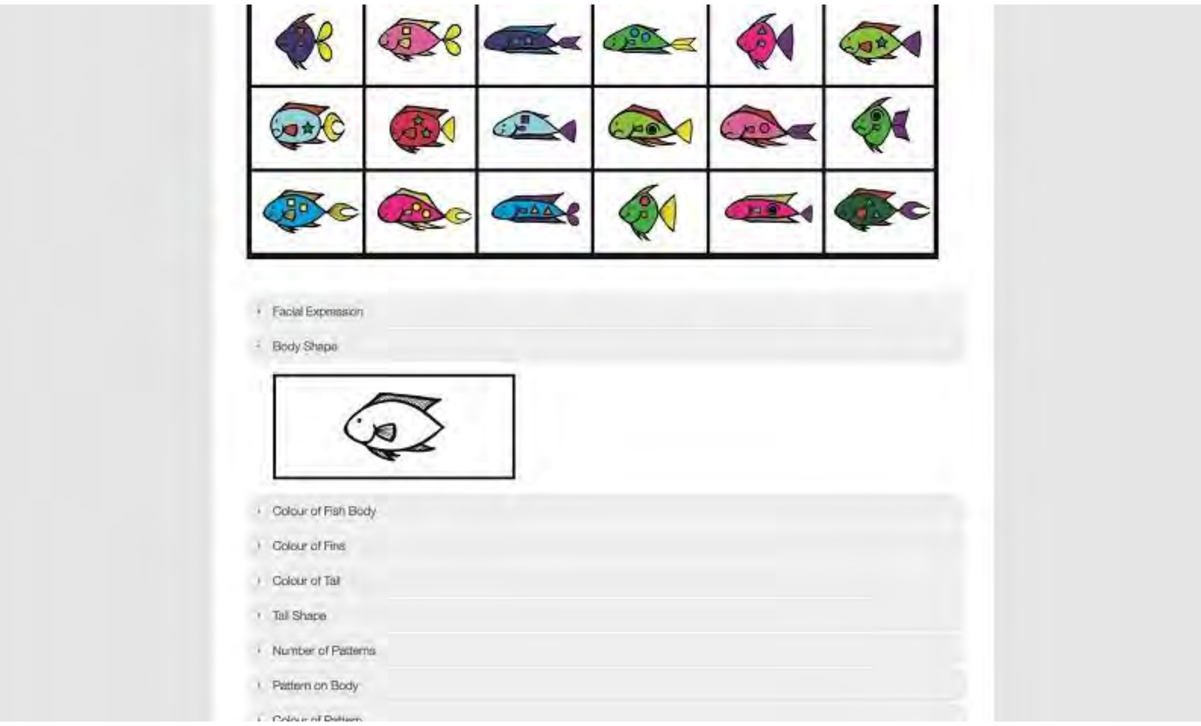

Figure 10. Details of the unknown fish were presented within tabs. In the figure above, the tab 'Body Shape' was accessed to reveal the body shape of the unknown fish.

The order in which information was accessed was measured on all trials, and converted into a single ratio metric by calculating the ratio of tabs accessed sequentially by the total number of tabs accessed. A sequential access of tabs suggests that no pre-determined priority was given to any information when accessing information, whereas a non-sequential access should suggest that some information was prioritised over others, indicating that cues may have been acquired to guide information search.

Cue acquisition scores were calculated based on the differences between the mean ratio of the first and last three trials. Smaller cue acquisition scores were expected to reflect a higher level of cue acquisition.

Rail Control Task

A low-fidelity, simulated rail control task served as a measure to assess the construct validity of the cue acquisition tasks developed in the present study. The task was presented to the participants on a computer screen within a laboratory setting.

The overall goal of the rail control task is for the participants to ensure that trains arrive at the correct destinations, rerouting trains where necessary. The rail control task consists of four horizontal green lines, representing railway tracks. At one end of each track, an intersection divides the track to two different destinations, labelled either 'Odd' or 'Even'. This intersection is depicted with white lines, and is controlled by a circular button labelled 'Change' located at the top of each track (see Figure 11).

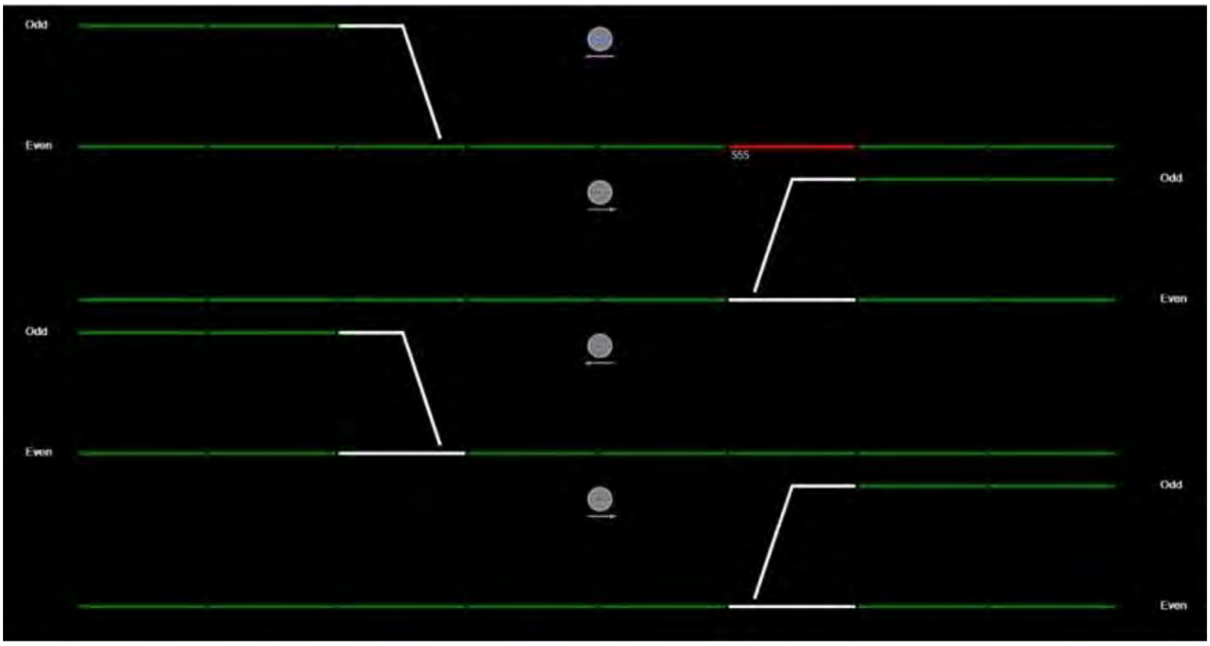

Figure 11. Visual display of the novel rail control task

A red horizontal bar, representing a train, appears on the track at the opposite end from the intersection, and travels towards the intersection (see Figure 11). Every train is labelled with a three-digit number that is either odd or even (e.g., 333, 888).

When a train appears on a track, one of the two white lines at the intersection turns green, depicting the programmed route of the train. By clicking on the 'Change' button, the white track will turn green, and the train is diverted onto the alternative track.

The participants' task is to ensure that trains arrive at the destination that corresponds with their labelled numbers (e.g., 333 to Odd destination). The programmed route does not always match the numbers on the train, and in these instances, participants are required to select the 'Change' button to ensure that the train runs on a consistent route, and thus, arrive at the correct destination. Feedback is presented to the participants in the form of red text that displays either 'Correct' or 'Incorrect' when the trains arrive at their destination.

When a train appears on a track, participants have seven seconds to change the intersection tracks, and the intersection can only be changed once for each train. Seventy-one of the total 154 trains presented in the task did not need rerouting. In the present study, the duration of the rail control task was 20 minutes.

Programmed into the task was a pattern of train movement so that trains running on two of the four tracks always required rerouting (e.g., tracks one and two), while trains on the other two would never require rerouting (e.g., tracks three and four). If participants successfully derived these patterns of train movement, accuracy would be greater and response latency lower.

Three measurement outcomes were recorded. Response latency was measured in milliseconds from the appearance of the train to the selection of the 'Change' button. The accuracy of responses was measured and determined based on whether the trains arrived at the correct destinations. A measure of pattern recognition was also derived from participants' verbal responses to the question "Did you notice a pattern in the rail control task?", and coded as either positive (correct pattern identification) or negative (incorrect pattern identification) based on their response.

Procedure

On arrival at the testing laboratory, participants were provided with an Information sheet and Consent form. All of the participants were required to complete a demographic questionnaire. Participants proceeded to complete the five tests of cue acquisition, which took approximately 30 minutes, after which they commenced the simulated rail control task. Participants were first given a 5-minute practice session on the rail control task, during which they were provided with standardised verbal instructions by the researcher. After the 5-minute practice session and once the participants indicated that they had sufficiently understood the task, the simulated rail control task was initiated. The duration of the rail control task was 20 minutes.

When the rail control task was completed, participants were asked by the researcher if they had noticed a pattern in the rail control task. If participants indicated that they had, they were asked to describe the pattern to ensure that the pattern had been derived accurately. Each session took approximately 60 minutes to complete and each participant was tested individually.

Results

Rail Control Performance

The accuracy of responses, together with response latency, constituted the measures of rail control performance. The frequency of errors ranged from zero to 14 ($M = 3.48$, $SE = 0.37$). Given that error rate was higher than the rate reported by Brouwers et al. (2017), rail control performance scores were adjusted for accuracy. Responses were sectioned into four blocks by five-minute intervals across the 20 minutes duration to assess changes in performance across time. Mean response latency for correct responses in each block was divided by the total frequency of correct responses in the respective blocks to derive an adjusted performance score. Smaller scores reflected improved rail control performance. Rail control performance in each block was normally distributed.

In addition to rail control performance, pattern recognition was measured. Pattern recognition constituted a dichotomous outcome variable with two levels: correct or incorrect.

The pattern was identified by 40.20% of the participants (n = 33). An independent sample t-test was conducted on the error frequency to assess whether error frequency was influenced by pattern recognition. There was a statistically significant difference in error frequency, $t(80) = 2.42$, SE = 0.73, p = .018, between participants who identified the pattern (M = 2.42, SE = 0.39), and those who failed to identify the pattern (M = 4.18, SE = 0.54). Inspection of the means indicated that participants who identified the pattern made fewer errors.

Rail Control Performance and Pattern Recognition

Consistent with Brouwers et al. (2017), a 4 x 2 mixed-repeated ANOVA, incorporating two levels of pattern recognition as a between-groups variable and the four blocks as a within-groups variable, was conducted on the adjusted performance score to examine the relationship between recognition of pattern and rail control performance. Correcting for violation of Mauchly's Test of Sphericity (p < .001), a statistically significant main effect was evident for Blocks, $F(2.54, 202.81) = 3.90$, MSe = 183.85, p = .014, $\eta_p^2 = 0.05$. Pairwise comparisons revealed that the significant difference lay between the adjusted performance score in the second five-minute block (M = 50.64, SE = 2.66) and the last five-minute block (M = 44.10, SE = 2.18; p = .012). Rail control performance tended to improve across time, $F(1, 80) = 4.80$, MSe = 207.31, p = .031, $\eta_p^2 = 0.06$) consistent with a learning effect. There was no statistically significant interaction between blocks and recognition of pattern ($F(2.54, 202.81) = 0.42$, p = .703, $\eta_p^2 = 0.01$; Table 2 and Figure 12).

Consistent with Brouwers et al. (2017), a statistically significant main effect was evident for pattern recognition, $F(1, 80) = 5.13$, MSe = 1208.26, p = .026, $\eta_p^2 = 0.06$. Inspection of the means indicated that participants who identified the pattern were faster (M = 42.77, SE = 3.03) than participants who did not report recognising the pattern (M = 51.64, SE = 2.48).

Table 2.
Mean (and Standard Error) of Adjusted Performance Score by Blocks and Pattern Recognition

	Block of Trials			
Pattern Recognised	Block 1	Block 2	Block 3	Block 4
Yes	44.21 (3.40)	44.90 (4.11)	42.23 (3.31)	39.76 (3.37)
No	51.61 (2.79)	56.39 (3.37)	50.11 (2.72)	48.45 (2.77)

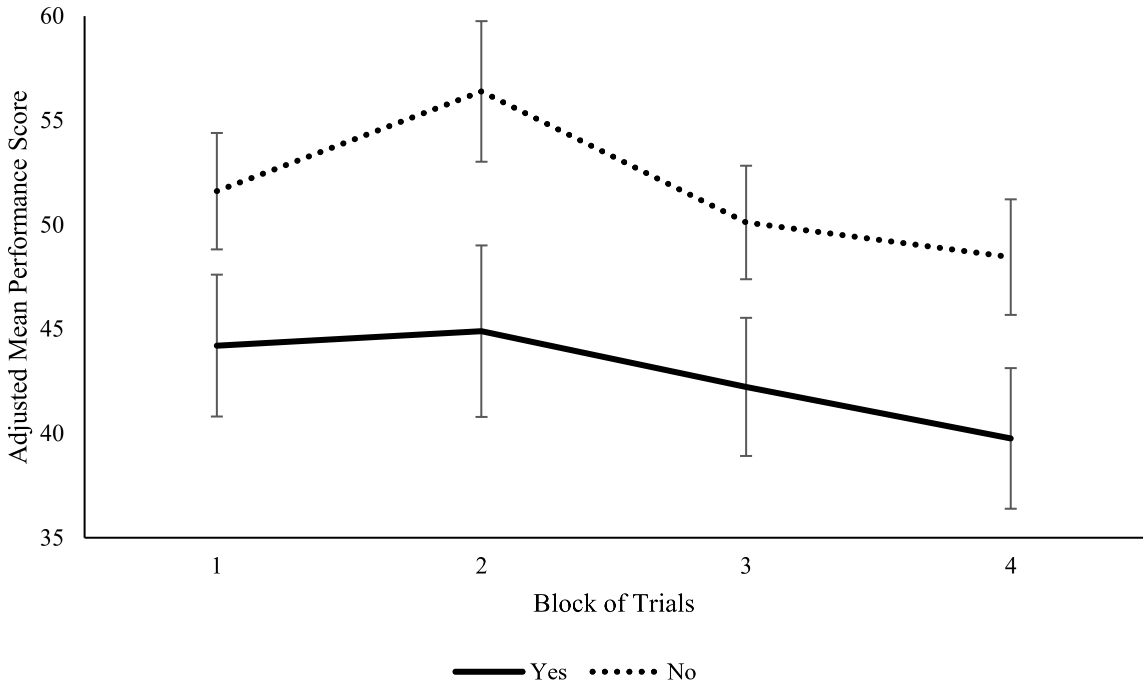

Figure 12. Mean Adjusted Performance Score (with Standard Error Bars) across the Four Blocks of Trials by Pattern Recognition.

Cue Acquisition and Rail Control Performance

The relationship between rail control performance and cue acquisition performance was analysed to examine the construct validity of the five tasks. Backward regression analysis was conducted on the mean adjusted performance score on the rail control task to test whether performance on any of the five cue acquisition tasks predicted performance on the rail control task.

Cue acquisition explained a significant proportion of variance in mean adjusted performance score, $F(5, 76) = 3.16$, $R^2 = 0.17$, $p = .012$. However, only the Timed-Search Task significantly predicted mean adjusted performance score, $\beta = -0.46$, $t(76) = -3.81$, $p < .001$. Mean adjusted performance score decreased by 0.46 for each increase in accuracy score on Timed-Search

Task. Table 2 presents the zero-order correlations and regression coefficients of all five cue acquisition tasks against mean adjusted performance score.

Table 2.
Regression Coefficients of the Cue Acquisition Tasks against Mean Adjusted Performance Score (n = 82)

	Zero-order *r*					
Cue Acquisition	Mean Rail RT	*B*	*SE B*	*β*	*t*	*P*
Search Task	-0.05	-2.37E-5	0.00	-0.04	-0.36	.719
Timed-Search Task	-0.38	-0.63	0.17	-0.46	-3.81	.000
Explicit Association-Learning Task	0.11	-0.67	1.14	-0.07	-0.59	.557
Problem Solving Task	-0.01	-0.52	0.40	-0.14	-1.28	.203
Classification Task	0.04	0.78	1.58	0.05	0.49	.626

Cue Acquisition and Pattern Recognition

To assess the predictive effects of the five cue acquisition tasks on pattern recognition, a logistic regression was conducted with pattern recognition as the dependent variable. The independent variables were scores on the five cue acquisition tests. Scores on all five tasks were normally distributed.

The intercept-only model correctly classified 59.80% of the participants. Chi-square goodness-of-fit test failed to reveal any statistically significant improvement in the model with the inclusion of the scores on the five cue acquisition tasks ($R^2 = 0.06$, $p = .629$). Only 5.60% of the variation in pattern recognition could be explained by the model incorporating the scores on the five cue acquisition tasks. Table 3 presents the Wald scores and odds ratios of the five cue acquisition tasks.

Table 3.
Logistic Regression Results of Cue Acquisition Tests on Recognition of Pattern

	B	*SE B*	*Wald*	*Exp(B)*	*95% CI*
Search Task	0.00	0.00	0.01	1.00	1.00 – 1.00
Timed-Search Task	0.10	0.09	1.37	1.11	0.94 – 1.31
Explicit Association-Learning Task	0.39	0.57	0.46	1.47	0.48 – 4.54
Problem Solving Task	0.34	0.21	2.52	1.40	0.92 – 2.13

Classification Task	0.21	0.80	0.07	1.23	0.26 – 5.93

Rail Control Performance, Cue Acquisition and Pattern Recognition

Given that only performance on the Timed-Search Task reliably predicted performance in the novel rail control task, the relationship between rail control performance, pattern recognition and Timed-Search Task scores was analysed separately.

A 4 x 2 mixed ANCOVA was conducted on the adjusted performance score on the rail control task. The analysis incorporated two levels of recognition of pattern as a between-groups variable, the four blocks of rail control performance as a within-groups variable, and the Timed-Search Task scores as a covariate score.

The Greenhouse-Geisser correction was adopted due to a violation of Mauchly's Test of Sphericity (p < .001). There was no statistically significant main effect for blocks, F(2.54, 200.61) = 1.50, MSe = 184.47, p = .221, η_p^2 = 0.02, and no significant interaction between blocks and pattern recognition (F(2.54, 200.61) = 0.45, p = .687, η_p^2 = 0.01), and blocks and Timed-Search Task performance, F(2.54, 200.61) = 0.60, p = .590, η_p^2 = 0.01.

A statistically significant main effect for pattern recognition was observed after controlling for the effect of Timed-Search Task performance, F(1, 79) = 4.90, p = .030. The covariate TST performance was significantly related to the adjusted performance score, F(1, 79) = 13.76, p < .001.

Pairwise comparison revealed a statistically significant difference between mean adjusted performance scores in Block 2 (M = 50.72, SE = 2.58) and Block 4 (M = 44.18, SE = 2.07; p = .013). Simple main effects analysis revealed a significant difference in Block 2 adjusted performance score between individuals who identified the pattern (*M* = 45.35, *SE* = 3.99) and those who did not (*M* = 56.08, *SE* = 3.27; *p* = .041; see Table 4).

Table 4
Mean (and Standard Error) of Adjusted Performance Score by Blocks and Pattern Recognition

	Block of Trials			
Pattern Recognition	Block 1	Block 2	Block 3	Block 4
Yes	44.80 (3.10)	45.35 (3.99)	42.64 (3.18)	40.22 (3.20)
No	51.21 (2.54)	56.08 (3.27)	49.83 (2.61)	48.14 (2.63)

Discussion

The aim of the present study was to design and evaluate the construct validity of five psychometric instruments that were designed to measure individual differences in cue acquisition. The tasks varied by exposure period and salience of characteristics between stimuli. Performance on each task was evaluated against performance on a novel, simulated rail control task.

Amongst the five tasks, only performance on the Timed-Search Task was associated with performance on the novel rail task. Greater scores on the Timed-Search Task were associated with faster response latency on the novel rail task. This finding suggests that higher cue acquisition, measured by the Timed-Search Task, was associated with improved performance on the novel rail control task.

In the novel rail control task, participants were tasked to ensure that trains were programmed to run on the correct tracks by rerouting trains as necessary. Embedded in the task was a pattern of train movements. The identification of the pattern of train movements was intended to facilitate task performance. In the present study, participants who verbally reported that they had identified the pattern of train movements recorded faster response latency in rerouting trains where necessary than participants who failed to verbally report that they had identified the pattern.

The recognition and utilisation of patterns underlies learning across a range of contexts (Eraut, 2000; Reber, 1989). For example, in learning a synthetic language, Reber et al. (1980) noted that participants were able to derive the abstract structures underlying the array of nonsense grammatical strings of letters without being informed that there was a structure in the array. Importantly, when participants were instructed about the presence of a pattern in the stimuli, they performed at a significantly lower level than participants for whom the presence of the pattern was not made explicit. Reber et al. (1980) argued that the absence of information pertaining to the pattern may have engaged an inherent capability for pattern acquisition, whereas the explicit reference to the pattern may have impeded this capability.

When presented with explicit instructions to search for patterns or associations within the stimuli, participants were more likely to construct inappropriate inferences about associations than those who were not presented with explicit instructions (Reber et al., 1980). Evidently, explicit knowledge of the presence of associations interferes with the capability for pattern acquisition by leading individuals to derive erroneous associations (D. Howard & Howard, 2001). Consequently, the outcome measure of an instrument incorporating explicit stimuli may not reflect inherent individual differences in cue acquisition.

In the present study, the Search Task and Timed-Search Task were the only tasks where the presence of features and events or objects that form the basis of cues, were not made salient to the participants. In both tasks, instructions pertaining to the presence of a pattern within the stimuli were not made explicit. The remaining tasks involved an explicit reference to the pattern to be identified during the task. Therefore, it might be argued that the Search Task and Timed-Search Task drew on capabilities that were also necessary for successful performance during the rail control task.

The capacity to identify patterns is presumed to occur spontaneously in the absence of conscious awareness (J. Howard, Jr. & Howard, 1997; Reber et al., 1980). Individual differences in pattern identification correspond to differences in performance on tasks where patterns must be discerned from an array (Reber et al., 1991; Vicari, Marotta, Menghini, Molinari, & Petrosini, 2003). These differences are evident in the timely and accurate detection of targets.

The outcome variable for both the Search Task and the novel rail control task was response latency, while the measurement outcome for the Timed-Search Task was accuracy. Although the Search Task and the novel rail control task shared the same measurement outcome, the results of the present study did not reveal any statistically significant relationship between performance on the Search Task and the novel rail control task. However, this observation might be attributed to differences in a design feature inherent in the two tasks.

Both the Search Task and the rail control task required participants to match the target objects, held in memory, to the array of stimuli until a match was identified. In the rail control task, this matching process occurred within a specified duration. By contrast, there was no time pressure in the Search Task.

Time pressure during decision making will result in increased cognitive load due to the restriction in the time available to process the same amount of information (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007). Consequently, changes in cognitive processes are necessary to adapt to the cognitive demands. This adaptation to increased cognitive demands is often reflected in increased selective attention to fewer cues or pieces of information that have been attributed greater importance (Gonzalez, 2004; Rothstein, 1986). Therefore, it might be argued that the differences in time pressure between the novel rail control task and Search Task may necessitate that participants adopt different cognitive processes, which was reflected in a lack of significant relationship between performances on the two tasks.

Given that the rail control task incorporated a time constraint, performance on the task should reflect individual differences in the cues selected and utilised during task performance. However, the selection and utilisation of cues must be preceded by the acquisition of cues since the task was novel to all the participants. Therefore, variations in performance on the novel rail task should reflect individual differences in cue acquisition. On the other hand, the lack of time constraint in the Search Task allowed for unrestrained cognitive processing, and hence, performance on the task may reflect other cognitive processes beyond cue acquisition.

Consistent with the rail control task, the Timed-Search Task incorporated a time constraint. The statistically significant relationship between performance on the Timed-Search Task and performance on the rail control task suggests that performance on the Timed-Search Task is likely to reflect individual differences in cue acquisition. This observation constitutes evidence for the construct validity of the Timed-Search Task as a measure of cue acquisition.

Overall, the findings from the present study demonstrated the validity of the Timed-Search Task as a measure of individual differences in cue acquisition. However, the instrument was administered and assessed within a controlled setting and with a non-operational sample. This meant that the validity of the Timed-Search Task within any operational or industrial settings is not clear, and must be further assessed for generalisability.

In high-risk operational settings, such as aviation, an instrument that measures individual differences in cue acquisition is a valuable supplementary tool during recruitment. Superior cue acquisition is expected to translate into a faster rate of learning, and in turn, reduce the costs of training. While the novel rail control task can be utilised for this purpose, the domain-general framework of the Timed-Search Task allows for a more widespread application within various operational settings.

To conclude, the present study was designed to construct and assess an instrument to assess individual differences in cue acquisition. Five tests of cue acquisition were designed and evaluated against performance on the novel rail control task. Of the five tasks tested, the Timed-Search Task, which incorporated restricted exposure period and less salient characteristics, emerged as the only predictive task of rail control performance.

References

Abernethy, B. (2008). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences, 8*(1), 17-34. doi:10.1080/02640419008732128

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 570. doi:10.1037/0278-7393.33.3.570

Bilalić, M., Langner, R., Erb, M., & Grodd, W. (2010). Mechanisms and neural basis of object and pattern recognition: A study with chess experts. *Journal of Experimental Psychology: General, 139*(4), 728. doi:10.1037/a0020756

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics*, 1-55. doi:10.1080/00140139.2017.1330494

Brouwers, S., Wiggins, M. W., Helton, W., O'Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. *Frontiers in Psychology, 7*. doi:10.3389/fpsyg.2016.00435

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55-81.

Chung, P. H., & Byrne, M. D. (2008). Cue effectiveness in mitigating postcompletion errors in a routine procedural task. *International Journal of Human-Computer Studies, 66*(4), 217-232. doi: 10.1016/j.ijhcs.2007.09.001

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology, 70* (1), 113-136. doi: 10.1348/000709900158001

Fracker, M. L. (1988). A theory of situation assessment: implications for measuring situation awareness. *Proceedings of the Human Factors Society Annual Meeting, 32*(2), 102-106. doi:10.1177/154193128803200222

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors: The Journal of Human Factors and Ergonomics Society, 46*(3), 449-460. doi:10.1518/hfes.46.3.449.50395

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. doi:10.1177/154193120104500411

Howard, D. V., & Howard, J. H., Jr. (2001). When it does hurt to try: Adult age differences in the effects of instructions on implicit pattern learning. *Psychonomic Bulletin & Review, 8*(4), 798-805. doi:10.3758/Bf03196220

Howard, J. H., Jr., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging, 12*(4), 634-656. doi:10.1037/0882-7974.12.4.634

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology: Section A, 54*(4). 1105-1124. doi: 10.1080/713756001

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456-460. doi:10.1518/001872008x288385

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making, 4*(3), 186-209. doi:10.1518/155534310X12844000801203

Loveday, T., Wiggins, M. W., & Searle, B. J. (2014). Cue utilization and broad indicators of workplace expertise. *Journal of Cognitive Engineering and Decision Making, 8*(1), 98-113. doi:10.1177/1555343413497019

Mann, D. T. Y., Williams, M. A., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology, 29*(4), 457-478. doi:10.1123/jsep.29.4.457

Maule, A. J., Hockey, G. R., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica, 104*(3), 283-301. doi: 10.1016/S0001-6918(00)00033-0

McCormack, C., Wiggins, M. W., Loveday, T., & Festa, M. (2014). Expert and competent non-expert visual cues during simulated diagnosis in intensive care. *Frontiers in Psychology, 5*, 949-955. doi:10.3389/fpsyg.2014.00949

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*(1), 115-119. doi:10.1037/h0027454

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-235. doi:10.1037/0096-3445.118.3.219

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 492. doi:10.1037/0278-7393.6.5.492

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(5), 888-896. doi: 10.1037/0278-7393.17.5.888

Renshaw, P. F., & Wiggins, M. W. (2017). The predictive utility of cue utilization and spatial aptitude in small visual line-of-sight rotary-wing remotely piloted aircraft operations. *International Journal of Industrial Ergonomics, 61,* 47-61. doi: 0.1016/j.ergon.2017.05.014

Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes, 37*(1), 83-92. doi: 10.1016/0749-5978(86)90045-2

Simon, H., & Chase, W. (1988). Skill in chess. In *Computer Chess Compendium* (pp. 175-188). Springer: New York, NY.

Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia, 41*(1), 108-114. doi:10.1016/s0028-3932(02)00082-9

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science, 65*, 118-124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., & O'Hare, D. (2003). Expert and novice pilot perceptions of static in-flight images of weather. *The International Journal of Aviation Psychology, 13*(2), 173-187. doi:10.1207/S15327108IJAP1302_05

Bridging Section

Study 1 was designed to construct a series of instruments and evaluate their effectiveness in assessing individual differences in cue acquisition. Five distinct tasks were designed, varying on exposure period and saliency of characteristics, and evaluated against task performance on a novel rail control task that incorporated an implicit pattern of train movements. Of the five tasks, only performance scores on the Timed-Search Task (TST), which incorporated restricted exposure period and less salient feature-event/object associations within the stimuli, was associated with task performance on the novel rail control task. Given that performance on the novel rail control task is influenced by the extraction and utilisation of an implicitly embedded patterns of associations, the positive association observed between TST scores and performance on the novel rail task provided some support for the TST as an instrument that captures individual differences in the acquisition of task-relevant feature-event/object associations.

The aim of Study 2 was to extend the outcomes of Study 1 by establishing the construct validity of the TST against performance on a composite measure of cue utilisation in the context of general aviation. To support effective situation assessment, higher cue utilisation relies on the availability of relevant feature-event/object associations in memory. The rate at which relevant feature-event/object associations are acquired is presumed to be influenced by differences in the capacity and opportunities for cue acquisition (Crane et al., 2018; Loveday, Wiggins, Searle, Festa, & Schell, 2012; Todd & Thomas, 2012). In general aviation, opportunities for cue acquisition can be compared based on the number of flight hours that pilots have accumulated. This tends to be a more robust measure of exposure than subjective estimates of experience.

The author of the present thesis contributed approximately 80% to the preparation of this study in data collection, data analysis, and writing of the paper.

CHAPTER FIVE

The Relationship between Cue Acquisition and Cue Utilisation in General Aviation

Abstract

Cue utilisation during situation assessment supports the construction of more accurate and rapid responses. However, the relationship between the initial acquisition of cues and their later utilisation is not yet clear. In the present study, performance on the Timed-Search Task, a previously developed measure of cue acquisition, was assessed against a composite measure of cue utilisation in the context of general aviation. Forty-four pilot trainees from three universities in Australia and New Zealand completed the cue acquisition and cue utilisation tasks online. The results indicated that cue acquisition was only related to performance on the Feature Identification Task and the Feature Recognition Task that comprised the broader measure of cue utilisation. This outcome suggests that individual differences in cue acquisition are possibly associated with the more rudimentary aspects of the process of cue utilisation.

Keywords: cue acquisition, cue utilisation, cues, general aviation

The Relationship between Cue Acquisition and Cue Utilisation in General Aviation

In complex domains, effective decision-making is reliant upon the accurate assessment of often ambiguous and dynamic situations. Situation assessment is characterised by the derivation and integration of information in the environment to construct an understanding or make sense of a series of events (Horrey & Wickens, 2001). This understanding acts as the basis on which action plans are derived or retrieved from memory (Klein, Calderwood, & Clinton-Cirocco, 2010). Therefore, more accurate situation assessment tends to generate more appropriate action plans, resulting in superior decision performance (Wiegmann, Goh, & O'Hare, 2002).

Within high-risk environments such as aviation, the accuracy and timeliness of situation assessment is associated with the levels of cognitive load that are experienced (Sweller, van Merrienboer, & Paas, 1998; Wickens, 2008). Cognitive load is the interaction between the cognitive demands imposed by the situation and the cognitive resources available to operators at the time of the assessment (Barrouillet, Bernardin, & Camos, 2004). A combination of high cognitive demands and low cognitive resources will generate greater levels of cognitive load, which in turn, increases the likelihood of operator errors in response to system changes (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007).

The demands on cognitive load can be further compounded by increases in time pressure that often coincide with unexpected changes in the environment (Maule, Hockey, & Bdzola, 2000; Vine, Uiga, Lavric, Moore, Tasneva-Atanasova, & Wilson, 2015). Under time pressure, there is an inclination to engage heuristic-based strategies as they obviate the requirement for slower and more effortful analytical approaches. However, accurate heuristic-based strategies depend upon a repertoire of cue-based associations in memory to support inductive reasoning and pattern recognition (Alison, Doran, Long, Power, & Humphrey, 2013; Gigerenzer & Gaissmaier, 2011; Svenson, Edland, & Slovic, 1990).

Cues are associations in long term memory (LTM) between features in the environment and events or objects (Kaempf, Klein, Thordsen, & Wolf, 1996). The utilisation of cue-based associations supports situation assessment by affording operators the opportunity for rapid assessments by accurately and rapidly matching perceived features against existing schemas in LTM (Klein, 2008). When a match is detected, the operator forms an understanding of the presenting situation based on the information within the matched framework without the need to undertake an extensive and time-consuming analysis of the situation.

Schemas in LTM are templates comprising a variety of feature-event/object associations that are relevant to the situations within the domain-specific environment (Ghosh & Gilboa, 2014). These feature-event/object associations are acquired through individual experiences and interactions with the domain-specific environment (Wiggins, 2012). Repeated and consistent exposure to the co-occurrence of features and events in the environment results in the construction of paired-associations between the features and events or objects (Kim, Seitz, Feenstra, & Shams, 2009).

The domain-specificity of feature-event associations is such that cue utilisation that supports performance in one domain would not normally be expected to influence cue-dependent performance in a different domain. However, recent evidence suggests that cue utilisation in one domain does, in some cases, relate to performance in another domain (Brouwers, Wiggins, Griffin, Helton, & O'Hare, 2017; Brouwers, Wiggins, Helton, O'Hare, & Griffin, 2016; Renshaw & Wiggins, 2017). For example, Renshaw and Wiggins (2017) observed that individuals with behaviours consistent with greater cue utilisation in operating a motor vehicle, acquired cue-based relationships more rapidly in learning to fly an unmanned aerial vehicle. Similarly, Brouwers et al. (2017) noted a positive relationship between cue utilisation in operating a motor vehicle and performance in a novel rail control task. These relationships suggest the existence of individual differences that regulate the acquisition of feature-event/object associations across domains, particularly during the initial stages of skill acquisition.

The capacity to acquire cue-based associations rapidly in a novel domain may constitute a trait that draws on intrinsic cognitive processes such as pattern recognition. Since the natural environment is often characterised by a range of features and events – the associations between which are not always clearly apparent – it is the ability to distinguish patterns of associations from 'noise' that will be most advantageous in the acquisition of cues.

Consistent with this proposition, Brouwers et al. (2017) examined the relationship between cue utilisation in operating a motor vehicle against performance in a novel rail control task, which incorporated implicit patterns of train movement. The derivation of these patterns of train movement from the stimuli was expected to facilitate task performance. Brouwers et al. (2017) noted that individuals with greater cue utilisation in operating a motor vehicle were faster in responding to the novel rail control task. More importantly, individuals who recorded faster response latency in the rail control task were also significantly more likely to verbally report having recognised the patterns embedded in the task.

Given the novelty of the rail control task, performance could not be contingent on any rail control associations in memory, and therefore, was not the result of cue utilisation. Instead, variations in performance were attributed to individual differences in the capacity to identify patterns of associations in the environment. Consequently, it might be argued that individual differences in the recognition of associations regulate the acquisition of critical cues, on which future cue utilisation functions.

The recognition of patterns of associations in the environment appear to underlie the acquisition of cues at the initial stages of learning (Brouwers et al., 2017). Given that the capacity to identify patterns of associations occurs spontaneously (Reber, Kassin, Lewis, & Cantor, 1980), variations in performance outcomes are, arguably, the result of individual differences in the rate at which associations in the environment are isolated and identified. This individual difference in the recognition of patterns ultimately determines the rate of cue acquisition, and consequently, the accuracy of situation assessment.

In high-risk industries such as aviation, assessing individual differences in cue acquisition may assist the recruitment of operators who might require less training to achieve requisite levels of performance. This, in turn, may reduce the cost of training and decrease the rates and/or severity of accidents and incidents.

In Study 1 (Chapter 4), the Timed-Search Task (TST) was developed as an instrument to measure individual differences in cue acquisition. The TST is a domain-general instrument with feature-event/object associations implicitly incorporated into the task. The derivation of these associations is expected to result in greater TST scores. The initial construct validity of the TST was established by evaluating these scores against performance in a novel rail control task.

The present study was designed to further examine the construct validity of the TST. While Study 1 (Chapter 4) assessed the construct validity of the TST against performance on a novel task, the present study evaluated performance on the TST against cue utilisation where participants had already acquired experience. General aviation was selected as the context for cue utilisation on the basis that measures of exposure are relatively accurate and reliable. To demonstrate construct validity, it was hypothesised that performance on the TST would be positively associated with cue utilisation within the general aviation domain, taking into account hours of flight experience.

Methodology

Participants

The participants comprised 44 undergraduate pilot trainees recruited from three universities in Australia and New Zealand. The age range was 18 to 27 (M = 20.20, SE = 0.33). Table 1 illustrates the gender distribution across the three universities.

The inclusion criterion comprised students enrolled in a Bachelor of Aviation (or equivalent) course at either of the three universities. Following approval by the ethics committees at the three universities, pilot trainees were recruited through self-selection, and were compensated with a chance to win one of three iPad minis.

Table 1.
Gender Distribution across the Three Universities

	University 1 (n = 25)	University 2 (n = 11)	University 3 (n = 8)	Total
Male	23	11	6	40
Female	2	0	2	4

Pilot trainees from three universities in Australia and New Zealand reported a range of flight experience from 10.50 to 327 hours (M = 112.17, SE = 13.31). Twenty-six pilot trainees had flown their first solo, while 18 pilot trainees had not flown their first solo at the time of testing. Table 2 illustrates the distribution of pilot trainees who had and had not flown solo across the three universities.

Table 2.
Distribution of Pilot Trainees who Had and Had Not Flown Solo

	University 1 (n = 25)	University 2 (n = 11)	University 3 (n = 8)	Total
No	13	4	1	18
Yes	12	7	7	26

Materials

Cue utilisation in general aviation was assessed via the EXPERT Intensive Skills Evaluation online platform (EXPERTise 2.0; Wiggins, Loveday & Auton, 2015). The platform is a shell software comprising five different tasks relevant to the measurement of domain-specific cue utilisation, including a Feature Identification Task, Feature Recognition Task, Feature Association Task, Feature Discrimination Task, and Feature Prioritisation Task.

As shell software, EXPERTise 2.0 has been used to assess cue utilisation in a range of different operational domains, including audiology (Watkinson, Bristow, Auton, McMahon & Wiggins, 2018) and power control (Loveday, Wiggins, Harris, Smith, & O'Hare, 2013). The tasks administered in the present study were relevant to the domain of general aviation.

The classification of higher and lower cue utilisation based on EXPERTise scores has demonstrated good test-retest reliability (Loveday, Wiggins, Festa, Schell, & Twigg, 2013; Watkinson et al., 2018), construct validity (Loveday et al., 2014; Loveday, Wiggins, Harris, et al., 2013; Wiggins et al., 2014), and predictive validity (Watkinson et al., 2018).

Feature Identification Task (FIT)

During the Feature Identification task, participants were presented with an image of an electronic flight instrument display (Figure 1). Across 15 trials, they were tasked to identify and select a feature of concern as quickly as possible. The aim of the Feature Identification Task is to assess individual differences in the ability to extract critical features from a visual display.

Response latency was measured during each trial. Mean response latency was calculated over the 15 trials. Lower mean response latency is presumed to be associated with higher cue utilisation, while greater mean response latency is associated with lower cue utilisation (Loveday et al., 2013).

Figure 1. Example image of an electronic flight display presented on each trial of the Feature Identification Task

Feature Recognition Task (FRT)

On each of the 15 trials in the Feature Recognition Task, participants were presented with an image of an electronic flight display, similar to the stimuli presented in FIT, for a period of 1000 milliseconds. On a separate screen, participants were asked to select the orientation of the aircraft from nine multiple-choice options: climbing left turn, climbing right turn, wings level climb, wings level descent, descending left turn, descending right turn, straight level flight, level left turn, or a level right turn.

The aim of the Feature Recognition Task is to assess individual differences in the ability to extract critical features from a visual display, under time constraint. In the present study, accuracy was measured in each trial and summed across the 15 trials. Greater accuracy is presumed to be associated with higher cue utilisation, while lower accuracy score is associated with lower cue utilisation (Wiggins & O'Hare, 2003).

Feature Association Task (FAT)

During the Feature Association Task, participants were presented with two terms related to the context of general aviation. The terms were presented simultaneously for 1000 milliseconds. On a separate screen, participants were tasked to rate on a Likert scale from 1 (Extremely unrelated) to 10 (Extremely related), the extent to which they thought the two terms were related. Participants were asked to respond to 15 trials. Example terms included 'Piston' and 'Propeller', and 'Pressure' and 'Temperature'.

Variance in ratings and mean response latency were measured across the 15 trials. Each participant received an overall score that represented variance in ratings as a proportion of mean response latency. A greater score is typically associated with higher cue utilisation, while lower score is associated with lower cue utilisation (Morrison et al., 2013).

Feature Discrimination Task (FDT)

In the Feature Discrimination Task, participants were presented with two written scenarios. In each scenario, participants were advised that they had been requested to fly to specified destinations, and were provided with flight-related information, such as the nature of the terrain at the destination, weather conditions, and the type of aircraft that they would be operating. They were then asked to select their initial response to each situation from four possible options. The options included 'Plan the flight as requested', 'Seek advice from colleagues before making a decision', 'Collect more information before making a final decision', or 'Reject the request to conduct the flight'.

Once the decisions were selected, participants were asked to rate on a Likert scale from 1 (Extremely irrelevant) to 10 (Extremely relevant), the extent to which each of the 11 features that were included in each scenario, such as wind conditions, cloud cover, and terrain, were relevant to their decision.

Variance in ratings across the 11 features was measured for each scenario. Mean variance ratings across the two scenarios was calculated where a greater mean variance is associated with higher cue utilisation, and a lower mean variance is associated with lower cue utilisation (Pauley, O'Hare, & Wiggins, 2009).

Feature Prioritisation Task (FPT)

In the Feature Prioritisation Task, participants were provided a brief context to a scenario, and were subsequently asked to select further information from a drop-down list to inform their decision response. The drop-down list from which participants gained further information comprised 13 tabs which were labelled with relevant features, such as ‘Weather’, ‘Loadsheet’, and ‘Map’. When a tab was selected, information pertaining to that feature was displayed (see Figure 2). Participants were limited to the selection of 10 tabs, and 120 seconds to access the information. Once the available time is up and/or 10 tabs have been selected, participants were required to select, on a separate screen, their decision response from a list of four options: ‘Fly the flight as planned’, ‘Postpone the flight’, ‘Cancel the flight’ and ‘Fly to the destination via a different route’.

The order in which tabs were accessed and the total number of tabs accessed were recorded. A single ratio metric was calculated by computing the ratio of the number of pairs of tabs accessed sequentially as a proportion of the total number of pairs of tabs accessed. A smaller ratio metric is presumed to be associated with higher cue utilisation, while greater ratio metric is associated with lower cue utilisation. These scoring values were based on the observations that higher cue utilisation is associated with accessing fewer but critical information, whereas lower cue utilisation is associated with accessing information in the order it is presented (Wiggins & O'Hare, 1995).

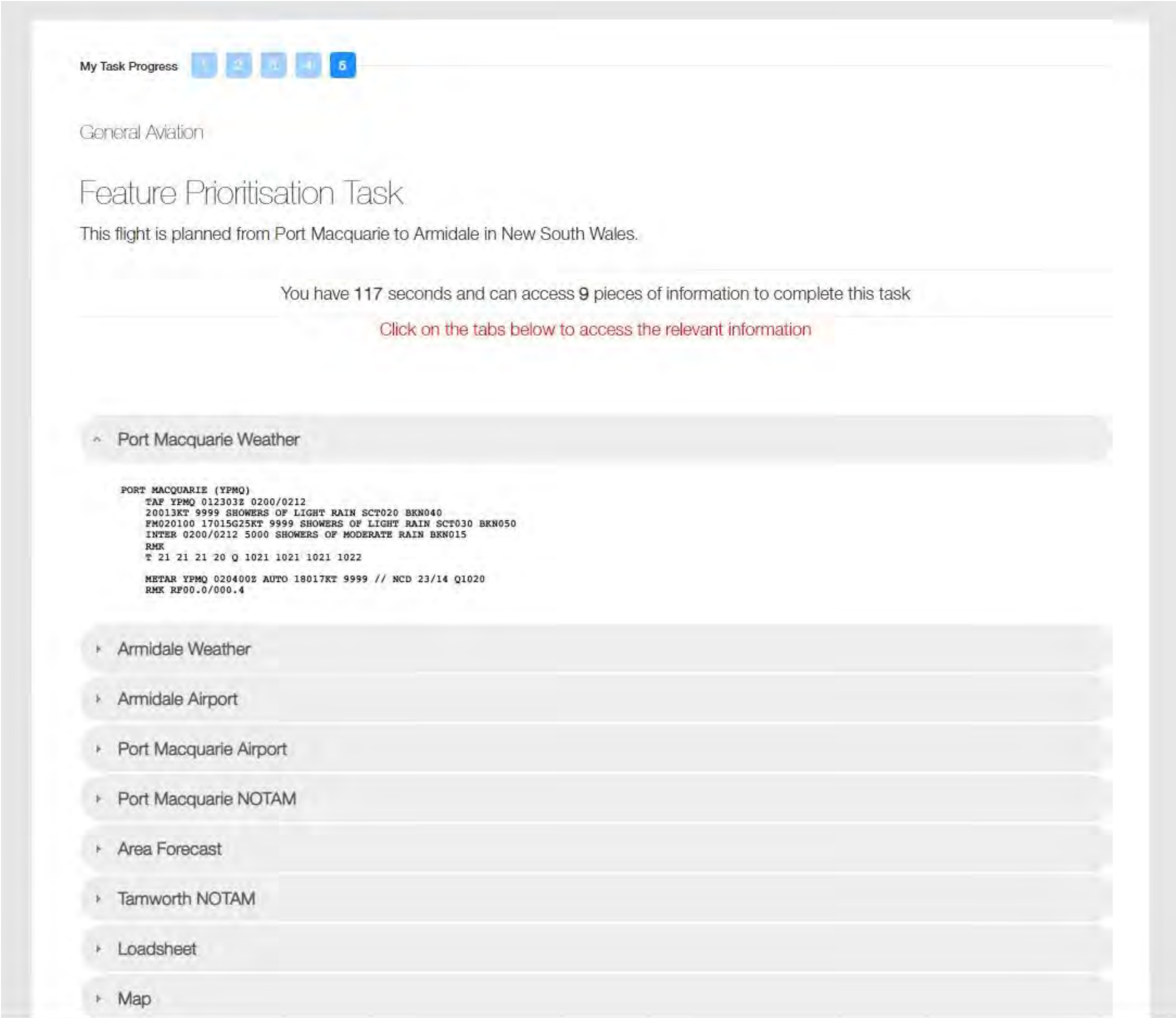

Figure 2. Example drop down list from which participants can acquire information to make their decision

Timed-Search Task (TST)

The Timed-Search Task was set within a fictitious domain not influenced by individual experience. The TST comprised 12 trials. On each trial, participants were required to visually search for a target item in a display illustrating eight schools of fish (see Figure 3). The target item was a small fish with an angry face.

Once the target fish was identified, the participants were asked to determine the shape that was imprinted on its body. The display was presented for 10 seconds. On a separate screen, the participants were asked to select, from a list of five choices, including ‘Square’, ‘Circle’, ‘Triangle’, ‘Star’ and ‘Did not see’, the shape that they had observed on the target fish. The option ‘Did not see’ was available in situations where participants failed to identify the target fish or the pattern imprinted on its body.

The location of the target fish was associated with two features within the display such that the acquisition of one or more of these features should be associated with an increase in response accuracy. The first feature associated with the location of the target fish was the lead fish. The target fish was always in the school led by the right-facing lead fish. Second, the target fish was always located towards the rear of the school.

Response accuracy was measured for each trial. Accuracy in the last four trials was divided by the total accuracy score to derive the rate of learning score. The TST score was then calculated by summing the learning score to the total accuracy score. This was intended to account for more rapid cue acquisition which may occur in the first few trials. Greater TST scores were presumed to be associated with a higher level of cue acquisition.

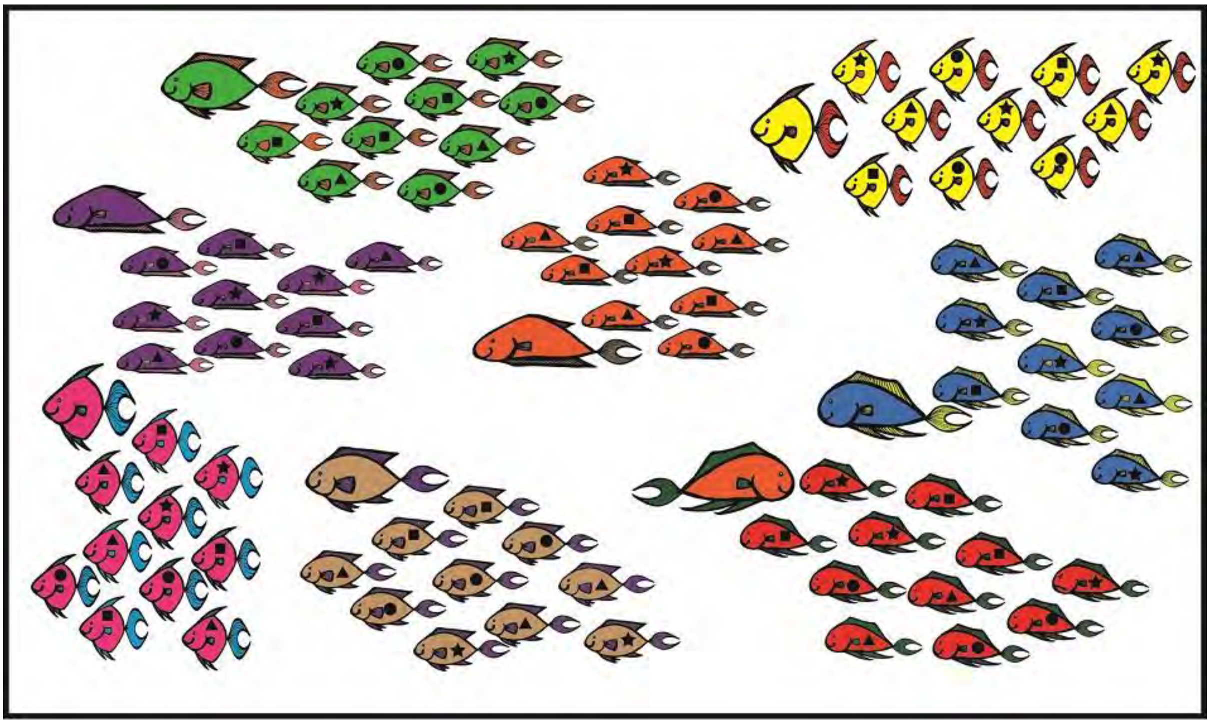

Figure 3. Example image presented on each trial of the Timed-Search Task.

Procedure

An electronic invitation was circulated to all students enrolled in the Bachelor of Aviation programs at the three universities. The invitation included a brief summary of the study, instructions on how to access the study, and a link to access the online software EXPERTise 2.0. Internet access and a desktop or laptop were required to participate in the study.

The TST was incorporated into the EXPERTise 2.0 platform so that participants were only required to complete one battery of tests. Participants first completed a demographic questionnaire, comprising of age and gender, and a flight-related questionnaire, comprising of affiliated institutions, total flight hours and solo flight experiences, followed by the TST, and finally the cue utilisation tasks comprising the FIT, FRT, FAT, FDT and FPT in the given order. All of the participants participated in the study in their own time. The study took approximately 30 to 45 minutes to complete.

Results

To assess the construct validity of the TST against the construct of cue utilisation, performance scores on the TST were evaluated in relation to the individual dimensions of cue utilisation, and the overall level of cue utilisation. Cue utilisation within the domain of general aviation was operationalised as: (1) performance on the five general aviation cue utilisation tasks, and (2) typologies representing either higher or lower cue utilisation. Performance scores on the five cue utilisation tasks were normally distributed.

To establish two distinct cue utilisation typologies, a *K-means* cluster analysis was conducted on the standardised performance scores on the five general aviation cue utilisation tasks. This approach is consistent with the standard calculation of cue utilisation typologies (Brouwers et al., 2017; Wiggins, Azar, Hawken, Loveday, & Newman, 2014). $K = 2$ clusters was utilised based on sample size and previous research. Performance on the general aviation

EXPERTise tasks yielded two typologies. Table 3 presents the mean centroids of the five cue utilisation tasks for both clusters.

Cluster 1 consisted of 17 participants with relatively slower response latency on Feature Identification Task, higher accuracy scores on Feature Recognition Task, greater variance as a proportion of response latency on the Feature Association Task, greater variance score on the Feature Discrimination Task, and a smaller ratio score on the Feature Prioritisation Task. With the exception of performance on the Feature Identification Task, this pattern of performance is generally consistent with higher cue utilisation. Cluster 2 consisted of 27 participants with pattern of performance consistent with relatively lower cue utilisation.

Independent sample *t*-tests were conducted on cue utilisation scores for each of the five cue utilisation tasks to assess whether cue utilisation performance differed between the two clusters. Scores on all five cue utilisation tasks were significantly different between the lower and higher cue utilisation typologies (see Table 3).

Table 3.
Mean Centroids of the Five Cue Utilisation Tasks

	Cluster 1 (n = 17)	Cluster 2 (n = 27)			
	Higher cue utilisation	Lower cue utilisation	*t*	Df	*p*
FIT	0.56	-0.35	3.24*	42	.002
FRT	0.47	-0.30	2.66*	42	.011
FAT	0.54	-0.34	3.09*	42	.004
FDT	0.49	-0.31	2.80*	42	.008
FPT	-0.48	0.30	-2.69*	42	.010

*Significant at the .05 level (two-tailed)

The relationship between flight experience and cue utilisation typology was examined to assess whether levels of cue utilisation varied as a function of flight experience. A Log10 transformation was conducted on hours of experience to achieve normality. An independent sample *t*-test was conducted with cue utilisation typology (higher or lower) as the between subjects variable and hours of flight experience as the dependent variable. No statistically significant difference in flight experience was evident on the basis of cue utilisation, $t(42) = 0.15$, $p = .882$.

An ANCOVA was undertaken on the TST scores to evaluate the relationship between cue acquisition and cue utilisation taking into account hours of flying experience. The independent variable comprised the cue utilisation typology in general aviation. The covariate was flight experience in hours and the dependent variable comprised scores on the TST. There was no statistically significant difference in TST scores between higher ($M = 5.14$, $SE = 0.81$) and lower ($M = 5.57$, $SE = 0.65$) cue utilisation typologies, $F(2, 40) = 0.17$, MSe = 11.12, $p = .680$, $\eta_p^2 = 0.00$.

Given that cue acquisition was not associated with general aviation cue utilisation typology, the relationships between cue acquisition and the five dimensions of cue utilisation, accounting for hours of experience, were analysed to evaluate whether cue acquisition was related to one or more of the underlying dimensions of cue utilisation. Multiple linear regression was conducted on the TST scores with scores on the five cue utilisation tasks and hours of experience serving as the predictor variables.

Cue utilisation accounted for 28.50% of the variance in TST scores, $F(5, 37) = 2.40$, MSe = 8.94, $p = .047$. However, only performance on the FIT ($\beta = -0.42$, $t(37) = -2.62$, $p = .013$) and the FRT ($\beta = 0.36$, $t(37) = 2.38$, $p = .023$) significantly predicted TST scores. Scores on the TST decreased by 0.42 for each increase in a unit of response latency on the FIT, and increased by 0.36 for each increase in accuracy score on the FRT. Table 4 presents the zero-order correlations and regression coefficients of all five cue utilisation tests and hours of experience against TST scores.

Table 4.
Regression Coefficients of the Cue Utilisation Tests and Flight Hours against TST (n = 43)

	Zero-order *r*			
Cue Acquisition	TST	*β*	*t*	*p*
FIT	-0.34	-0.42	-2.62	.013
FRT	0.27	0.36	2.38	.023
FAT	0.26	0.24	1.48	.147
FDT	0.09	-0.12	-0.75	.456
FPT	0.01	-0.18	-1.03	.311
Hours	0.09	-0.03	-0.19	.848

Discussion

The present study was designed to evaluate the construct validity of the TST as a measure of individual differences in cue acquisition. Performance on the TST was assessed against a composite measure of cue utilisation in the context of general aviation. Measures of cue utilisation included: (1) higher or lower cue utilisation typologies based on the patterns of performance on the general aviation edition of EXPERTise 2.0, and (2) performance scores on each of the five cue utilisation tasks. Overall, there was no association between scores on the TST and the cue utilisation typologies. However, a negative relationship was evident between scores on the TST and response latency on the FIT, while a positive relationship was evident with accuracy on the FRT.

The lack of a relationship between scores on the TST and the composite measure of cue utilisation may have been due to the nature of the typologies, since the observed pattern of performance was not completely consistent across the five cue utilisation tasks (Brouwers et al., 2017; Loveday, Wiggins, Festa, & Schell, 2013). For example, participants classified with higher cue utilisation would normally be expected to be characterised with a lower response latency on the FIT, greater accuracy on the FRT, a greater variance per unit time on the FAT, greater variance on the FDT, and a low ratio on the FPT. However, in the present study, participants who, in other tasks displayed performance consistent with higher cue utilisation, recorded a centroid for the FIT that was relatively higher than participants whose performance across the other tasks was more consistent with lower cue utilisation. This suggests that the edition of EXPERTise 2.0 that was developed in the context of general aviation failed to discriminate effectively between participants on the basis of their cue utilisation.

Separate analyses of performance across the five cue utilisation tasks suggested that the nature of the relationship between the FIT and performance on the TST was consistent with expectations, since a higher TST score was associated with a lower response latency on the FIT. The nature of this relationship corresponds to the outcomes of previous research (See Chapter 4) where greater performance on the TST is associated with a lower response latency in response to mis-routed trains. Similarly, the relationship between performance on the TST and scores on the FRT were consistent with expectations, with higher scores on the TST associated with greater accuracy on the FRT. Therefore, higher cue acquisition was related to faster response latency in identifying areas of concern within an electronic flight instrument display, and to a greater ability in extracting critical features from an electronic flight instrument display under time pressure.

The pattern of the results associated with the present study suggests that the TST may hold utility for specific markers of cue utilisation, rather than as a measure of overall performance. Since these associations occurred independent of hours of exposure, the findings suggest that there may be underlying capabilities which were employed when performing both the TST and the cue utilisation tasks. However, it may also reflect a common-method bias associated with the two instruments, since the TST incorporate measures of both accuracy and response latency.

Importantly, in the present study, the domain was familiar and therefore, performance on the FIT and FRT should have been associated with hours of exposure. However, an inspection of the responses indicated that, for the FRT, a ceiling effect was evident, which was likely to have restricted the range of responses. This may have influenced the classification of cue utilisation typologies.

The methodological issues that emerged in the present study likely limit the extent to which the outcomes are generalisable. Nevertheless, it does appear to be evident that a domain-general measure of cue acquisition is associated, at least in part, with some measures of a task-specific version of cue utilisation. Future initiatives are necessary to test measures of performance beyond cue utilisation in isolation, and determine whether cue acquisition predicts operational performance.

Common-method biases associated with the measure of cue utilisation and cue acquisition can be reduced by administering the battery of tests within a controlled laboratory setting. Given that the TST and cue utilisation tasks incorporated elements of time constraint, performance on these tasks is more susceptible to bias in a self-administration environment compared to measures that are not time constrained. Therefore, future investigations that rely on measures of cue acquisition and cue utilisation can benefit from administering the tasks within a regulated environment where exposure to distractor stimuli, such as noise and social interruption, is limited.

The benefit of controlled laboratory settings extends to a more precise evaluation of operational performance. In determining the predictive influence of cue acquisition and cue utilisation on operational performance, a laboratory-based assessment allows for greater experimental control over the operational tasks undertaken by participants. This will, in turn, maintain consistency of exposure across participants, and thereby, reduce the impact of extenuating factors on operational performance.

Finally, it may be useful to examine whether performance on the domain-general measure of cue acquisition is similarly associated with aspects of domain-specific cue utilisation in a domain distinct from general aviation. Consistency in findings across different domains, after controlling for methodological issues, would suggest the generalisability of the association between cue acquisition and cue utilisation.

If individual differences in cue acquisition is observed to be a good predictor of cue utilisation within general aviation in future studies, the TST could be a beneficial supplementary instrument when selecting pilots. Pilots who possess a greater capacity to acquire cues would be more adaptable to novel conditions and are more likely to utilise cues when assessing unexpected situations. Consequently, the TST, as an instrument that is relatively easy to administer, could offer utility in the context of aviation.

In conclusion, the present study was designed to assess the construct validity of the TST as a measure of individual differences in cue acquisition. Performance on the TST was measured against general aviation cue utilisation. Cue acquisition was positively associated with the ability to identify domain-specific critical features, but did not predict cue utilisation typologies more generally.

References

Alison, L., Doran, B., Long, M. L., Power, N., & Humphrey, A. (2013). The effects of subjective time pressure and individual differences on hypotheses generation and action prioritization in police investigations. *Journal of Experimental Psychology: Applied, 19*(1), 83. doi:10.1037/a0032148

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83-100. doi:10.1037/0096-3445.133.1.83

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 570. doi:10.1037/0278-7393.33.3.570

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics, 60*(11), 1-55. doi:10.1080/00140139.2017.1330494

Brouwers, S., Wiggins, M. W., Helton, W., O'Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. *Frontiers in Psychology, 7*, 435. doi:10.3389/fpsyg.2016.00435

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81. doi:10.1016/0010-0285(73)90004-2

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104-114. doi:10.1016/j.neuropsychologia.2013.11.010

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451-482. doi:10.1146/annurev-psych-120709-145346

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. doi:10.1177/154193120104500411

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making, 4*(3), 186-209. doi:10.1518/155534310x12844000801203

Loveday, T., Wiggins, M., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In Latorre C. P., Sánchez J., Fred A. (eds) *Pattern Recognition - Applications and Methods. Advances in Intelligent Systems and Computing*, *204* (pp. 1-11). Springer, Berlin, Heidelberg.

Loveday, T., Wiggins, M., Harris, J. M., Smith, N., & O’Hare, D. (2013). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors, 55*(1), 90-107. doi: 10.1177/0018720812450911

Maule, A. J., Hockey, G. R., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica, 104*(3), 283-301. doi: 10.1016/S0001-6918(00)00033-0

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 492. doi:10.1037/0278-7393.6.5.492

Renshaw, P. F., & Wiggins, M. W. (2017). The predictive utility of cue utilization and spatial aptitude in small visual line-of-sight rotary-wing remotely piloted aircraft operations. *International Journal of Industrial Ergonomics, 61,* 47-61. doi: 0.1016/j.ergon.2017.05.014

Svenson, O., Edland, A., & Slovic, P. (1990). Choices and judgments of incompletely described decision alternatives under time pressure. *Acta Psychologica, 75*(2), 153-169. doi:10.1016/0001-6918(90)90084-s

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping, 28*(4), 467-477. doi: 10.1080/10615806.2014.986722

Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology, 57*(9), 681-687. doi: 10.1080/14992027.2018.1476782

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455. doi: 10.1518/001872008X288394

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 189-197. doi:10.1518/0018720024497871

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science, 65*, 118-124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology-Applied, 1*(4), 305-320. doi:10.1037/1076-898x.1.4.305

Bridging Section

The construct validity of the TST as a measure of individual differences in cue acquisition was tested in Studies 1 (Chapter 4) and 2 (Chapter 5). In Study 1 (Chapter 4), TST scores were positively associated with task performance in a novel rail task. This outcome suggests that TST scores reflect individual differences in the ability to identify relevant implicit feature-event associations.

In Study 2 (Chapter 5), TST scores were only associated with the Feature Identification Task (FIT) and Feature Recognition Task (FRT) scores in a cue utilisation measure within the domain of general aviation. Greater TST scores were associated with faster response latency on the FIT and greater accuracy on the FRT, suggesting that higher cue acquisition is associated with faster speed of identifying relevant features and a more accurate recognition of relevant features. Therefore, despite a number of methodological limitations, the outcomes of Study 2 (Chapter 5) provide limited support for the proposition that individual differences in cue acquisition may contribute to the development of cue utilisation through the identification and recognition of relevant feature-event associations.

Study 3 (Chapter 6) was designed to investigate the relationship between cue acquisition and performance in driving. Since previous experience or exposure to domain-specific environment facilitates the acquisition of relevant feature-event associations (Randel, Pugh, & Reed, 1996), it was reasoned that participants, matched for exposure, would show differences in the cue utilisation and performance that would be related to their capacity for cue acquisition. A version of this study has been accepted for publication in the Journal of Safety Research. The author of the present thesis contributed approximately 75% to the preparation of this study in data collection, data analysis, and writing of the paper.

CHAPTER SIX

Cue Acquisition, Cue Utilisation, and Performance on a Simulated Driving Task

Abstract

The influence of cue utilisation on domain-specific performance has been well established across a number of domains. However, the influence of the initial acquisition of cues on later domain-specific task performance remains unclear. In the present study, the relationships between cue acquisition, cue utilisation and task performance were examined in the context of driving. Performance measures included driving errors and visual search behaviours, including fixations and saccades. Seventy-one undergraduate psychology students with at least one-year driving experience were recruited via self-selection. Higher cue utilisation was associated with fewer driving errors, fixations, and saccades after accounting for cue acquisition. Cue acquisition, however, was not associated driving errors, fixations, or saccades. The results suggest that the influence of cue acquisition may not be evident on later task performance following the initial acquisition of feature-event associations.

Keywords: Cue acquisition; Cue utilisation; Driving; Fixation; Saccade

Cue Acquisition, Cue Utilisation, and Performance on a Simulated Driving Task

Road accidents are, in part, the result of a mismatch between the drivers' driving capabilities and the demands imposed by the traffic conditions (Fuller, 2005). This mismatch occurs when drivers fail to adapt to task demands that are greater than their driving capabilities. To effectively adapt to difficult situations, drivers can influence the complexity of task demands by adjusting their behaviours, including reducing their speed and/or increasing the distance from other vehicles (de Craen, Twisk, Hagenzieker, Elffers, & Brookhuis, 2008).

The selection of driving behaviours that influence task demands is, in part, dependent upon the precise assessments of the presenting situations (de Craen et al., 2008). In comparison to experts, novice drivers tend to have less precise judgments of situations and task demands, which is presumed to culminate in poorer driving performance (Engström, Gregersen, Hernetkoski, & Keskinen, 2003). More specifically, they appear to differ from expert drivers in the accuracy and efficiency with which they assess situations (Chapman & Underwood, 1998).

Situation assessment is a cognitive process involving the extraction and assimilation of information in the environment and in memory to form an evaluation of a situation (Horrey & Wickens, 2001). This evaluation guides the selection or construction of behavioural responses appropriate to the demands of the various situations (Klein, Calderwood, & Clinton-Cirocco, 2010). More precise situation assessments tend to generate more suitable behavioural responses, resulting in improved performance.

Given that the process of situation assessment involves the interaction between the individual and the environment, the precision and accuracy with which situations are assessed is contingent upon individual factors, including exposure to the environment and inherent ability (Fracker, 1988; Wiegmann, Goh, & O'Hare, 2002). Exposure in particular, enables the acquisition of feature-event/object associations (McCormack, Wiggins, Loveday, & Festa, 2014). Through repeated and consistent exposure to the co-occurrence of features and events or objects, paired-associations between the features and events or objects are developed and retained in memory (Kim, Seitz, Feenstra, & Shams, 2009).

Feature-event/object associations facilitate the rapid construction of assessments by retrieving previously experienced events and learned behaviours associated with the perceived feature (Klein, 2008). During situation assessment, features in the environment are matched against representations of features in memory. When a match occurs, information associated with the perceived feature is retrieved to form an understanding of the presenting situation based on previously experienced situations (Lipshitz, Klein, Orasanu, & Salas, 2001).

For instance, when experienced drivers perceive the brake lights of a leading car illuminate (feature), they are signalled to the expectation that the car is decelerating (event). Amongst naïve drivers, the association between the brake lights and deceleration may not exist or may be less well established. As a consequence, the construction of assessments is delayed, and the assessments may be unsuitable for the specific situation, resulting in inaccurate responses.

When features in the environment trigger the retrieval of feature-event/object associations from memory, they function as cues (Kaempf, Klein, Thordsen, & Wolf, 1996). Wiggins (2012) proposed that the utilisation of cues requires at least five distinct processes. These include the capacity to rapidly identify key features, the capability to accurately recognise key features, the ability to rapidly and precisely differentiate between features, the capability to discriminate between relevant and irrelevant features, and the capacity to prioritise the acquisition of key features.

The utilisation of cues during situation assessment can be inferred from visual search behaviours, including fixations and saccades, typically measured with eye tracking devices (Duchowski, 2002). Fixations are stable states during which the eyes are motionless in order to acquire visual information, while saccades are rapid movements of the eyes between fixation points (Rayner, 1998). Individual variations in fixations and saccades are thought to be contingent on domain-specific experience and the complexity of the visual scenes (Crundall & Underwood, 1998; Tatler, Baddeley, & Vincent, 2006).

In various domains, such as driving (Crundall, Underwood, & Chapman, 1999) and construction (Dzeng, Lin, & Fang, 2016), experts and novices differ in the characteristics of their visual fixations in response to task demands of varying complexities. In comparison to novices, experts tend to record greater fixations when undertaking more complex tasks (Crundall & Underwood, 1998), and fewer fixations when performing less complex tasks (Dzeng et al., 2016; Krupinski, Graham, & Weinstein, 2013; Tien et al., 2015).

When performing a routine, less complex task, experts appear to rely on existing associations or heuristics in memory to direct their visual attention towards fewer, but more relevant features (Smuc, Mayr, & Windhager, 2010). By contrast, novices may not have acquired the feature-event associations necessary to distinguish greater from less relevant features, culminating in greater number of visual fixations to acquire sufficient information to construct appropriate assessments.

The distinction between experts and novices is also evident in the characteristics of their visual saccades. Experts tend to record fewer visual saccades than novices (Arthur, Khuu, & Blom, 2016). Given that no visual information is presumed to be encoded during saccadic eye movements (Mann, Williams, Ward, & Janelle, 2007), the presentation of fewer saccades suggests a greater scope for the acquisition of information in response to changes in the environment. Consequently, variations in cue utilisation can be inferred from the frequency of fixations and saccades when identifying cues from an array of features in the environment.

For features in the environment to function as cues, it is necessary for operators to have acquired feature-event/object associations relevant to the situations in which they will be utilised. Consequently, feature-event/object associations that bolster cue utilisation in one domain may not be applicable in another domain.

The domain-specific nature of feature-event/object associations suggests that the advantages afforded by cue utilisation should not be associated with performance in another domain. However, previous research has demonstrated a relationship between cue utilisation in one domain and performance outcome in a different domain (Brouwers, Wiggins, Griffin, Helton, & O'Hare, 2017; Brouwers, Wiggins, Helton, O'Hare, & Griffin, 2016; Renshaw & Wiggins, 2017). Higher cue utilisation in a familiar domain has been associated with improved performance outcomes in a separate domain, suggesting a fundamental trait of cue acquisition that underlie cue utilisation across domains.

The present study was designed to investigate the influence of individual differences in cue acquisition on cue utilisation and performance in the context of driving. To assess cue acquisition, the Timed-Search Task (TST) was developed in Study 1 (Chapter 4) as a domain-general instrument. The TST incorporates implicit patterns of associations, which when derived, were expected to result in improved TST scores, reflecting faster rates of cue acquisition. The construct validity of the TST has been demonstrated previously in Studies 1 (Chapter 4) and 2 (Chapter 5).

In the current study, performance was assessed in a driving simulator and cue utilisation was assessed based on activities relating to driving, including hazard awareness and wayfinding. Based on the outcomes of Study 2 (Chapter 5), it was hypothesised that cue acquisition would moderate the relationship between cue utilisation in the context of driving and driving performance in a simulator. It was also hypothesised that cue acquisition would moderate the relationship between cue utilisation and the frequency of fixations and saccades during a simulated driving task.

Methodology

Participants

Following approval by the Macquarie University Human Research Ethics Committee (Appendix C), 71 first year Psychology students were recruited through convenience sampling to participate in the study. Two participants reported experiencing simulator sickness during the driving simulator task and their data were excluded from subsequent analysis. The remaining participants comprised 13 males and 56 females with an age range of 18 to 25 (M = 19.19, SE = 0.16). Participants were required to have acquired a minimum of one-year experience driving a motor vehicle, and be no older than 25 years of age. These constraints were imposed to control for domain-specific experience. Participants were excluded if they self-reported colour blindness.

Participants ranged in driving experience from 14 to 76 months (M = 38.55, SE = 1.55), and spent between one to 21 hours a week driving (M = 7.57, SE = 0.55). The total hours spent playing video games in a week ranged from zero to 15 hours (M = 1.14, SE = 0.32).

Materials

The battery of instruments included the Timed-Search Task and the driving edition of EXPERTise 2.0. A driving simulator was used to present visual stimuli, and eye tracking glasses were used to capture visual behaviour metrics.

Timed-Search Task (TST)

The TST is a domain-general instrument developed to assess cue acquisition. The task incorporates implicit associations, and the stimuli are fictitious to ensure that task performance is not influenced by individual experience. Any variations in performance were presumed to be a reflection of individual differences in the acquisition of cue-based associations.

For each of the 20 trials, participants were asked to search for a target object, which was a small fish with an angry face, in a visual display consisting of eight schools of fish (see Figure 1).

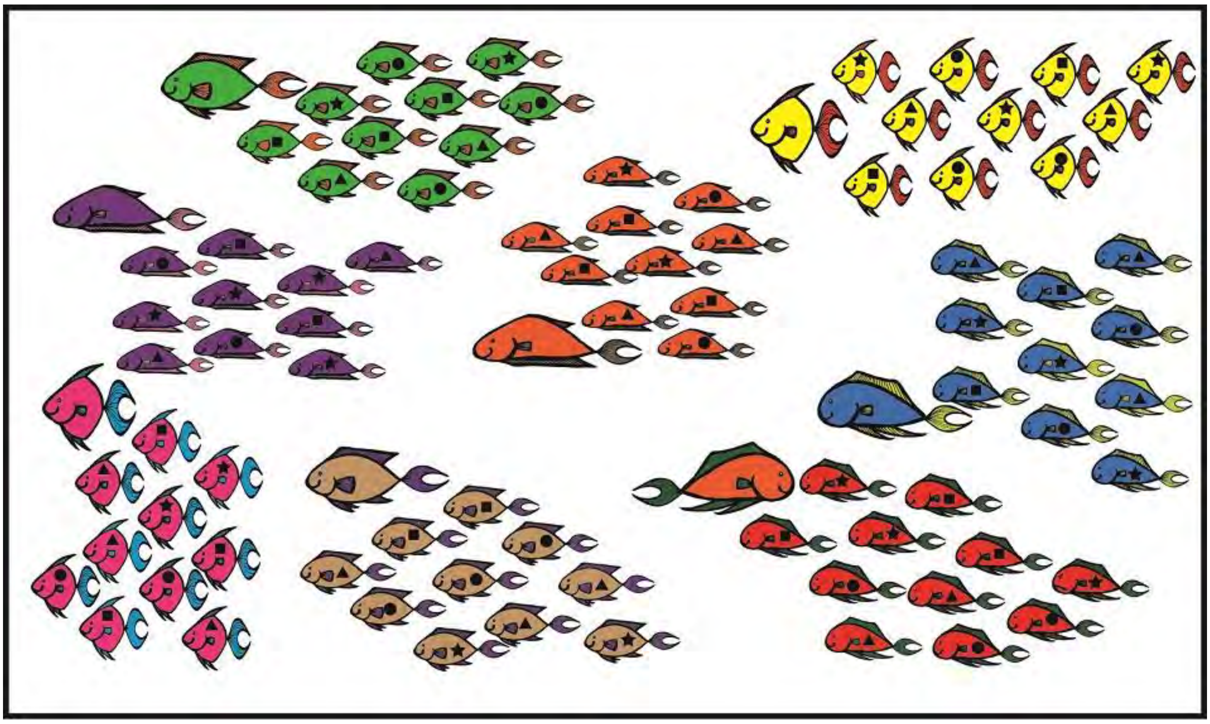

Figure 1. Example image presented on each trial of the Timed-Search Task.

When participants successfully located the target fish, they were required to retain in memory, the shape that was imprinted on the body of the target fish. Each visual display was presented for 10 seconds, after which, the participants were asked to select, from a list of five options, the shape that they had identified on the target fish. The options included ‘Circle’, ‘Square’, ‘Triangle’, ‘Star’ and ‘Did not see’.

The location of the target fish was associated with two features in the visual display such that the acquisition of one or both the features would lead to the development of a cue association in memory and arguably result in greater accuracy. The first feature was the lead fish. The target fish was always located within the school that was led by the right-facing lead fish. The second feature was the location of the target fish within the school. The target fish was always at the back of the school.

Response accuracy was measured for each trial. A learning score was calculated by dividing the total response accuracy of the last four trials by the total response accuracy of the 20 trials. The TST score was calculated by summing the total response accuracy of the 20 trials with the learning score. By adding the learning score to the total score, the TST score would account for any swift cue acquisition, which may occur in the initial few trials. Greater TST scores are presumed to be associated with higher levels of cue acquisition.

EXPERTise 2.0

Cue utilisation was assessed using the driving edition of EXPERTise 2.0 (Wiggins, Loveday & Auton, 2015). The five tasks include the Feature Identification Task, the Feature Recognition Task, the Feature Association Task, the Feature Discrimination Task, and the Feature Prioritisation Task. In the present study, the tasks were designed within the domain of operating a motor vehicle. The validity and reliability of EXPERTise 2.0, as a measure of cue utilisation, are described in Study 2 (Chapter 5).

Feature Identification Task (FIT)

Performance on the FIT reflects individual differences in the ability to extract critical features from a visual display. The FIT consisted of 21 trials. On each trial, participants were presented with an image of a left-hand traffic road scene viewed from inside a right-hand drive vehicle (see Figure 2). Participants were asked to identify and select the area of greatest concern, such as a road cyclist, school zone, and reversing truck, as quickly as possible.

On each trial, response latency was measured from the initial presentation of the stimulus to the participants' mouse click response. An FIT score was established by calculating the mean response latency across the 21 trials. A lower mean response latency is associated with higher cue utilisation (Loveday, Wiggins, Festa, & Schell, 2013).

Figure 2. An example of a road scene presented in one of the trials within the FIT

Feature Recognition Task (FRT)

Performance on the FRT reflects individual differences in the ability to recognise key features accurately under time pressure. The FRT consisted of 17 trials. On each trial, participants were presented with an image of a left-hand traffic road scene viewed from inside a right-hand drive vehicle (see Figure 3). The image was presented for a period of 1000 milliseconds, after which participants were asked to select their estimate of the speed limit from a list of four options: 50 or 60 km/h, 70 or 80 km/h, 90 or 100 km/h, or 110+ km/h.

On each trial, accuracy was measured. An FRT score was calculated by summing the number of correct trials. Greater total accuracy is associated with higher cue utilisation (Wiggins & O'Hare, 2003).

Figure 3. An example of a road scene presented in one trial within the FRT. The correct speed limit in this instance is 50 km/h.

Feature Association Task (FAT)

Performance on the FAT reflects individual differences in the precision and speed in differentiating associations between features. The FAT comprised two components with 17 trials in each component. In both components, participants were presented, for a period of 1000 milliseconds, with two terms that relate to the context of driving a motor vehicle, such as 'Merge' and 'Give way'. In the first component, the two terms were presented sequentially, while in the second component, the two terms were presented simultaneously.

Following the display of each pair of terms, participants were asked to rate the extent to which they considered that the terms were related. Responses were recorded on a Likert scale ranging from 1 (Extremely Unrelated) to 6 (Extremely Related).

Variance in ratings and mean response latency were measured across the 17 trials for both components. The FAT score for each component was calculated by dividing the variance in responses by the mean response latency. The FAT score was determined by calculating the mean FAT score of the two FAT components. A greater FAT score is associated with higher cue utilisation (Morrison, Wiggins, Bond, & Tyler, 2013).

Feature Discrimination Task (FDT)

Performance on the FDT reflects individual differences in the ability to discriminate relevant from less relevant features in the environment. The FDT consisted of one written scenario, where participants were asked to make a route-choice decision. The scenario included information, such as the traffic conditions in the area, the distance to a location, and the approximate time to the destination. Participants were encouraged to read the scenario for as long as necessary, after which, they were asked to select, from a list of four decision options, their initial response to the scenario.

Once participants had selected their preferred decision, they were asked to rate the extent to which each feature, such as time of day, traffic congestion, and local radio reports, was relevant to their decision. Responses were recorded on Likert scales ranging from 1 (Not important at all) to 10 (Extremely important). Variance in ratings across the 14 features was measured. Greater mean variance is thought to be associated with higher cue utilisation (Pauley, O'Hare, & Wiggins, 2009).

Feature Prioritisation Task (FPT)

Performance on the FPT reflects individual differences in the capacity to prioritise the acquisition of key feature-related information. The FPT consisted of one brief written scenario, where participants were asked to make a decision in relation to the mode of transportation that they would use to reach a specified destination. Before a decision could be made, participants were required to access additional information from a drop-down list of menu tabs that included information such as ‘Current time’, ‘Current weather’, and ‘Availability of bicycle parking’. When a tab was selected, information pertaining to the feature was displayed. Participants were permitted 120 seconds to access as many tabs as they considered necessary to formulate their decision. The menu tabs were listed randomly (see Figure 4).

The order in which the tabs were accessed, and the total number of tabs accessed were measured. A ratio metric was calculated by dividing the number of pairs of tabs accessed sequentially by the total number of pairs of tabs accessed. A smaller ratio is thought to be associated with higher cue utilisation (Wiggins & O'Hare, 1995).

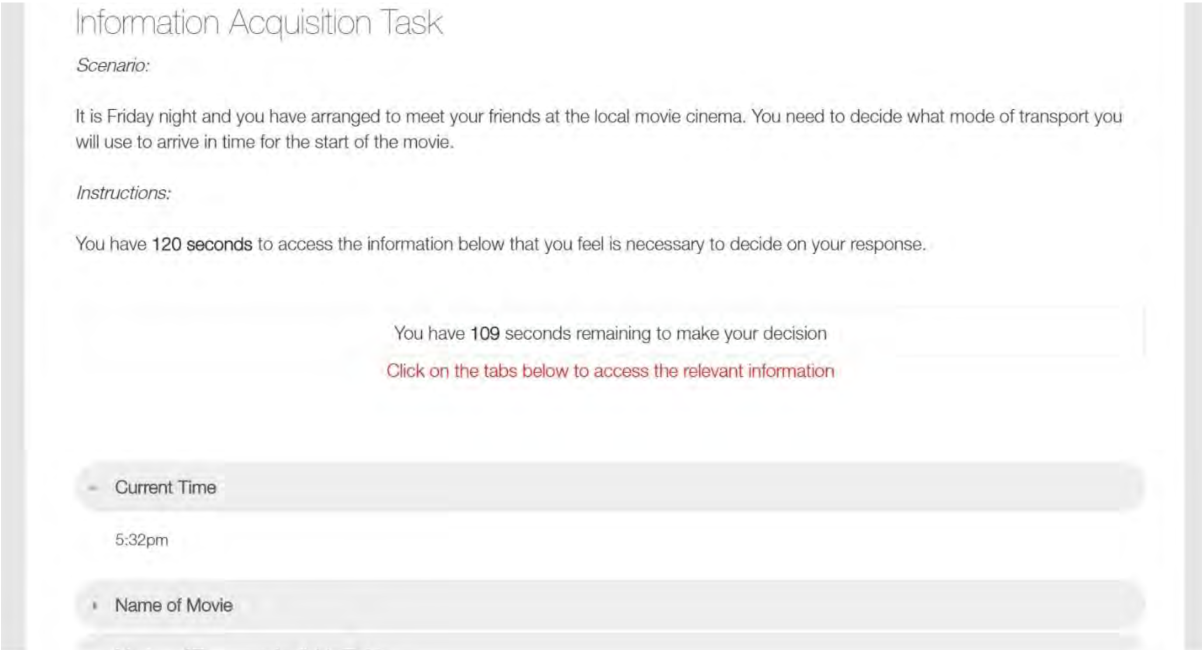

Figure 4. Example drop-down list from which participants can acquire information to make their decision

Driving Simulator

On the basis of previous assessments of driving performance (Campagne, Pebayle, & Muzet, 2004; Shechtman, Classen, Awadzi, & Mann, 2009), driving performance was evaluated based on the frequency of errors and violations committed during a 12-km drive in a driving simulator. The frequency of errors and violations represents a measure relevant to applied outcomes as it has been positively associated with accident involvement (Parker, Reason, Manstead, & Stradling, 1995).

The driving simulation task was conducted using a medium-fidelity driving simulator, which ran on the software STISIM (Version 8, Model 100). Participants controlled the driving simulator using a steering wheel, brake pedal, and accelerator. The view through the windscreen was displayed on three computer monitors, which supported a 135° view. Left and right side-view mirrors were displayed on the left and right monitors respectively, and the rear-view mirror was displayed on the centre monitor.

The experimental scenario comprised a 12-kilometre drive on a straight road in an urban setting. Features in the scenario included traffic lights, pedestrians, and parked cars, typical of a suburban driving environment. All the participants encountered the same features in the same conditions, such as red lights at fixed intersections, and pedestrians at fixed pedestrian crossings. If participants crashed into an object, they were able to continue driving. The total driving time was approximately 20 minutes if participants adhered to the road rules.

There were two measures of driving performance: the frequency of missed traffic signal errors and the frequency of collisions. Missed traffic signal errors included failing to stop where necessary at a red light, stop sign or a pedestrian crossing. Collisions included any crash resulting from a collision with an oncoming vehicle, stationary vehicle, leading vehicle or pedestrians. Any attempts to drive in a different direction other than the straight road would result in a crash, but would not be recorded as a collision.

Two measures of speed were recorded: average speed during the 12-kilometre drive and speed violations. Each time participants exceeded the speed limit, a speed violation was recorded.

Eye Tracker

SMI Eye Tracking Glasses 2.0 (SMI ETG) were utilised to collect visual behaviour metrics. The SMI ETG 2 records natural gaze behaviours based on 60 Hz binocular sampling rate, 60°

horizontal × 46° vertical range of vision, and a resolution of 1280 × 960 pixels at 24 frames per second. Visual input was calibrated with a three-point calibration.

Participants' visual behaviour metrics were recorded during the driving simulation task, and analysed with the BeGaze software. Visual behaviour metrics collected were fixation rates and saccades. Fixation rates comprised the mean number of fixations per minute, where fixation is measured with a 60Hz sampling rate.

Procedure

On arrival at the laboratory, participants were first provided with the information sheet to read and were given an opportunity to ask any questions. Participants were then directed to a room in which a driving simulator and desktop computer were located. The demographics questionnaire, the TST and the battery of cue utilisation tasks were administered online on the desktop computer. Participants completed the desktop-based tasks independently. On completion, participants were briefed on the driving simulator task, where they were instructed to drive on the main road as they would normally do in actual driving, and that they should stop driving should they experience any motion sickness.

During the driving simulation, participants first drove for approximately 1.3-km to familiarise themselves with the simulator. Following the practice trial, participants were fitted with the eye tracker and drove the 12-km experimental trial. Participants completed the driving simulator task without the researcher present in the room.

Results

Driving Cue Utilisation Clusters

Consistent with previous analysis of cue utilisation typologies (Brouwers, Wiggins, & Griffin, 2018; Crane et al., 2018; Watkinson, Bristow, Auton, McMahon, & Wiggins, 2018), a *K-means* cluster analysis was conducted on the standardised scores of the five cue utilisation tasks, including FIT, FRT, FAT, FDT and FPT, to establish two cue utilisation typologies in activities relating to driving a motor vehicle, including wayfinding, and hazard awareness.

Higher cue utilisation is typically associated with a faster response latency in the FIT, higher accuracy in the FRT, greater variance against response latency in the FAT, greater variance in the FDT, and a lower ratio in the FPT. By contrast, a shorter response latency on the FIT, lower accuracy on the FRT, lower variance against response latency on the FAT, lower variance on the FDT, and a higher ratio on the FPT normally characterises lower cue utilisation.

Following a *K-means* cluster analysis, the pattern of centroids indicated that, for the majority of tasks, the two groups recorded patterns of performance consistent with higher or lower levels of cue utilisation. However, the centroids for the Feature Association Task (FAT) were reversed so that, for the nominally higher cue utilisation group, the variance to response latency ratio was lower than the nominally lower cue utilisation group. Centroids for the tasks retained in the cluster analysis are listed in Table 1. Amongst the participants, 23 recorded performance that reflected lower cue utilisation, while 46 recorded performance that reflected higher cue utilisation.

To evaluate the differences in performance scores between the two clusters, independent sample *t*-tests were conducted on the scores of the four cue utilisation tasks. The lower and higher cue utilisation clusters significantly differed on the FIT and FPT scores, but only marginally differed on the FRT, FAT and FDT scores.

Table 1.
Mean centroids for higher and lower cue utilisation groups across each of the four cue utilisation tasks

	Cluster 1 (*n* = 23)	Cluster 2 (*n* = 46)			
	Lower cue utilisation	Higher cue utilisation	*t*	df	*p*
FIT	0.67	-0.33	3.70	29.26	.001
FRT	-0.27	0.14	-2.00	67	.113
FAT	0.17	-0.09	1.01	67	.313
FDT	-0.28	0.14	-1.85	67	.106
FPT	1.05	-0.52	9.41	67	<.001

To ensure that there were no systematic differences between the two cue utilisation groups on any other variables, independent samples *t*-tests were conducted on age, length of driving experience (in months), total hours in a week spent driving, and total hours in a week spent playing video games. There were no statistically significant differences between higher and lower cue utilisation groups in length of driving experience ($t(67) = 0.78$, $p = .436$, $d = 0.21$), total hours in a week spent driving ($t(67) = .11$, $p = .913$, $d = 0.03$), and total hours in a week spent playing video games ($t(67) = 0.09$, $p = .930$, $d = 0.03$). Given that the assumption of homogeneity of variance was not met for age, a Mann-Whitney Test was conducted. There was no statistically significant difference between higher and lower cue utilisation groups in age ($Z = 1.81$, $p = .070$).

Driving Cue Utilisation Clusters and Cue Acquisition

A binary logistic regression was conducted on driving cue utilisation clusters comprising two levels, higher cue utilisation and lower cue utilisation with TST scores as the predictor. The intercept-only model correctly classified 66.70% of the participants. Cue acquisition was not a significant predictor of cue utilisation cluster, χ^2 (1) = 0.61, $p = .437$. Cue acquisition explained 1.20% of the variability in cue utilisation clusters, Wald (1) = 0.60, $p = .439$. The odds ratio for cue acquisition was 1.04 (95% CI = 0.94 – 1.15). Therefore, cue acquisition was unlikely to be associated with cue utilisation clusters in activities relating to driving a motor vehicle.

Driving Cue Utilisation Components and Cue Acquisition

The relationship between cue acquisition and the five dimensions of driving cue utilisation were examined to establish whether cue acquisition was associated with one or more components of cue utilisation.

A multiple linear regression was conducted on TST scores with scores on the five dimensions of driving cue utilisation, FIT, FRT, FAT, FDT and FPT, as the predictor variables.

Cue utilisation accounted for 15.10% of the variance in TST scores, $F(5, 63) = 2.24$, MSe = 25.43, $p = .061$. However, only FDT ($\beta = -0.34$, $t(63) = 2.77$, $p = .007$) significantly predicted TST scores. Scores on the TST decreased by 0.34 for each increase in variance score on the FDT. Greater cue acquisition was associated with decreased variance in discriminating between features. The observed findings were inconsistent with the first hypothesis, which predicted that higher TST scores would be associated with lower FIT and higher FRT scores. Table 2 presents the zero-order correlations and regression coefficients of all five cue utilisation tests against TST scores.

Table 2.
Regression Coefficients of the Cue Utilisation Tests against TST (n = 69)

	Zero-order *r*				
Cue Acquisition	TST	*p*	*β*	*t*	*p*
FIT	-0.10	.260	-0.02	-0.18	.862
FRT	-0.08	.233	-0.05	-0.42	.679
FAT	0.12	.084	0.16	1.29	.202
FDT	-0.27	.011	-0.34	-2.77	.007
FPT	-0.16	.135	-0.23	-1.85	.069

Driving Cue Utilisation Clusters, Cue Acquisition and Driving Performance by Distance

Missed Signal Errors

To establish whether there were differences in driving performance across the duration of the driving simulation scenario, the scenario was divided into four blocks based on distance. A 4 x 2 mixed ANCOVA was undertaken to test the relationship between cue utilisation and the frequency of missed signal errors accounting for cue acquisition. The analysis incorporated two levels of driving cue utilisation (higher or lower) as a between-groups variable, the four blocks of driving performance as a within-groups variable, and the TST scores as a covariate.

The Greenhouse-Geisser correction was adopted due to a violation of Mauchly's Test of Sphericity ($p = .016$). There was no statistically significant main effect for blocks ($F(2.68, 176.77) = 1.62$, MSe = 2.78, $p = .191$, $\eta_p^2 = 0.02$), and no statistically significant interaction evident between blocks and driving cue utilisation (F(2.68, 176.77) = 1.17, p = .321, η_p^2 = 0.02), and blocks and TST scores (F(2.68, 176.77) = 0.24, p = .844, η_p^2 = 0.00).

A statistically significant effect for driving cue utilisation was evident after controlling for the effect of TST scores, F(1, 66) = 7.00, p = .010. Lower cue utilisation was associated with a greater frequency of missed signal errors (M = 2.85, SE = 0.42) compared to participants with higher cue utilisation (M = 1.50, SE = 0.29). TST scores were not significantly related to the frequency of missed signals, F(1, 66) = 0.35, p = .555 (see Figure 4).

Simple main effects analysis revealed statistically significant differences in missed signal errors committed between lower and higher cue utilisation individuals in Block 1 (p = .020), Block 3 (p = .049), and Block 4 (p = .012). Lower cue utilisation was associated with a greater frequency of missed signal errors in Blocks 1, 3 and 4 as compared to participants with higher cue utilisation (see Table 3).

Table 3.
Mean (and Standard Error) of Total Missed Signal Errors by Blocks and Cue Utilisation

	Block of Trials			
Cue Utilisation	Block 1	Block 2	Block 3	Block 4
Lower	3.13 (0.53)	2.15 (0.40)	2.38 (0.50)	3.75 (0.57)
Higher	1.59 (0.37)	1.34 (0.28)	1.14 (0.36)	1.93 (0.41)

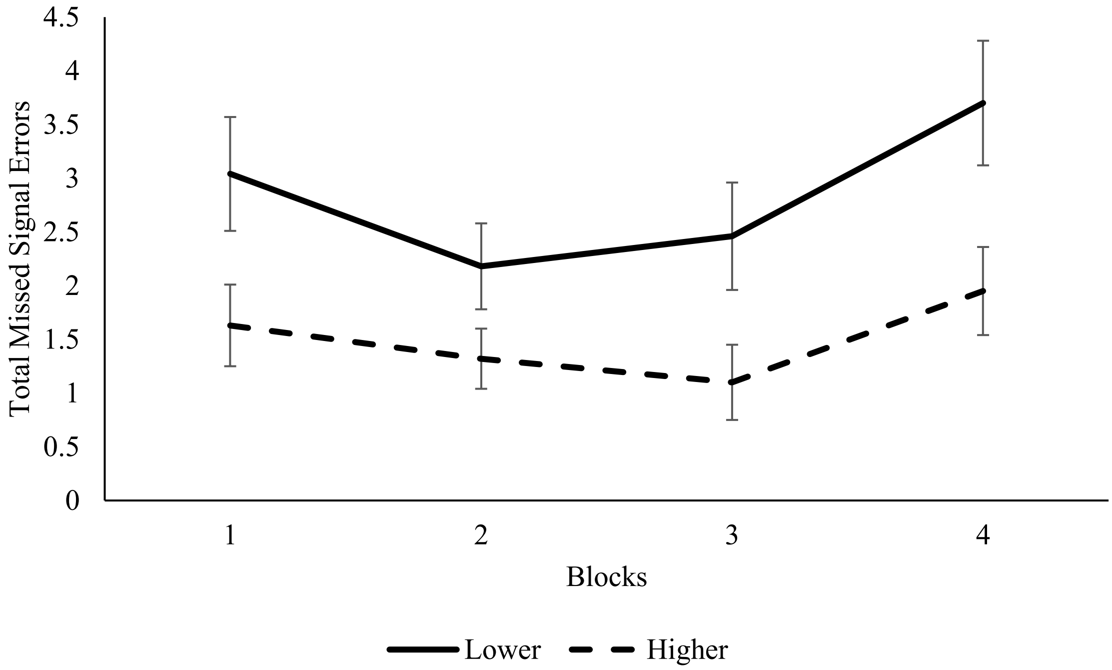

Figure 4. Total frequency of missed signal errors across the four time blocks by cue utilisation accounting for cue acquisition

In addition to evaluating the differences in driving performance between lower and higher cue utilisation, the mediating effect of cue utilisation on the relationship between cue acquisition and driving performance was evaluated. Four separate mediation analyses were conducted on the missed signal errors in each of the four blocks. Based on the results of Study 2 (Chapter 5), the analyses incorporated the FIT and FRT components of cue utilisation as mediators, and the TST scores as the independent variable.

In Block 1, there was no significant direct effect of TST scores (β = -0.01, $t(65)$ = -0.28, p = .777), nor indirect effect of FIT (ab = -0.01, 95% CI = -0.04 – 0.01) and FRT on missed signal errors (ab = 0.00, 95% CI = -0.01 – 0.03). Similarly, there was no significant direct effect of TST scores on missed signal errors in Blocks 2 (β = -0.04, $t(65)$ = -0.91, p = .366), Block 3 (β = -0.03, $t(65)$ = -0.53, p = .600), and Block 4 (β = -0.06, $t(65)$ = -0.85, p = .401). No significant indirect effect of FIT was observed in Block 2 (ab = -0.01, 95% CI = -0.03 – 0.01), Block 3 (ab = -0.01, 95% CI = -0.05 – 0.01), and Block 4 (ab = -0.00, 95% CI = -0.03 – 0.02), and no significant indirect effect of FRT was observed in Block 2 (ab = -0.00, 95% CI = -0.02

– 0.01), Block 3 (*ab* = 0.00, 95% CI = -0.02 – 0.04), and Block 4 (*ab* = 0.00, 95% CI = -0.02 – 0.03). The cue utilisation components FIT and FRT had no mediating effect on the relationship between cue acquisition and missed signal error.

Collision

A second 4 x 2 mixed ANCOVA was undertaken to test the relationship between cue utilisation and the frequency of collisions accounting for cue acquisition. The analysis incorporated two levels of driving cue utilisation (higher or lower) as a between-groups variable, the four blocks of driving performance as a within-groups variable, and the TST scores as a covariate.

The results revealed a statistically significant main effect for cue utilisation, $F(1, 66) = 4.02$, $p = .049$, $\eta_p^2 = 0.06$, where lower cue utilisation was associated with a greater frequency of collisions ($M = 0.31$, $SE = 0.05$) compared to participants with higher cue utilisation ($M = 0.18$, $SE = 0.04$). No statistically significant main effect was evident for blocks, $F(3, 198) = 1.38$, $p = .249$, $\eta_p^2 = 0.02$), and no statistically significant interaction was evident between blocks and driving cue utilisation ($F(3, 198) = 0.28$, $p = .838$, $\eta_p^2 = 0.00$), and blocks and TST scores ($F(3, 198) = 1.17$, $p = .323$, $\eta_p^2 = 0.02$).

To ensure that the observed differences in missed signal errors and collisions were not due to a speed accuracy trade-off, mean driving speed and speed violations of participants with higher and lower cue utilisation were compared across the 12-kilometre drive. The variable mean driving speed was transformed using a square-root transformation to achieve normality ($p = .062$).

Two independent sample *t*-tests were conducted separately on the transformed mean driving speed and speed violations, comparing higher and lower cue utilisation. The results failed to reveal any statistically significant differences in driving speed, $t(67) = 1.68$, $p = .098$, $d = 0.42$, and frequency of speed violations between participants with higher and lower cue utilisation, *t*(67) = 1.23, *p* = .224, *d* = 0.30.

Visual Behaviour Metrics, Cue Utilisation and Cue Acquisition

Frequency of Fixations

A 4 x 2 mixed ANCOVA was conducted to test the relationship between driving cue utilisation and the total frequency of fixations, accounting for cue acquisition. The analysis incorporated two levels of driving cue utilisation as a between-groups variable (higher or lower), the four blocks of total fixations as a within-groups variable, and the TST scores as a covariate score.

The Greenhouse-Geisser correction was adopted due to a violation of Mauchly's Test of Sphericity (p < .001). There was no statistically significant main effect for blocks (F(2.18, 97.96) = 2.64, MSe = 4874.37, p = .072, η_p^2 = 0.06), and no significant interaction between blocks and TST scores, F(2.18, 97.96) = 1.78, p = .155, η_p^2 = 0.04. However, a statistically significant interaction was evident between blocks and driving cue utilisation, F(2.18, 97.96) = 3.26, p = .039, η_p^2 η_p^2 = 0.07. A statistically significant, quadratic relationship was observed between blocks and cue utilisation (F(1, 45) = 5.76, MSe = 4575.11, p = .021, η_p^2 = 0.11). Lower cue utilisation individuals exhibited fewer total fixations at the beginning and end of the driving route compared to the middle of the driving route, whereas higher cue utilisation was associated with greater fixations at the beginning and end of the driving route compared to the middle of the driving route (see Figure 5).

A statistically significant effect of driving cue utilisation was evident after controlling for the effect of TST scores, F(1, 45) = 5.13, p = .028. Lower cue utilisation (M = 910.67, SE = 74.25) was associated with a greater frequency of total fixations compared to participants with higher cue utilisation (*M* = 713.80, *SE* = 45.25). TST scores were not significantly related to total fixations ($F(1, 45) = 1.81$, $p = .185$; see Figure 5).

Simple main effects analyses revealed statistically significant differences in frequency of fixations between lower and higher cue utilisation individuals in Block 1 ($p = .030$), Block 2 ($p = .017$), and Block 3 ($p = .025$). Lower cue utilisation was associated with a relatively greater frequency of fixations in Block 1, Block 2 and Block 3 (see Table 4).

Table 4.
Mean (and Standard Error) of Total Fixations by Blocks and Cue Utilisation

	Block of Trials			
Cue Utilisation	Block 1	Block 2	Block 3	Block 4
Lower	923.49 (72.66)	939.62 (77.65)	921.04 (81.94)	858.52 (69.62)
Higher	732.25 (44.28)	715.23 (47.32)	698.96 (49.94)	708.78 (42.43)

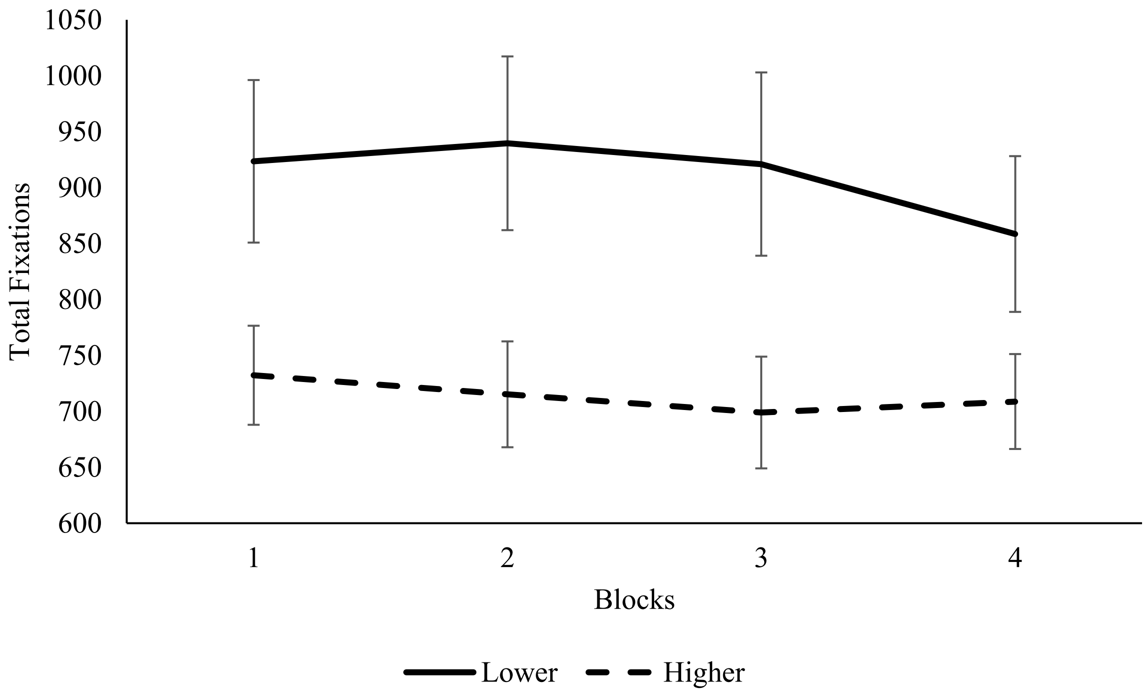

Figure 5. Total fixations across the four time blocks by cue utilisation accounting for cue acquisition

The mediating effect of cue utilisation on the relationships between cue acquisition and visual search behaviours were assessed. Four separate mediation analyses were conducted on the frequency of fixations in each of the four blocks. The analyses incorporated the FIT and

FRT components of cue utilisation as mediators, and the TST scores as the independent variable.

In Block 1, there was no significant direct effect of TST scores (β = 9.07, $t(46)$ = 1.27, p = .210), nor indirect effect of FIT (*ab* = -1.10, 95% CI = -4.99 – 3.11) and FRT on frequency of fixations (*ab* = 0.26, 95% CI = -4.20 – 4.26). Similarly, there was no significant direct effect of TST scores on frequency of fixations in Blocks 2 (β = 9.31, $t(46)$ = 1.19, p = .239), Block 3 (β = 13.49, $t(46)$ = 1.66, p = .103), and Block 4 (β = 9.08, $t(46)$ = 1.33, p = .190). No significant indirect effect of FIT was observed in Block 2 (*ab* = -0.96, 95% CI = -4.59 – 3.29), Block 3 (*ab* = -1.11, 95% CI = -5.05 – 3.13), and Block 4 (*ab* = -0.80, 95% CI = -3.86 – 2.51), and no significant indirect effect of FRT was observed in Block 2 (*ab* = 0.28, 95% CI = -4.56 – 4.78), Block 3 (*ab* = 0.30, 95% CI = -5.10 – 4.36), and Block 4 (*ab* = 0.24, 95% CI = -4.14 – 3.58). The cue utilisation components FIT and FRT had no mediating effect on the relationship between cue acquisition and frequency of fixations.

Frequency of Saccades

A 4 x 2 mixed ANCOVA incorporating two levels of driving cue utilisation (higher or lower) as a between-groups variable, the four blocks of time as a within-groups variable, and the TST scores as a covariate score was conducted to test the relationship with the total frequency of saccades.

The Greenhouse-Geisser correction was adopted due to a violation of Mauchly's Test of Sphericity (p < .001). Although a statistically significant main effect was evident for blocks ($F(2.01, 90.29)$ = 4.74, MSe = 3978.75, p = .011, η_p^2 = 0.10), no statistically significant interaction was evident between blocks and driving cue utilisation ($F(2.01, 90.29)$ = 1.85, p = .162, η_p^2 = 0.04), and blocks and TST scores ($F(2.01, 90.29)$ = 2.07, p = .127, η_p^2 = 0.04).

Pairwise comparisons revealed a statistically significant difference in the frequency of saccades between Block 1 (M = 685.81, SE = 40.46) and Block 3 (M = 652.79, SE = 43.72; p

= .017) and Block 4 (*M* = 632.65, *SE* = 37.90; *p* = .003), and Block 2 (*M* = 673.26, *SE* = 42.37) and Block 4 (*p* = .049). Therefore, the total frequency of saccades decreased over time (see Figure 6).

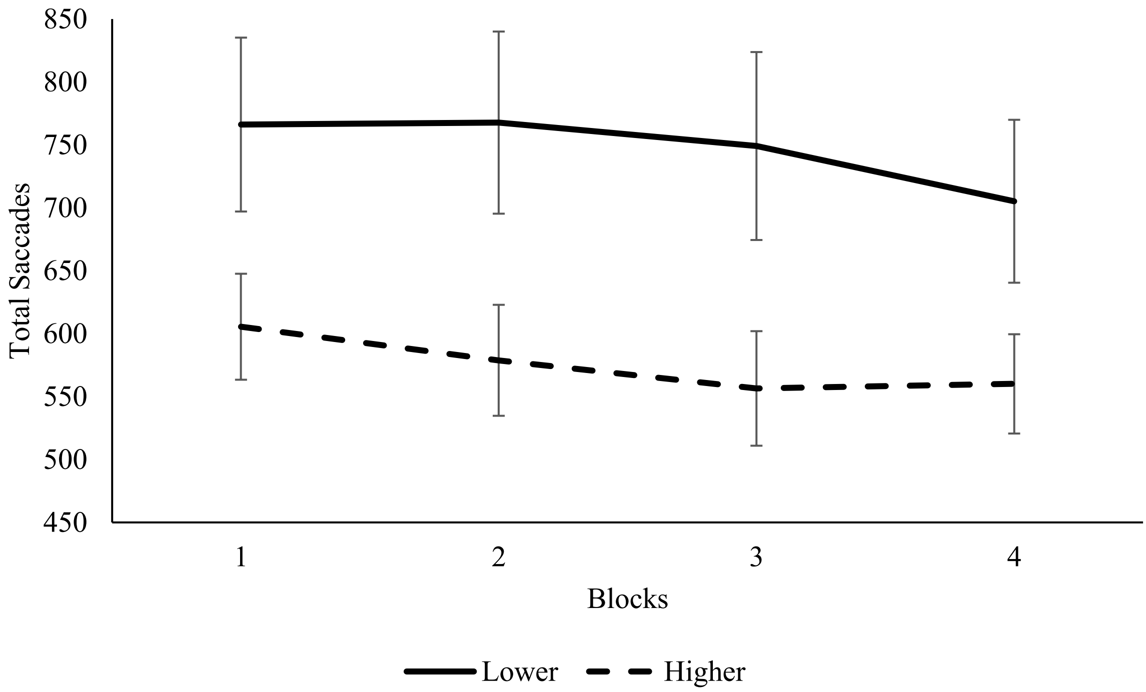

Figure 6. Total saccades across the four time blocks by cue utilisation accounting for cue acquisition

A statistically significant effect for driving cue utilisation was evident after controlling for the effect of TST performance, $F(1, 45) = 4.49$, $p = .040$. Lower cue utilisation (*M* = 747.03, *SE* = 69.21) was associated with a relatively greater frequency of saccades in comparison to higher cue utilisation (*M* = 575.22, *SE* = 42.18). TST performance was not significantly related to total saccades ($F(1, 45) = 2.43$, $p = .126$).

Simple main effects analysis revealed statistically significant differences in the total frequency of saccades between lower and higher cue utilisation individuals in Block 2 (p = .031) and Block 3 (p = .033). Lower cue utilisation individuals recorded a greater total frequency of saccades in the second and third blocks compared to individuals with higher cue utilisation (see Table 5). Differences in the total frequency of saccades between lower and higher cue utilisation in Block 1 (p = .053) and Block 4 (p = .062) were not statistically significant.

Table 5.
Mean (and Standard Error) of Total Frequency of Saccades by Blocks and Cue Utilisation

	Block of Trials			
Cue Utilisation	Block 1	Block 2	Block 3	Block 4
Lower	766.14 (69.09)	767.70 (72.36)	749.09 (74.67)	705.20 (64.73)
Higher	605.49 (42.11)	578.83 (44.10)	556.48 (45.51)	560.10 (39.45)

Four separate mediation analyses were conducted on the frequency of saccades in each of the four blocks. The analyses incorporated the FIT and FRT components of cue utilisation as mediators, and the TST scores as the independent variable.

In Block 1, there was no significant direct effect of TST scores (β = 8.77, t(46) = 1.27, p = .210), nor indirect effect of FIT (ab = -0.52, 95% CI = -3.31 – 2.54) and FRT on frequency of saccades (ab = 0.23, 95% CI = -3.55 – 3.91). Similarly, there was no significant direct effect of TST scores on frequency of saccades in Blocks 2 (β = 10.06, t(46) = 1.37, p = .177), Block 3 (β = 13.35, t(46) = 1.77, p = .084), and Block 4 (β = 9.65, t(46) = 1.48, p = .145). No significant indirect effect of FIT was observed in Block 2 (ab = -0.36, 95% CI = -2.85 – 2.71), Block 3 (ab = -0.45, 95% CI = -3.42 – 2.50), and Block 4 (ab = -0.16, 95% CI = -2.26 – 2.56), and no significant indirect effect of FRT was observed in Block 2 (ab = 0.24, 95% CI = -3.70 – 3.53), Block 3 (ab = 0.24, 95% CI = -3.70 – 3.70), and Block 4 (ab = 0.20, 95% CI = -3.35 – 2.83). Consequently, the cue utilisation components FIT and FRT had no mediating effect on the relationship between cue acquisition and frequency of saccades.

Discussion

The aim of the present study was to evaluate the influence of cue acquisition and cue utilisation on driving performance and visual search behaviours. It was hypothesised that the relationship between cue utilisation in driving and performance on a simulated driving task would be moderated by individual differences in cue acquisition. Further, it was also hypothesised the relationship between cue utilisation and the frequency of fixations and saccades during the simulated driving task would be moderated by cue acquisition.

Consistent with the proposed hypothesis, participants with higher cue utilisation committed fewer missed signal errors and collisions, and recorded fewer fixations and saccades compared to participants with lower cue utilisation. However, scores on the TST did not significantly contribute to the associations between cue utilisation, and driving performance and visual search behaviours. Further, cue acquisition was not significantly associated with scores on four of the five cue utilisation tasks, and was negatively associated with FDT scores. Greater variance on the FDT component of cue utilisation was associated with lower propensity for cue acquisition.

Given that self-reported driving experience was controlled, it might be argued that any differences in driving performance and visual search behaviours between the two cue utilisation clusters were due to variations in cue utilisation. This observation is consistent with previous findings, and suggests that, while domain-specific experience is necessary for the acquisition of feature-event/object associations, there are individual variations in the application of cue-based associations during situation assessment (Loveday, Wiggins, Searle, Festa, & Schell, 2013; Wiggins, Brouwers, Davies, & Loveday, 2014). As a consequence, variations in performance outcomes amongst individuals of similar experience should be influenced by individual variations in domain-specific cue utilisation.

Over the short driving scenario, all of the participants showed a similar deterioration in visual search behaviours, reflected in reduced fixations and saccades across time. The observed overall decline in visual search performance over time is consistent with the impact of sustained attention based on a Resource Theory interpretation (Helton & Warm, 2008). According to Resource Theory, extended periods of sustained attention increases cognitive demand during information processing, thereby depleting resources to a point where there are insufficient resources to meet the demands of the task. In the present study, this degradation appears more prominent amongst drivers with lower cue utilisation, who consistently committed more driving errors than drivers with higher cue utilisation. Based on this observation, it might be argued that higher cue utilisation is associated with a reduction in cognitive demands during sustained attention tasks such as driving, leading to greater residual resources per unit exposure and ultimately, a lower rate of error.

Further support for the proposition that higher cue utilisation is associated with the consumption of fewer resources per unit time is evident from the differences in visual search behaviours. Fewer fixations and saccades were observed amongst drivers with higher cue utilisation, possibly reflecting more efficient search behaviours (Smuc et al., 2010). Fixating on fewer, relevant features, arguably, decreases the amount of information to be managed, and reduces the number of saccades, where no visual information is acquired.

In routine tasks, where features in the environment are relatively familiar to the perceiver, fewer fixations presumably reflect the acquisition of visual information based on cues that match existing feature-event associations in memory. If features in the environment are relatively unfamiliar, cue utilisation would be ineffective in guiding fixations. Instead, the perceiver would record a greater frequency of fixations searching for visual information that corresponds to existing feature-event/object associations in memory.

This visual search strategy, characterised by greater fixations and saccades, is typically adopted by less experienced individuals during the process of situation assessment (Arthur et al., 2016; Krupinski et al., 2013). A greater frequency of fixations is associated with the need to increase recognition capacity (Duchowski, 2007), while greater saccadic eye movements is presumed to reflect a more comprehensive examination of a visual array (Goldberg & Kotval, 1999). In the context of the present study, a greater frequency of fixations and saccades might reflect the absence of or inadequate relevant feature-event/object associations in memory and the search for familiar information.

While driving performance and visual search behaviours were associated with cue utilisation, cue acquisition was not related to driving errors, fixations or saccades. This suggests that variations in driving performance and visual search behaviours are related to task-related cue utilisation, but that this effect is not due to individual differences in generalised cue acquisition as measured by the TST.

With increased experience in a domain, the influence of cue acquisition on performance outcomes is likely to become less evident than the influence of task-related cue utilisation. According to Ackerman (1988), the acquisition of learned behaviours occurs over three consecutive stages, where the acquisition and strengthening of associations occur in the first two stages, and the automated execution of the learned behaviours occurs in the final stage. The first two stages reflect the functions of cue acquisition at the early stages of learning, and the final stage reflects the application of cue utilisation through increases in practice and experience.

The outcomes of Study 1 (Chapter 4) and Study 2 (Chapter 5) of this programme of research suggest that individual differences in cue acquisition are evident during the initial stages of learning. However, following the acquisition of domain-specific, feature-event/object associations, the influence of cue acquisition diminishes as cue utilisation becomes more informative as a predictor of task performance. Given that the feature-event/object associations required to support cue utilisation have been acquired, cue acquisition is unlikely to be predictive of domain-specific task performance.

Cue acquisition is most likely to be evident in the context of novel tasks where previous associations are unlikely to be applicable. In the present study, the task was relatively familiar so that participants were drawing on cue-based associations that were already resident in memory. Therefore, the nature of the task may have obviated the acquisition of cue-based associations. To further assess this proposition, there is a need to examine performance during both familiar and novel tasks within the same context.

There is also a need to consider the relationship between cue acquisition and cue utilisation from a longitudinal perspective to assess the changes in the effects of cue acquisition and cue utilisation over the three stages of skill acquisition. For cue acquisition to function as a predictive measure of later performance, it is necessary to establish whether the rate of cue acquisition at the initial stages of learning would be reflected in the later stages of task performance when experiencing novel conditions. This measure is especially useful in high-risk domains, such as aviation and firefighting, where responses to novel conditions must be made under time pressure.

In assessing both cue acquisition and cue utilisation from a longitudinal perspective, it will also be possible to establish the dependencies between the variables. For example, in Study 2 (Chapter 5) of this programme of research, TST scores were associated with the FIT and the FRT in the context of general aviation. However, in the present study, the association was with the FDT. This suggests that either the TST lacks reliability or that there are different levels of dependency between cue acquisition and cue utilisation across domains and/or extent of experience.

In Study 2 (Chapter 5), the participants were relatively inexperienced. However, in Study 3, the participants were more familiar with the environment and this may have changed the dependency between measures of cue acquisition and cue utilisation. Future research might match levels of experience across different domains to better test the reliability of the relationship and establish whether the relationships were spurious or reflected dependencies in reality.

The nature of the experimental stimuli is also likely to have an impact on the relationship between cue utilisation and cue acquisition. For example, the 12-kilometre drive in the present study was an artificial reproduction of real-world driving conditions, where drivers must adapt to changing decision situations, such as intersections, roadworks, and/or reduced visibility. To build a more complete understanding of the relationship between cue utilisation and cue acquisition, more complex situations should be developed as the basis for evaluation.

On the basis of the present results, training schemes that are designed to support and improve new drivers' performance to reduce the frequency of accidents should consider the role of cue utilisation. Cue-based training, which emphasises the development of associations between features and events or objects, is an approach that could enhance cue utilisation (McCammon & Hägeli, 2007; Wiggins & O'Hare, 2003). By teaching operators cognitive strategies that highlight critical cues, operators are better able to maintain appropriate vigilance and attention when monitoring system state on sustained attentions tasks (Potter, Blickensderfer, & Bouquet, 2014).

In conclusion, this study was designed to evaluate the influence of cue acquisition and cue utilisation on driving performance and visual search behaviours amongst drivers with similar levels of driving experience. Higher cue utilisation was associated with improved driving performance and more efficient visual search behaviours. Cue acquisition, however, was not associated with cue utilisation and driving performance, and it was proposed that the influence of cue acquisition may not be as apparent following the initial acquisition of feature-event associations.

References

Ackerman, P. L. (1988). Determinants of individual-differences during skill acquisition - cognitive-abilities and information-processing. *Journal of Experimental Psychology-General, 117*(3), 288-318. doi:Doi 10.1037/0096-3445.117.3.288

Alison, L., Doran, B., Long, M. L., Power, N., & Humphrey, A. (2013). The effects of subjective time pressure and individual differences on hypotheses generation and action prioritization in police investigations. *Journal of Experimental Psychology: Applied, 19*(1), 83. doi:10.1037/a0032148

Arthur, P., Khuu, S., & Blom, D. (2016). Music sight-reading expertise, visually disrupted score and eye movements. *Journal of Eye Movement Research, 9*(7). doi:ARTN 110.16910/jemr.9.7.1

Brouwers, S., Wiggins, M., & Griffin, B. (2018). Operators who readily acquire patterns and cues, risk being miscued in routinized settings. *Journal of Experimental Psychology: Applied*. doi:10.1037/xap0000151

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics, 60*(11), 1-55. doi:10.1080/00140139.2017.1330494

Brouwers, S., Wiggins, M. W., Helton, W. S., O'Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. *Frontiers in Psychology, 7*, 435. doi:10.3389/fpsyg.2016.00435

Campagne, A., Pebayle, T., & Muzet, A. (2004). Correlation between driving errors and vigilance level: influence of the driver's age. *Physiology & Behavior, 80*(4), 515-524. doi:10.1016/j.physbeh.2003.10.004

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception, 27*(8), 951-964. doi:10.1068/p270951

Crane, M. F., Brouwers, S., Wiggins, M. W., Loveday, T., Forrest, K., Tan, S. G. M., & Cyna, A. M. (2018). “Experience isn't everything”: How emotion affects the relationship between experience and cue utilization. *Human Factors, 60*(5), 685-698. doi:10.1177/0018720818765800

Crundall, D., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics, 41*(4), 448-458. doi:Doi 10.1080/001401398186937

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception, 28*(9), 1075-1087. doi:10.1068/p281075

de Craen, S., Twisk, D., Hagenzieker, M. P., Elffers, H., & Brookhuis, K. A. (2008). The development of a method to measure speed adaptation to traffic complexity: Identifying novice, unsafe, and overconfident drivers. *Accident Analysis & Prevention, 40*(4), 1524-1530. doi:10.1016/j.aap.2008.03.018

Duchowski, A. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455-470. doi:10.3758/bf03195475

Duchowski, A. (2007). *Eye Tracking Methodology, Theory and Practice*. Springer International Publishing. doi:10.1007/978-1-84628-609-4

Dzeng, R.-J., Lin, C.-T., & Fang, Y.-C. (2016). Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification. *Safety Science, 82*, 56-67. doi:10.1016/j.ssci.2015.08.008

Engström, I., Gregersen, N. P., Hernetkoski, K., Keskinen, E., & Nyberg, A. (2003). *Young novice drivers, driver education and training: Literature review*. Linköping, Sweden: Swedish National Road and Transport Research Institute.

Fracker, M. (1988). A theory of situation assessment: Implications for measuring situation awareness. *Proceedings of the Human Factors Society - 32nd Annual Meeting, 32*(2), 102-106. doi:10.1177/154193128803200222

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention, 37*(3), 461-472. doi:10.1016/j.aap.2004.11.003

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451-482. doi:10.1146/annurev-psych-120709-145346

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631-645. doi:10.1016/s0169-8141(98)00068-7

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychologica, 129*(1), 18-25. doi:10.1016/j.actpsy.2008.04.002

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. doi:10.1177/154193120104500411

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Klein, G. (2008). Naturalistic decision making. *Human factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making, 4*(3), 186-209. doi:10.1518/155534310x12844000801203

Krupinski, E. A., Graham, A. R., & Weinstein, R. S. (2013). Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Human Pathology, 44*(3), 357-364. doi:10.1016/j.humpath.2012.05.024

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331-352. doi:10.1002/bdm.381

Loveday, T., Wiggins, M. W., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In L. C. P., S. J., & F. A. (Eds.), *Pattern Recognition - Applications and Methods* (Vol. 204, pp. 1-11): Springer, Berlin, Heidelberg.

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(1), 125-137. doi:10.1177/0018720812448475

Mann, D. T. Y., Williams, M. A., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: a meta-analysis. *Journal of Sport and Exercise Psychology, 29*(4), 457-478. doi:10.1123/jsep.29.4.457

McCammon, I., & Hägeli, P. (2007). An evaluation of rule-based decision tools for travel in avalanche terrain. *Cold Regions Science and Technology, 47*(1-2), 193-206. doi:10.1016/j.coldregions.2006.08.007

McCormack, C., Wiggins, M. W., Loveday, T., & Festa, M. (2014). Expert and competent non-expert visual cues during simulated diagnosis in intensive care. *Frontiers in Psychology, 5*, 949. doi:10.3389/fpsyg.2014.00949

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making, 7*(2), 211-226. doi:10.1177/1555343412459192

Parker, D., Reason, J. T., Manstead, A. S. R., & Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics, 38*(5), 1036-1048. doi:10.1080/00140139508925170

Pauley, K., O'Hare, D., & Wiggins, M. (2009). Measuring expertise in weather-related aeronautical risk perception: The validity of the cochran–weiss–shanteau (cws) index. *The International Journal of Aviation Psychology, 19*(3), 201-216. doi:10.1080/10508410902979993

Potter, B. A., Blickensderfer, E. L., & Boquet, A. J. (2014). Training monitoring skills in helicopter pilots. *Aviation, Space, and Environmental Medicine, 85*(5), 543-549. doi: 10.3357/ASEM.3771.2014

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi: 10.1037/0033-2909.124.3.372

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 492. doi:10.1037/0278-7393.6.5.492

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(5), 888-896.

Renshaw, P. F., & Wiggins, M. W. (2017). The predictive utility of cue utilization and spatial aptitude in small visual line-of-sight rotary-wing remotely piloted aircraft operations. *International Journal of Industrial Ergonomics, 61,* 47-61. doi: 0.1016/j.ergon.2017.05.014

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica, 127*(2), 258-276. doi:10.1016/j.actpsy.2007.05.004

Shechtman, O., Classen, S., Awadzi, K., & Mann, W. (2009). Comparison of driving errors between on-the-road and simulated driving assessment: A validation study. *Traffic Injury Prevention, 10*(4), 379-385. doi:10.1080/15389580902894989

Smuc, M., Mayr, E., & Windhager, F. (2010). The game lies in the eye of the beholder: The influence of expertise on watching soccer. *Cognition in Flux*, 1631-1636.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46*(12), 1857-1862. doi:10.1016/j.visres.2005.12.005

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2015). Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical Endoscopy, 29*(2), 405-413. doi:10.1007/s00464-014-3683-7

Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology*, 1-7. doi:10.1080/14992027.2018.1476782

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human factors, 44*(2), 189-197. doi:10.1518/0018720024497871

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014). Trait-based cue utilization and initial skill acquisition: Implications for models of the progression to expertise. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00541

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology-Applied, 1*(4), 305-320. doi: 10.1037/1076-898x.1.4.305

Wiggins, M. W., & O'Hare, D. (2003). Expert and novice pilot perceptions of static in-flight images of weather. *The International Journal of Aviation Psychology, 13*(2), 173-187. doi:10.1207/S15327108IJAP1302_05

Bridging Section

Study 3 (Chapter 6) was designed to investigate the influence of individual differences in cue acquisition on the relationship between cue utilisation and task performance in the context of driving. While higher cue utilisation was associated with reduced error rate in a driving task and more efficient visual search behaviour, individual differences in cue acquisition appeared to contribute little in explaining the relationship between cue utilisation and task performance in the context of motor vehicle driving.

The results that emerged from Study 1 (Chapter 4), however, demonstrated an association between cue acquisition and task performance in a novel domain, suggesting that individual differences in cue acquisition may only influence task performance in some instances. In combination, the findings of Studies 1 (Chapter 4) and 3 (Chapter 6) suggest that the sensitivity of cue acquisition in predicting task performance is likely to change with experience. Therefore, Study 4 (Chapter 7) was intended to investigate the extent to which individual differences in cue acquisition predict task performance following initial exposure.

The author of the present thesis contributed approximately 75% to the preparation of this paper in data collection, data analysis, and the writing of the paper.

CHAPTER SEVEN

Absence of Direct Associations between Cue Acquisition and Performance in General Aviation

Abstract

Effective situation assessment is important for the construction of accurate assessments of the presenting situation, and therefore, the selection of appropriate decision responses. Higher cue utilisation has previously been demonstrated to be a better predictor of improved situation assessment than experience, suggesting that there are individual differences in the acquisition of cues that support cue utilisation. A measure of cue acquisition was completed by 42 pilot trainees, and assessed against actual flight performance data, accounting for flight experience. Cue acquisition was not associated with flight control measures and instructor ratings, accounting for flight experience. While the findings suggest that cue acquisition is not associated with performance, there were some methodological factors that may have impacted the results.

Keywords: cue acquisition, flight performance, cues, general aviation

Absence of Direct Associations between Cue Acquisition and Performance in General Aviation

Situation assessment is a cognitive process that involves the extraction and consolidation of relevant information to form a working understanding of a situation or problem (Horrey & Wickens, 2001). It informs the selection and/or construction of action plans, and thereby ensures sustained, higher levels of performance, particularly in dynamic, high consequence environments, such as aviation piloting and motor vehicle driving (Klein, Calderwood, & Clinton-Cirocco, 2010).

More effective situation assessment has been associated with higher levels of domain-specific performance amongst expert operators (Goh & Wiegmann, 2001; Mueller & Trick, 2012; Wiegmann, Goh, & O'Hare, 2002). For instance, Wiegmann et al. (2002) observed greater accuracy in the assessments of visibility and cloud ceiling amongst pilots who elected to divert from deteriorating weather conditions compared to pilots who continued to fly into deteriorating weather conditions. Similarly, novice drivers who recorded greater rates of collision, tended to be less accurate in identifying hazards than expert drivers (Mueller & Trick, 2012).

According to the Dual-Processing account of cognition (Evans, 2003, 2008), the slower, less effective situation assessment of novice operators functions using System 2 processing, where information in the environment is analysed slowly and consciously. In contrast, situation assessment amongst expert operators is usually a process that is rapid, nonconscious, and accurate, consistent with a System 1 processing. The rapid assessments, generated by System 1 processing are dependent upon pre-existing schemas in memory.

Schemas are templates comprising representations of features and events or objects that are drawn from experiences in previous, similar situations (Ghosh & Gilboa, 2014). Within schemas, representations of features and events or objects exist in associations that are formed through related experiences. Repeated and consistent exposure to the co-occurrence of features and events or objects across varying situations in domain-specific environments results in the acquisition of feature-event/object associations in memory (Kim, Seitz, Feenstra, & Shams, 2009).

Where feature-event/object associations in memory support the production of rapid and accurate assessments by triggering both the perception and interpretation of features from the environment, they are known as cues (Klein, 2008). During situation assessment, operators identify matches between perceived features in the environment and representations of features in memory. A match will prompt an assessment of the situation supplemented by associated information, including expectancies, possible goals, and previously successful responses.

The importance of the application of cues is evident in the context of an urban fireground setting, where critical information needs to be discerned quickly from an array of dynamic, and often obscured features. For example, smoke and fire may not be immediately evident to a firefighter. However, heat radiating from an unvented area (feature) constitutes a critical precursor to a situation where the re-introduction of oxygen by opening a door or breaking a window can cause a rapid and often uncontrolled increase in combustion (event; Okoli, Watt, & Weller, 2017).

In the ideal situation, the radiation of heat is matched to existing feature-event/object associations in memory, prompting the retrieval of related events or objects, including that the fire could advance quickly with a sudden increase in the availability of oxygen. Where the association between radiated heat and rapid combustion is unavailable in memory in the form of a cue, firefighters may be unable to recognise the significance of the feature, and therefore, may respond inappropriately.

The acquisition of relevant cues that support effective situation assessment is, in part, influenced by individual differences in domain-specific experiences (Ericsson & Charness, 1994). Arguably, experience constitutes a necessary condition for operators to encounter relevant features and events or objects. However, associations between features and events or objects exist within environments that are characterised by various other irrelevant information, and therefore, operators must identify and extract meaningful patterns of associations during cue acquisition.

The ability to extract patterns of associations from noisy environments is characteristic of human cognition and is necessary to ensure optimal performance (Reber, Kassin, Lewis, & Cantor, 1980). However, exposure to similar experiences does not guarantee that different operators will extract similar patterns of associations. Instead, there is evidence to suggest that there are individual differences in the rate at which cues are acquired (Crane et al., 2018; Todd & Thomas, 2012).

For example, Loveday, Wiggins, Harris, O'Hare, and Smith (2012) and Crane et al. (2018) have demonstrated that performance on diagnostic tasks is better predicted by differences in cue utilisation than it is by differences in exposure to the domain. Individual differences in cue utilisation is reflected in behaviours demonstrating the extraction, integration and utilisation of cues during situation assessment (Wiggins, 2012). Given that cue utilisation is a better predictor of performance outcome than experience, it might be inferred that there are individual differences in cues that support situation assessment. Importantly, this dissociation between exposure and performance, consistent with previous research (Hambrick et al., 2014; Meinz & Hambrick, 2009), suggests that there are individual differences in the rate at which cues are acquired in memory.

In domains where the costs of training are relatively high, assessments of the capacity to acquire cues have the potential to differentiate applicants and facilitate the selection of candidates with a greater propensity for cue acquisition. In turn, this reduces the costs associated with training and/or retraining, and potentially yields improvements in operator performance.

The Timed-Search Task (TST) is an assessment tool that is designed to evaluate individual differences in the acquisition of cues. It is a context-independent instrument, the construct validity of which was evaluated against cue utilisation within the domain of general aviation in Study 2 (Chapter 5) of this programme of research. Accounting for domain-specific experience, TST scores were associated with two components of the general aviation measure of cue utilisation: The Feature Identification Task (FIT) and Feature Recognition Task (FRT). Given that performance on the TST is not dependent on previous experience, the findings suggest that TST scores reflect individual differences in the acquisition processes of identifying and recognising relevant cues in the environment.

Extending the findings of Study 2 (Chapter 5), the present study was designed to assess the relationship between cue acquisition and performance in the context of general aviation flight training, accounting for domain-specific experience. Actual aircraft operations and instructors' ratings were employed as measures of flight performance that represent more objective and more subjective measures respectively. On the basis of the outcomes of Study 2, it was predicted that, accounting for hours of flight experience, greater scores on the TST would be associated with improved flight control performance and higher subjective ratings.

Methodology

Participants

Forty-two pilot trainees enrolled in the Bachelor of Aviation at a tertiary institution in New Zealand were recruited through convenience sampling. Participation in the study was voluntary with a chance to win an iPad mini. Approval to recruit participants for the study was granted by the Human Research Ethics Committee of the tertiary institution (see Appendix B).

The participants consisted of 40 males and two females, and ranged in age from 17 to 27 years (*M* = 20.29, *SE* = 0.35). Flight experience at the tertiary institution ranged from 2.70 hours to 745.10 hours (*M* = 98.37, *SE* = 20.79). Flight experience prior to starting at the tertiary institution ranged from zero to 40 hours (*M* = 6.41, *SE* = 1.64).

To be included in the study, participants were required to be enrolled in, or recently graduated from, the Bachelor of Aviation at the tertiary institution, and having had at least two hours of flight experience at the institution.

Materials

Timed-Search Task

The Timed-Search Task (TST) is an instrument developed and validated in Studies 1 (Chapter 4) and 2 (Chapter 5) of this programme of research. The stimuli presented in this task were fictitious, and therefore, responses were not expected to be influenced by experience (see Figure 1). The version of the TST administered in the present study was the same as the version utilised in the previous three studies (Chapters 4, 5, and 6).

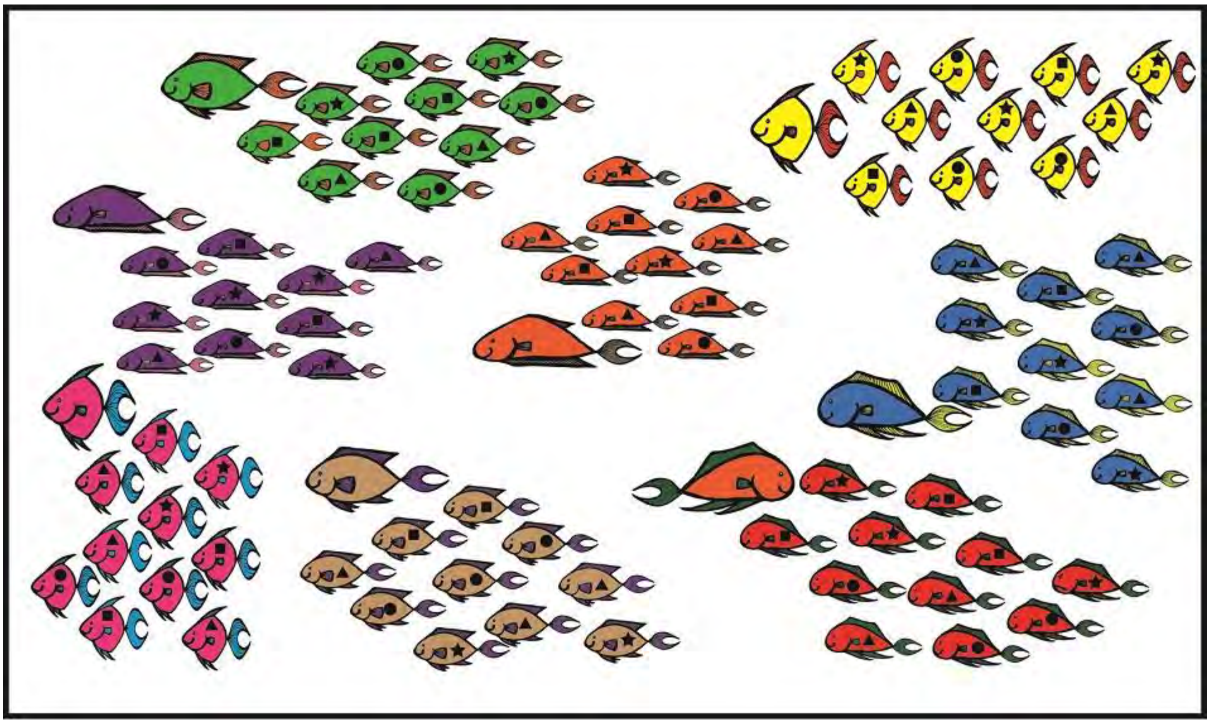

Figure 1. Example display observed by participants on a single TST trial

Flight Performance

Flight performance was operationalised as: (1) instructor ratings; and (2) flight behaviour during a selected flight lesson.

Instructor Ratings of Selected Flight Lesson

Lesson 11 of the flight training syllabus had been selected to establish a performance outcome for both instructor ratings and actual flight operations. This lesson is one of the standard lessons undertaken prior to pilot trainees undertaking their first solo flight. During the flight, pilot trainees are required to perform specified manoeuvres and procedures.

The instructor ratings are completed based on a rating list, consisting of 19 – 21 items. Differences in the number of items in the rating list are due to updates and improvements made to the rating list across student cohorts. The items relate to pre-flight, flight and post-flight activities. Example items include 'Weather briefing', 'Steep turns', and 'Flapless landing'.

The items are rated on a scale of two, three, four or five points. For each scale, the minimum score is one, with performance deemed to be below the minimum score rated as negative one or two. Ratings for each item were calculated as a proportion of the total points on the given scale. For instance, if a participant was awarded a rating of one on an item with a five-point scale, the rating score for the given item was computed as 0.20. Similarly, if a participant was given a rating of negative one on an item with a five-point scale, the rating score for the given item was computed as –0.20. If a value for an item could not be completed due to extenuating circumstances, the item was removed from the overall calculation.

An overall instructor rating was computed for each pilot trainee by calculating the mean score of all items. Greater scores reflected a greater performance rating by the instructor, while lower scores reflected a poorer rating.

Actual Flight Operations during Selected Flight Lesson

The flight time for Lesson 11 was between 0.80 hours and 2.20 hours (M = 1.63, SE = 0.04). For each participant, two, two-minute segments from the total flight period were extracted. The first segment was selected by identifying a period of stable, level flight. Level flight was operationalised as a period during which the aircraft was maintaining straight and level flight. For every participant, the level flight segment was identified as the first period of straight and level flight following take-off. The second two-minute segment occurred during the final approach phase immediately prior to landing. This segment was referred to as the landing segment.

Once the two segments of the flight were identified, the angles of pitch and roll were recorded for each segment. Pitch constitutes a measure of movement about the aircraft's transverse axis, while roll is a measure of movement about the aircraft's longitudinal axis. Variances in pitch and roll have been demonstrated to be reliable indicators of flight performance associated with various factors, including age and expertise (Kennedy, Taylor, Reade, & Yesavage, 2010; Taylor, Kennedy, Noda, & Yesavage, 2007), and cue utilisation (Wiggins, 2014). Pitch and roll were recorded as degrees from horizontal. The mean variance in pitch and roll were calculated for the selected two-minute segments.

Procedure

On arrival at the designated testing room, participants were presented with the information sheet to read, which indicated that some of the participants' flight data would be accessed by the researchers. After consenting to participate in the study, the participants completed a demographics questionnaire and the TST, administered on a laptop computer. The computer-based tasks took approximately 15 to 20 minutes to complete.

As part of the demographic questionnaire, participants were asked to supply their student IDs as a means of accessing the relevant flight data and linking up the flight data with performance on the cue acquisition measure. Flight data were accessed from a secure desktop and aircraft at the flight training school.

Results

Flight Performance

Flight performance was operationalised using two measures: (1) mean pitch and roll variance during the two segments extracted from Lesson 11, and (2) subjective instructor ratings of Lesson 11. Variance in pitch and roll during level flight and variance in pitch during landing were not normally distributed, and were transformed to achieve normality. Table 1 presents the transformation function and the Shapiro-Wilk tests of significance.

Table 1.
Flight Performance Transformation and Normality

	Transformation	Shapiro-Wilk
Level Flight Pitch	Square Root	.126
Level Flight Roll	Log10	.106
Landing Pitch	Square Root	.069
Landing Roll	-	.127
Instructor Ratings	-	.094

Cue Acquisition and Flight Control

The relationships between cue acquisition, flight experience and flight control were evaluated by analysing the predictive associations of TST scores and total flight hours on flight control performance. Four separate multiple linear regressions were conducted with variance in pitch during level flight, variance in roll during level flight, variance in pitch during landing, and variance in roll during landing as the outcome variables in each analysis. TST scores and total flight hours were the predictors in each analysis.

Analysis of flight controls during level flight revealed that the model explained 1.70% of the variance in pitch variance, $F(2, 34) = 0.29$, $p = .751$. Cue acquisition ($\beta = 0.12$, $t(36) = 0.69$, $p = .496$) and total flight hours ($\beta = -0.07$, $t(36) = 0.38$, $p = .707$) were not predictive of pitch variance during level flight. Further, the model explained 3.70% of the variance in roll variance during level flight, $F(2, 34) = 0.65$, $p = .770$. Cue acquisition ($\beta = 0.18$, $t(36) = 1.05$, $p = .302$) and total flight hours ($\beta = -0.09$, $t(36) = 0.53$, $p = .599$) were not predictive of roll variance during level flight.

During landing, the model explained 3.70% of the variance in pitch variance, $F(2, 33) = 0.63$, $p = .538$. Cue acquisition ($\beta = -0.08$, $t(35) = 0.48$, $p = .632$) and total flight hours ($\beta = 0.18$, $t(35) = 1.06$, $p = .299$) were not predictive of pitch variance during landing. Furthermore, the model explained 0.90% of the variance in roll variance, $F(2, 33) = 0.15$, $p = .550$. Cue acquisition ($\beta = 0.09$, $t(35) = 0.51$, $p = .611$) and total flight hours ($\beta = -0.04$, $t(35) = 0.23$, $p = .820$) were not predictive of roll variance during landing. Therefore, there was no linear association between cue acquisition and total flight hours and variance in pitch and roll during level flight and landing.

Cue Acquisition, Flight Experience, and Instructor Ratings

The relationships between cue acquisition, flight experience and instructor ratings of performance were evaluated by analysing the predictive associations of TST scores and total flight hours on instructor ratings. A multiple linear regression was conducted with instructor ratings of performance as the outcome variable, and TST scores and total flight hours as the predictors.

The model explained 21.20% of the variance in instructor ratings, $F(2, 33) = 4.43$, MSe = 0.01, $p = .020$. Cue acquisition ($\beta = 0.05$, $t(35) = 0.31$, $p = .763$) was not predictive of instructor ratings, but flight experience ($\beta = 0.46$, $t(35) = 2.94$, $p = .006$) was significantly predictive of instructor ratings. Greater total flight hours was associated with higher instructor ratings.

Cue Acquisition, and Flight Controls and Instructor Ratings by Flight Experience

To assess the role of experience in the relationship between cue acquisition and flight performance, correlation analyses were conducted on TST scores and measures of flight performance amongst less and more experienced participants. Given that pilot trainees in the tertiary education institution accumulate approximately 100 flight hours by the end of their first year of training, participants who reported 100 or less flight hours were categorised as less experienced, while participants who reported 101 or more flight hours were categorised as more experienced.

Amongst less experienced participants, TST scores were not significantly correlated with pitch (r = 0.15, p = .462) and roll during level flight (r = 0.07, p = .728), pitch (r = -0.16, p = .463) and roll during landing (r = -0.05, p = .809), and flight instructor ratings (r = 0.03, p = .888). Amongst more experienced participants, no significant correlation was observed between TST scores and pitch (r = 0.04, p = .900) and roll during level flight (r = 0.35, p = .259), pitch (r = 0.11, p = .745) and roll during landing (r = 0.31, p = .321), and flight instructor ratings (r = 0.13, p = .681). The relationship between cue acquisition and flight performance appeared to be similar irrespective of the experience of the pilot trainees.

Discussion

This study was designed to evaluate the relationship between cue acquisition and flight performance, controlling for total flight experience. Flight performance was operationalised as flight control, including the variance in pitch and roll during a selected flight, and instructor ratings of flight performance. It was hypothesised that greater TST scores would be associated with a lower variance in pitch and roll, and higher instructor ratings of flight performance, controlling for flight experience.

Contrary to the hypothesis, no relationship was evident between TST scores and variance in pitch and roll, and instructor ratings. Individual differences in cue acquisition appeared not to be predictive of domain-specific performance in the context of exercising operational control of an aircraft. However, consistent with previous findings (Bellenkes, Wickens, & Kramer, 1997; Li et al., 2003), flight experience appeared to be predictive of flight performance.

Nevertheless, this predictive effect of experience was only observed for instructor ratings, and not for operational control of an aircraft.

Cue acquisition was proposed to represent individual differences in the ability to acquire cues or feature-event/object associations that support improved cue utilisation, and subsequently, more effective situation assessment. Differences in cue utilisation, despite similar exposure to domain specific experiences (Hambrick et al., 2014), suggest that there may be differences in the availability of cues in memory. These observations led to the inference that the apparent disparity may stem from individual differences in the propensity to identify and acquire relevant and meaningful cues that ultimately contribute to performance outcome.

The outcomes of the present study suggest that cue acquisition may not contribute to performance in the context of flight control. However, there were a number of issues associated with the assessment of performance in the present study that may have resulted in difficulties establishing clear relationships. In particular, actual, real-world performance data is likely to be influenced by a range of factors, many of which are uncontrolled experimentally.

In the present study, flight performance data were derived from a single lesson to maintain consistency in both the structure of the flight operation, and the extent of exposure to domain-specific experiences at the time of testing. While flight performance was derived from the same lesson, there were variations in the nature of the flight, including the route, weather and time taken to complete the flight. During the selected lesson, pilot trainees were required to demonstrate a set of manoeuvres satisfactorily, but were not restricted in their flight decision-making nor in performing other manoeuvres given the need to adapt to different atmospheric or operational conditions. Consequently, these variations in the nature of individual flights may have resulted in considerable noise in the data.

Noisy data and variations in flight conditions necessitated the selection of measurement outcomes that allowed comparisons across participants. In the operational context, there are relatively few variables that remain consistent across participants and from which comparative assessments might be made. Therefore, periods of stable, level flight were selected on the assumption that maintaining straight and level flight through aircraft pitch and roll should constitute a measure of performance that occurs irrespective of the nature of flights.

Maintaining level flight, however, is also an activity that is impacted by the application of cues. Since cyclical variations in air movement cause disruptions to straight and level flight, cue utilisation can support the anticipation of aircraft responses to air movement, and thereby minimising any disruption and reducing the variability in aircraft pitch and roll (Wiggins, Azar, Hawken, Loveday, & Newman, 2014). Consequently, individual differences in cue utilisation may be one of the many factors that may have impacted the selected variables which were presumed to be constant across participants.

Although there is evidence to suggest that variations in aircraft pitch and roll might correspond to differences in performance on the basis of cue utilisation (e.g., Wiggins, 2014), it is also possible that there was insufficient variability in flight performance in the present study given that all pilot trainees undertook the same flight training. In a previous study conducted by Schriver, Morrow, Wickens and Talleur (2008), qualified pilots ranging in flight experiences, based on flight hours, type of ratings and certifications, and general aviation knowledge, were recruited from both university and surrounding community. Therefore, evaluating the performance of pilots with varying flight experiences and training structures might provide an opportunity to test, more explicitly, the relationship between cue acquisition and flight performance, accounting for differences in experience.

Like the flight performance data, the variability in instructor rating data was limited due to a restriction of range. Ratings tended to be clustered due to both limitations in the scale, and the propensity to record performance at, or about, the mid-point. While greater flight experience predicted higher instructor ratings, it is important to note that instructor ratings pertain to individual performance on Lesson 11. On the other hand, the accumulated fight experience comprises cumulative hours at the point of testing, which may have occurred prior to or subsequent to Lesson 11 for different participants. Consequently, the predictive association of flight experience on instructor ratings remains unclear.

There is an opportunity, in future research, either to track the performance of pilots longitudinally, and/or to examine performance in a simulated environment that enables the exercise of greater experimental control. The assessment of performance using a flight simulator allows for a more standardised approach to assessment that overcomes variations in the time of the flight, weather conditions, the time of day, and/or aircraft performance. Further, a cross-sectional assessment of performance in the flight simulator will clarify the predictive effect of experience on flight performance.

In combination, the outcomes of the present study failed to provide support for the proposition that cue acquisition would predict flight performance amongst trainees undertaking a standard flight. However, difficulties associated with the nature of the data may have masked any relationship between the propensity for cue acquisition and operational performance. Therefore, future research needs to be directed towards testing the relationship in a context that maintains sufficient realism, but which standardises more effectively, the nature of the flight and types of data acquired.

References

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68*(7), 569-579.

Crane, M. F., Brouwers, S., Wiggins, M. W., Loveday, T., Forrest, K., Tan, S. G. M., & Cyna, A. M. (2018). "Experience isn't everything": How emotion affects the relationship between experience and cue utilization. *Human Factors, 60*(5), 685–698. doi:10.1177/0018720818765800

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747. doi:10.1037/0003-066X.49.8.725

Evans, J. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459. doi:10.1016/j.tics.2003.08.012

Evans, J. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*(1), 255-278. doi:10.1146/annurev.psych.59.103006.093629

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104-114. doi:10.1016/j.neuropsychologia.2013.11.010

Goh, J., & Wiegmann, D. (2001). An investigation of the factors that contribute to pilots' decisions to continue visual flight rules flight into adverse weather. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(2), 26-29. doi:10.1177/154193120104500205

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45*, 34-45. doi:10.1016/j.intell.2013.04.001

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. doi:10.1177/154193120104500411

Kennedy, Q., Taylor, J. L., Reade, G., & Yesavage, J. A. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. *Aviation, Space, and Environmental Medicine, 81*(5), 489-497. doi: 10.3357/ASEM.2684.2010

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Klein, G. A. (2008). Naturalistic decision making. *Human factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: the original study plus a postscript. *Journal of Cognitive Engineering and Decision Making, 4*(3), 186-209. doi:10.1518/155534310x12844000801203

Li, G., Baker, S. P., Grabowski, J. G., Qiang, Y., McCarthy, M. L., & Rebok, G. W. (2003). Age, flight experience, and risk of crash involvement in a cohort of professional pilots. *American Journal of Epidemiology, 157*(10), 874-880. doi:10.1093/aje/kwg071

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2012). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 90-107. doi:10.1177/0018720812450911

Meinz, E. J., & Hambrick, D. Z. (2009). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill. *Psychological Science, 21*(7), 914-919. doi:10.1177/0956797610373933

Mueller, A. S., & Trick, L. M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis & Prevention, 48*, 472-479. doi:10.1016/j.aap.2012.03.003

Okoli, J., Watt, J., & Weller, G. (2017). Towards the classification of fireground cues: A qualitative analysis of expert reports. *Journal of Contingencies and Crisis Management, 25*(4), 197-208. doi:10.1111/1468-5973.12129

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 492-502. doi:0096-I515/80/0605-0492S00.75

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors: The Journal of Human Factors and Ergonomics Society, 50*(6), 864-878. doi:10.1518/001872008x374974

Taylor, J. L., Kennedy, Q., Noda, A., & Yesavage, J. A. (2007). Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. *Neurology, 68*(9). doi: 10.1212/01.wnl.0000255943.10045.c0

Todd, M. A., & Thomas, M. J. W. (2012). Flight hours and flight crew performance in commercial aviation. *Aviation, Space, and Environmental Medicine, 83*(8), 776. doi:10.3357/ASEM.3271.2012

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human factors, 44*(2), 189-197. doi:10.1518/0018720024497871

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Wiggins, M. W. (2014). Differences in situation assessments and prospective diagnoses of simulated weather radar returns amongst experienced pilots. *International Journal of Industrial Ergonomics, 44*(1), 18-23. doi:10.1016/j.ergon.2013.08.006

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science, 65*, 118-124. doi:10.1016/j.ssci.2014.01.006

Bridging Section

The aim of Study 4 (Chapter 7) was to evaluate the direct association between cue acquisition and task performance amongst individuals with existing experience within the domain of general aviation. The outcomes demonstrated that individual differences in cue acquisition were not sensitive to differences in flight control nor instructors' subjective ratings of performance following the initial exposure to domain-specific environment. However, Study 4 (Chapter 7) relied on noisy flight data collected during actual flight conditions.

Data noise resulted from the limited control that was able to be exercised over variables and that may have influenced flight performance. The lack of experimental control over variables, including weather conditions, may have masked any association between individual differences in cue acquisition and task performance. Consequently, the aim of Study 5 (Chapter 8) was to extend the findings of Study 4 (Chapter 7) by assessing task performance within a controlled, simulated setting.

The author of the present study contributed approximately 80% to the preparation of this paper in data collection, data analysis, and writing up of this paper.

CHAPTER EIGHT

Absence of Any Direct Association between Cue Acquisition and Performance in a Simulated Flight

Abstract

The aim of Study 5 was to address the limitations of Study 4 pertaining to noisy data by testing flight performance within a controlled simulated setting. Twenty-six qualified pilots completed the measure of cue acquisition. Participants flew a flight simulator, which included a critical event involving engine failure. Measures of cognitive load, including blood pressure, heart rate and oxygen consumption, were taken following the engine failure. Cue acquisition was not associated with flight performance and measures of cognitive load, accounting for flight experience. The findings corroborated the findings of Study 4 and suggest that cue acquisition may not impact later performance.

Keywords: cue acquisition, flight performance, cues, general aviation

Absence of Any Direct Association between Cue Acquisition and Performance in a Simulated Flight

Situation assessment is a cognitive process that involves the derivation and integration of information in the environment with information in memory to construct a working understanding of the presenting situation (Horrey & Wickens, 2001). Skilled situation assessment is particularly important in high-risk domains, such as aviation and firefighting, where it forms the basis on which decisions and action plans are selected or formed (Klein, Calderwood, & Clinton-Cirocco, 1986).

More accurate assessments of presenting situations are generally associated with superior domain-specific performance (Goh & Wiegmann, 2001; Mueller & Trick, 2012; Wiegmann, Goh, & O'Hare, 2002). For instance, pilots who are more accurate in their assessments of visibility and cloud ceiling are more likely to divert from deteriorating weather conditions compared to pilots whose assessments are less accurate (Wiegmann et al., 2002). Similarly, pilots who show a greater capacity to acquire and interpret in-flight weather reports tend to divert from deteriorating weather conditions at an earlier phase of a flight than pilots who are less capable in assessing in-flight weather reports (Johnson & Wiegmann, 2016).

Superior in-flight performance has also been associated with greater domain-specific experience (Chase & Simon, 1973; Johnson & Wiegmann, 2016). Experience allows for the acquisition of feature-event/object associations or cues in memory that support effective situation assessment (Ericsson & Charness, 1994). Through repeated and consistent exposure to the co-occurrence of features and events or objects across different situations, paired associations between these features and events or objects are likely to be acquired and integrated into schemas in memory (Kim, Seitz, Feenstra, & Shams, 2009).

The acquisition and integration of features and events or objects is consistent with System 1 processing, which involves the utilisation of cues in memory to construct swift and more accurate assessments (Evans, 2003, 2008). When familiar features in the environment are identified as matches to representations of features in memory, assessments of the presenting situations are formed on the basis of the information associated with the matched features (Klein, 2008). These assessments constitute events, expectancies, goals, and responses that the operators had experienced previously in association with the perceived features.

In the absence of relevant cues in memory, operators must rely on System 2 processing to undertake an effortful, conscious analysis of the varying features and events or objects in the environment to form working assessments (Evans, 2003). This is especially prevalent amongst less experienced operators who have insufficient exposure to domain-specific environments to acquire relevant cues (Bruder, Eißfeldt, Maschke, & Hasse, 2013; Crundall & Underwood, 1998). Consequently, the construction of assessments by less experienced operators is often delayed and likely to be prone to errors, resulting in slower and less accurate decision outcomes and performance.

In addition to the accuracy of matched features, the utilisation of cues during situation assessment contributes to improved performance outcomes by reducing cognitive load (Brouwers, Wiggins, Griffin, Helton, & O'Hare, 2017; Brouwers, Wiggins, Helton, O'Hare, & Griffin, 2016; Wiggins, Whincup, & Auton, 2018). By relying on cues to form assessments, operators utilise fewer cognitive resources by extracting relatively less perceptual information from the environment (Rieskamp & Hoffrage, 2008). Consequently, the residual cognitive resources, which would otherwise have been be utilised to identify, evaluate and interpret features and events or objects in the environment, remain available to manage additional demands, such as responding to other tasks or monitoring changes in the system state.

Cognitive load during the performance of a task is normally assessed either through subjective reports (Brouwers et al., 2017; Rubio, Diaz, Martin, & Puente, 2004; Tsang, & Velazquez, 1996), or through psychophysiological measures, including eye movements, cardiovascular responses and brain activities (Ahlstrom & Friedman-Berg, 2006; Wiggins et al., 2018; Wilson, 2002). Individual differences in cognitive load result from the ratio of task demands to individual resources (Wickens, 2008; Wierwille & Eggemeier, 1993). Task demands that exceed resources will result in excessive cognitive load, while task demands that match or are lower than individual capabilities will result in moderate or low cognitive load.

A number of different approaches to the assessment of cognitive load are often required to provide a more comprehensive assessment (Tsang, & Vidulich, 2006). In dynamic environments, psychophysiological measures offer the advantages of capturing responses continuously at the time of task performance, and are influenced to a lesser extent by retrospective subjective perceptions of individual performance (Brookhuis & de Waard, 2010).

Cardiovascular measures, including heart rate and blood pressure, are commonly used psychophysiological measures that are sensitive to changes in task demands (Mansikka, Simola, Virtanen, Harris, & Oksama, 2016). Increased cognitive load as a result of increased task demand is associated with increased heart rate and blood pressure (Brookhuis, de Vries, & de Waard, 1991; Mehler, Reimer, Coughlin, & Dusek, 2009; Reimer & Mehler, 2011; Wilson, 2002). For instance, when Mehler et al. (2009) increased the demands of a secondary task during a driving simulation task, participants' average heart rate increased with the increasing demands. Therefore, it might be concluded that increases in cardiovascular responses are associated with increases in cognitive load during the performance of a task.

While blood pressure and heart rate represent volumetric cardiovascular responses to cognitive load, the amount of oxygen consumed during a task represents a cardiovascular response at the molecular level. Levels of blood oxygenation appear to provide reliable psychophysiological measures of cognitive load (Sassaroli et al., 2008). Changes in the blood oxygenation level in the prefrontal cortex can be measured using the non-invasive imaging method, Functional Near-Infrared Spectroscopy (fNIRS). fNIRS records both the oxygenated haemoglobin (oxy-haemoglobin) and deoxygenate haemoglobin (deoxyhaemglobin) which in combination, reflect the oxygen being consumed during various activities with different levels of task demand (Herff et al., 2014).

Increased oxygenation in the prefrontal cortex is associated with increasing task demand, and, as a consequence, increases in cognitive load (Causse, Chua, Peysakhovich, Campo, & Matton, 2017; Hirshfield et al., 2009; Tsunashima & Yanagisawa, 2009). Evidence to support this assertion can be drawn from Causse et al. (2017) who observed greater levels of oxy-haemoglobin and lower levels of deoxyhaemoglobin during more difficult landing scenarios in a flight simulator task. This suggests that higher levels of oxygen consumption reflect higher cognitive load during task performance.

In many operational domains, increased cognitive load is associated with impaired performance (Brookhuis et al., 1991; Gateau, Durantin, Lancelot, Scannella, & Dehais, 2015; Sauvet et al., 2009). For instance, increased heart rate variability and blood pressure were associated with reduced vigilance amongst military pilots during and following a cross-country flight (Sauvet et al., 2009). Similarly, pilots who demonstrated higher levels of oxygen consumption during a difficult flight simulator task were less accurate in responding to Air Traffic Control (ATC) messages (Gateau et al., 2015).

Given that cue utilisation supports improved performance by reducing cognitive load (Brouwers et al., 2017), individual differences in cue utilisation should be associated with differences in psychophysiological measures of cognitive load. Wiggins et al. (2018), for instance, observed that lower cue utilisation was associated with greater levels of blood pressure when performing a high load task. Further, increased oxygen consumption in the prefrontal cortex during task performance tends to be greater with lower cue utilisation compared to higher cue utilisation (Sturman, Wiggins, Auton, & Loft, 2019). Consequently, compared to relatively higher cue utilisation, lower cue utilisation is likely to be associated with an experience of higher cognitive load, reflected in greater increase in heart rate, blood pressure and cortical oxygen consumption.

For cue utilisation to effectively reduce cognitive load, it is necessary for operators to possess relevant cues, which are acquired through individual experiences in domain-specific environments. However, previous studies have established that domain-specific experience is a poor predictor of performance outcome (Crane et al., 2018; Hambrick et al., 2014; Meinz & Hambrick, 2009; Todd & Thomas, 2012), suggesting that individuals differ in the extent to which they acquire cues from similar experiences.

Real-world situations are often characterised by a diverse range of features and events that may or may not be associated with one another. For operators to acquire relevant cues, they must, therefore, rely on the ability to extract reliable associations between features and events or objects from noisy environments (Reber, Kassin, Lewis, & Cantor, 1980). Given that individuals differ in their levels of cue utilisation and performance outcome despite similar experiences, it might be inferred that individuals differ in their cue acquisition or ability to extract meaningful associations given similar levels of experience.

The Timed-Search Task (TST) is a domain-independent instrument developed and validated in the preceding four studies to measure individual differences in cue acquisition. In Study 2, cue acquisition was associated with the Feature Identification Task (FIT) and Feature Recognition Task (FRT) components of the general aviation cue utilisation, having accounted for flight experience. Higher cue acquisition was associated with lower response latency in identifying and higher accuracy in recognising relevant cues. These associations between cue acquisition and components of cue utilisation suggest that individual differences in cue acquisition reflect differences in the ability to identify and recognise relevant cues in novel environments.

Given that higher cue utilisation is associated with lower cognitive load, it might be argued that operators with higher cue acquisition are likely to experience lower cognitive load when acquiring cues in novel environments. A faster rate in identifying and recognising relevant cues, arguably, allows for cue utilisation that supports situation assessment. Therefore, higher cue acquisition should be associated with lower increase in heart rate, blood pressure and cortical oxygen consumption during, and following, task performance.

The outcomes of Study 4 indicated that individual differences in cue acquisition were not associated with measures of flight performance. However, there were a number of methodological factors which may have contributed to the null findings, including noisy data that is characteristic of actual, real-world flight performance. To rectify the methodological issues of Study 4, the present study was designed to assess, within a controlled setting, the relationship between individual differences in cue acquisition and performance in the context of general aviation, accounting for domain-specific experience.

Flight performance was evaluated within a flight simulator. Flight controls and landing success following an emergency were employed as measures of flight performance. Further, cognitive load was assessed through cardiovascular measures and oxygen consumption in the prefrontal cortex. On the basis of previous findings, it was predicted that, accounting for flight experience, higher cue acquisition would be associated with improved flight controls, greater landing success, and relatively lower increase in heart rate, blood pressure and cortical oxygen consumption.

Methodology

Participants

Twenty-six pilots based in New South Wales, Australia, were recruited through convenience sampling following ethical approval from the relevant Human Research Ethics Committees (see Appendix D). Participants were required to hold Pilot License Qualifications issued by the Civil Aviation Safety Authority. All participants were compensated with a AU$30.00 shopping voucher.

Participants comprised 24 male pilots and two female pilots ranged from 18 to 70 years of age (M = 32.35, SE = 2.93). Pilots ranged in the licenses that they held, including a Student Pilot License (n = 6), a Recreational Pilot License (n = 3), a Private Pilot License (n = 9), a Commercial Pilot License (n = 6), and an Air Transport Pilot License (n = 2). Table 1 illustrates the length of time, at the time of testing, the pilots had held their licenses.

Table 1.
Length of Time Pilots Have Held their Licenses

	Less than 6 months	6 months to 1 year	1 to 2 years	3 to 4 years	5 to 6 years	More than 10 years
Frequency	7	3	3	3	3	7

Most participants reported flying fewer than five hours a week in the last year (n = 20), while five participants reported flying between five to ten hours a week, and one flying between 15 to 20 hours a week. Flight experience ranged from 15 hours to 8000 hours (M = 666.04, SE = 340.43) with experience as pilot in command ranging from zero to 2000 hours (M = 215.33, SE = 83.00). Participants ranged in their Instrument Flight Rules experience from none to 4400 hours (M = 195.96, SE = 168.83). In the previous 90 days prior to testing, the number of hours flown by participants ranged between none to 200 hours (M = 16.46, SE = 7.68).

Materials

Timed-Search Task

The Timed-Search Task (TST) was developed and validated as a measure of cue acquisition in Studies 1 (Chapter 4) and 2 (Chapter 5). The same version as administered in preceding studies was employed in the present study.

Redbird FMX Flight Simulator

Flight performance was assessed using a Redbird FMX, a moderate-fidelity flight simulator. The Redbird FMX is equipped with a 3-axis electric motion platform to generate motion feedback with a 50° pitch, 40° roll and 60° yaw movements. The view through the cockpit is displayed on six monitors, which supports a 200° view.

An enhanced version of the Microsoft Flight Simulator is used to run the simulated scenarios. The simulated aircraft is configured according to the Cessna 172 with the Garmin G1000 employed as the electronic flight instrument display. Flight data, including pitch, roll, yaw and altitude, were recorded at every half second using the software 'Insight'.

All participants flew the same scenario involving Visual Flight Rules (VFR) from Taree, New South Wales to Coffs Harbour, New South Wales, a distance of 99 nautical miles. Participants were instructed to maintain flight at 7500 feet, and keep, as closely as possible, to the straight, direct route displayed on the navigation screen. At 32 nautical miles from departure, the engine of the aircraft was failed.

The dependent variables measured during the flight simulation were: (1) landing success, (2) final location (either the nearby Port Macquarie Airport or an open field), and (3) flight control performance. The variable landing success consisted of two levels, successful or unsuccessful. A flight outcome was considered as unsuccessful if the pilot conducted a manoeuvre that resulted in significant damage to the aircraft rendering it unflyable. Conversely, a successful landing was designated as an outcome where the aircraft remained flyable following landing.

Flight control performance was derived from two variables including the mean variance in pitch and roll during two selected two-minute segments during the flight. The first segment constituted the first two minutes when participants reached level flight at 7500 feet, while the second segment constituted of the two minutes immediately after the engine had been failed.

Flight control performance was calculated by computing the differences in the mean variance of pitch and roll between post-failure and level flight. Positive scores signified an increase in mean variance in pitch and roll post-failure compared to level flight, while negative scores signified decrease in variance of pitch and roll post-failure compared to level flight. Values further from zero signified greater change.

Near Infrared Spectroscopy (NIRS)

During the flight simulation, a Portalite Near Infrared Spectroscope (NIRS) sensor was fitted to the right side of the participants' forehead, one centimetre above the eyebrow. The Portalite NIRS uses light emitting diodes with 760 and 850 nm wavelength to measure cerebral consumption of oxygen at a 50 Hz sampling rate.

Oxyhemoglobin (O2Hb) and deoxyhemoglobin (HHb) are two measures of haemoglobin concentrations in the cerebral tissues captured by the Portalite NIRS. The level of oxygen consumption (rSO2) was calculated by computing the ratio of O2Hb to total haemoglobin (O2HB + HHb; Ekkekakis, 2009; Gratton & Fabiani, 2006). The variables measured were oxygen consumption during baseline and a two-minute segment following engine failure. During the baseline measure, participants were asked to sit quietly in a relaxed position for two minutes. rSO2 was measured for both segments.

The dependent variables in the present study consisted of a relative measure of rSO2, which represented the change in oxygen consumption. The relative measure was between post-failure and baseline, which were calculated by computing the differences between post-failure and baseline rSO2. Positive scores signified an increase in rSO2 post-failure compared to baseline, while negative scores signified decrease in rSO2 post-failure compared to baseline. Greater values away from zero signified greater change.

Blood Pressure Monitor

The Omron Wrist Blood Pressure Monitor is a non-invasive device used to measure blood pressure and heart rate. It was fitted around the wrist while participants were seated in a relaxed position with an elbow propped on the desk in 45° angle. During the measurement period, participants were required to keep body movement to a minimum to reduce error in measurement.

Systolic pressure, diastolic pressure and heart rate were recorded prior to, and following, the simulator task. Changes in systolic pressure, diastolic pressure and heart rate were calculated by subtracting the pre-flight values from the post-flight values. Positive scores signified increases in systolic pressure, diastolic pressure and heart rate post-flight compared to pre-flight, while negative scores signified decreases in systolic pressure, diastolic pressure and heart rate post-flight compared to pre-flight. Greater values away from zero signified greater change.

Procedure

Participants were recruited to attend a two-hour testing session at the University Simulation Hub. They were presented with the information sheet to read, were given the opportunity to ask questions, and were provided with a consent form to read and sign if they agreed to participate in the study. The demographic questionnaire and the TST was then completed on a desktop computer, which took approximately 15 to 20 minutes.

Having completed the TST, participants were presented with the simulator task paper-based materials, including the description of the scenario, flight plan, load sheet, charts, weather, and maps. Participants had approximately ten minutes to study the documents, after which they were permitted five minutes to familiarise themselves with the instruments in the flight simulator. During this five-minute familiarisation-period, the simulated scenario was not operating.

Before the commencement of the simulator task, participants were required to complete the Simulator Sickness Questionnaire to ensure that they were fit to participate in the simulator task. They were also advised to indicate to the researcher if they experienced any nausea during the course of the simulator task.

The first measure of blood pressure and heart rate was taken before participants were fitted with the Portalite NIRS and required to sit quietly for two minutes. The simulated scenario was then initiated, and participants flew for approximately 32 nautical miles or 20 minutes before the critical event was activated. All participants tried to identify the problem and attempted to land the aircraft in either an open field or the nearby Port Macquarie airport. On crashing or landing and securing the aircraft, the simulated scenario was terminated, and the participants were asked to step out of the flight simulator.

A second measure of blood pressure and heart rate was taken, and the Portalite NIRS was removed. The Simulator Sickness Questionnaire was administered again to ensure that there were no changes to reported symptoms. Participants were then debriefed regarding the critical event and given the opportunity to discuss their experiences of the flight simulation.

Results

Flight Performance

The three measures of flight performance were: (1) landing success, (2) final location, and (3) flight control performance. Ten participants landed the aircraft successfully, while 12 participants crashed the aircraft. Half the participants who landed successfully arrived at Port

Macquarie airport, while the other half landed on an open field. Table 2 illustrates the distribution of landing success by final location.

Table 2.
Distribution of Landing Success by Final Location

		Landing Success		
		Landed	Crashed	Total
Landing Location	Port Macquarie	8	3	11
	Open Field	2	10	12
Total		10	13	23

Flight Control Performance

A Shapiro-Wilk test indicated that variance in pitch during level flight was normally distributed (p = .289). However, variance in pitch post-failure (p < .001), variance in roll during level flight (p < .001), and variance in roll post-failure (p = .026) were not normally distributed.

Two separate sets of Wilcoxon signed-ranked tests were conducted independently on the variance in pitch and roll, comparing level flight and post-failure responses. There were statistically significant differences between ranks at level flight and ranks post-failure for variance in pitch, Z = 2.49, p = .013, and between level flight ranks and post-failure ranks for variance in roll, Z = 3.29, p = .001. Variance in pitch and roll were significantly greater post-failure compared to level flight.

Differences between level flight and post-failure variance in pitch and roll were computed to derive two dependent variables. A Shapiro-Wilk test indicated that changes in variance in pitch (p = .944), and variance in roll (p = .091) were normally distributed.

Psychophysiological Measures

Two psychophysiological measures were collected to assess changes in biological responses to the critical event. First, systolic pressure, diastolic pressure and heart rate were compared between pre-flight and post-flight. Second, rSO2 was compared between baseline and post-failure.

Blood Pressure and Heart Rate

A Shapiro-Wilk test indicated that measures of systolic pressure pre-flight (p = .291) and post-flight (p = .973), diastolic pressure pre-flight (p = .989) and post-flight (p = .963), and heart rate pre-flight (p = .845) and post-flight (p = .786) were normally distributed.

Three separate sets of paired-sample t-tests were conducted independently on systolic pressure, diastolic pressure and heart rate comparing pre-flight and post-flight responses. There were statistically significant differences between pre-flight (M = 124.45, SE = 2.40) and post-flight systolic pressure (M = 138.82, SE = 2.52, $t(21) = 6.78$, $p < .001$), between pre-flight (M = 81.26, SE = 1.91) and post-flight diastolic pressure (M = 92.96, SE = 2.81, $t(22) = 5.46$, $p < .001$), and between pre-flight (M = 77.33, SE = 2.44) and post-flight heart rate (M = 82.54, SE = 2.96, $t(23) = 3.18$, $p = .004$). Systolic pressure, diastolic pressure and heart rate were significantly higher post-flight compared to pre-flight.

Differences between pre-flight and post-flight systolic pressure, diastolic pressure and heart rate were computed to derive three dependent variables. A Shapiro-Wilk test indicated that changes in systolic pressure (p = .777), diastolic pressure (p = .697), and heart rate (p = .171) were normally distributed.

Level of Oxygen Consumption

A Shapiro-Wilk test indicated that measures of rSO2 at baseline (p = .473), and post-failure (p = .208) were normally distributed. A paired-sample t-test was conducted on the rSO2 comparing baseline responses and post-failure responses. There was no statistically significant difference between baseline (M = 0.74, SE = 0.02) and post-failure rSO2 (M = 0.76, SE = 0.02; $t(19) = 1.97$, $p = .064$). Blood oxygenation level following engine failure was not different from baseline. A strong correlation between baseline and post-failure rSO2 ($r = 0.90$, $p < .001$) suggests a case of multicollinearity.

The difference between baseline and post-failure rSO2 was computed to derive a dependent variable. A Shapiro-Wilk test indicated that post-failure rSO2 relative to baseline (*p* = .981) was normally distributed.

Cue Acquisition and Flight Performance

The relationship between cue acquisition and flight performance was evaluated accounting for flight experience. Individual analysis was conducted on each of the three measures of flight performance: (1) landing success, (2) final location, and (3) flight control performance.

Landing Success

To assess whether cue acquisition predicted landing success, accounting for flight experience, a binary logistic regression was conducted with participants' landing success as the outcome variable. The predictive variables were level of cue acquisition reflected in the performance scores on the Timed-Search Task and flight experience reflected in the total accumulated flight hours. A Shapiro-Wilk test indicated that TST scores (*p* = .149) were normally distributed. However, the variable flight hours was square root transformed to achieve normality (*p* = .088).

The logistic regression model accounted for 27.70% of the variance as to whether or not participants successfully landed or crashed the aircraft, χ^2 (2) = 4.79, *p* = .091, and correctly classified 66.70% of the participants. Cue acquisition was not a significant predictor of landing success, Wald (1) = 3.52, *B* = -0.19, *p* = .061, and neither was flight experience, Wald (1) = 0.20, *B* = -0.04, *p* = .654. The odds ratio for cue acquisition was 0.83 (95% CI = 0.68 – 1.01), and flight experience was 0.96 (95% CI = 0.81 – 1.14). Cue acquisition and flight experience were not predictive of landing success.

Final Location

The predictive effect of cue acquisition on final location, accounting for flight experience, was assessed with a binary logistic regression. Participants' final location of either Port

Macquarie airport or a field served as the dependent variable. The predictive variables were TST scores and transformed flight hours.

The logistic regression model accounted for 17.50% of the variance as to the final location of the aircraft, χ^2 (2) = 2.94, *p* = .230, and correctly classified 61.90% of the participants. Neither cue acquisition, Wald (1) = 2.35, *B* = -0.14, *p* = .126, nor flight experience, Wald (1) = 0.12, *B* = -0.03, *p* = .729, were significant predictors of final location. The odds ratio for cue acquisition was 0.87 (95% CI = 0.73 – 1.04), and flight experience was 0.97 (95% CI = 0.83 – 1.14). Cue acquisition, accounting for flight experience, was not predictive of the final location of the aircraft.

Flight Control Performance

The predictive effect of cue acquisition on flight control performance, accounting for flight experience, was evaluated using two separate multiple linear regressions. The outcome variables were the changes in variance of pitch, and the changes in variance of roll from level flight to post the engine failure. TST scores and transformed flight hours were the predictive variables for both analyses.

The analysis of pitch control revealed that the model explained 6.40% of the variance in pitch, $F(2, 15) = 0.51$, $p = .611$. Cue acquisition (β = -0.08, $t(17) = 1.00$, $p = .332$) and total flight hours (β = 0.02, $t(17) = 0.23$, $p = .825$) were not predictive of changes in variance of pitch.

The analysis of roll control revealed a similar model, which explained 8.20% of the variance in roll, $F(2, 17) = 0.76$, $p = .483$. Cue acquisition (β = -0.64, $t(19) = 1.23$, $p = .236$) and total flight hours (β = 0.01, $t(19) = 0.03$, $p = .980$) were not predictive of changes in variance of roll. Therefore, cue acquisition was not associated with changes in the variances of pitch and roll, accounting for flight experience, following the critical event.

The role of experience in the relationship between cue acquisition and flight control performance was evaluated. The mean total flight hours, excluding outliers, was used as the cut-off score to categorise participants into less and more experienced groups. Participants who reported 170 or less hours were categorised as less experienced, and those who reported 171 or more hours were categorised as more experienced.

Correlation analyses were conducted on TST scores and flight control performance in less and more experienced participants. Amongst less experienced participants, TST scores were not significantly correlated with changes in pitch (r = -0.31, p = .455), and roll variance (r = -0.17, p = .631). Similarly, TST scores were not significantly correlated with changes in pitch (r = -0.38, p = .222), and roll variances amongst more experienced participants (r = -0.42, p = .178). Cue acquisition was not associated with flight control performance amongst either relatively less, or more experienced pilots.

Cue Acquisition and Psychophysiological Measures

The relationships between cue acquisition and psychophysiological responses were evaluated accounting for flight experience. Individual analysis was conducted on each of the two measures of psychophysiological responses: (1) blood pressure and heart rate, and (2) level of oxygen consumption.

Blood Pressure and Heart Rate

The predictive effect of cue acquisition, accounting for flight experience, was evaluated against the two measures of blood pressure, and heart rate. Three separate multiple regression analyses were conducted with the changes in systolic pressure, changes in diastolic pressure and changes in heart rate as outcome variables. TST scores and transformed flight hours served as the predictive variables.

The model explained 6.30% of the variance in the change in systolic pressure, F(2, 17) = 0.57, p = .577. Neither cue acquisition (β = 0.19, t(19) = 0.41, p = .690) nor total flight hours

(β = 0.41, t(19) = 0.89, p = .384) were predictive of changes in systolic pressure. A similar outcome was evident in the analysis of changes in diastolic pressure, F (2, 17) = 1.21, R^2 = 0.12, p = .324. Cue acquisition (β = 0.69, t(19) = 1.55, p = .139) and total flight hours (β = -0.12, t(19) = -0.28, p = .786) were not predictive of changes in diastolic pressure.

The results also indicated that the model explained 4.00% of the variance in the change in heart rate, F(2, 18) = 0.04, p = .961. Similarly, neither cue acquisition (β = -0.09, t(20) = -0.26, p = .797) nor total flight hours (β = 0.05, t(20) = 0.14, p = .890) were predictive of changes in heart rate. Overall, it appeared that cue acquisition, accounting for flight experience, was not associated with changes in blood pressure and heart rate following the simulator task.

Level of Oxygen Consumption

A multiple regression analysis was undertaken on the relative rSO2 between baseline and post-failure to evaluate the predictive effect of cue acquisition, accounting for flight experience. TST scores and transformed flight hours served as the predictive variables.

The results indicated that the model explained 3.00% of the variance in relative rSO2, F(2, 16) = 0.24, p = .786. Consistent with the previous results, neither cue acquisition (β = 0.00, t(18) = 0.24, p = .810) nor total flight hours (β = 0.00, t(18) = 0.64, p = .529) were predictive of relative rSO2. Therefore, cue acquisition was not predictive of level of oxygen consumption following the critical event, accounting for flight experience.

Discussion

This study was designed to test the outcomes of Study 4 in a more experimentally controlled, simulator setting. The influence of cue acquisition on performance was investigated within the domain of general aviation. It was hypothesised that, controlling for flight experience, higher cue acquisition would be associated with: (a) smaller changes in the variance of pitch and roll control of the aircraft following an engine failure during a simulated flight, and (b) a greater likelihood in landing the aircraft safely. Further, it was predicted that higher cue acquisition would be associated with lower perceived cognitive load following a critical event as reflected in lower changes in heart rate and blood pressure, and a lower consumption of oxygen in the prefrontal cortex.

Contrary to the hypotheses, cue acquisition was not associated with the likelihood of landing the aircraft safely, differences in flight control, nor changes in psychophysiological responses. Therefore, the results suggest that individual differences in cue acquisition, as measured by the TST, were not associated with flight performance nor cognitive load following a critical event in a simulated cross-country flight task.

In contrast to Study 4, where real-world flight performance data were collected, the present study was conducted within a simulated context to control for variations in flying conditions experienced by the participants. However, despite the efforts to control for task demands, the outcomes of the present study confirm the conclusions drawn from Study 4, suggesting that cue acquisition, at least in the form assessed by the TST, may not contribute to differences in flight performance.

The consistency in the outcomes of Studies 4 and 5 provides support for the proposition that cue acquisition may constitute a trait that impacts the very early stages of skill acquisition with no direct, linear association with performance outcomes at the latter stages of practice. Following the successful acquisition of relevant cues, the influence of other processes, such as deliberate practice (Ericsson, 2004; Lehmann & Ericsson, 1997), cognitive biases (Gilbey & Hill, 2012), and cue utilisation (Loveday, Wiggins, Searle, Festa, & Schell, 2012) may become more prominent.

Support for this proposition can be drawn from Renshaw and Wiggins (2017) who investigated the influence of different predictors on skill acquisition in the context of learning to operate an Unmanned Aerial Vehicle (UAV). At the initial stages of skill acquisition, variables such as spatial visualisation and video game experience predicted performance.

However, they became less predictive of performance as further experience was acquired. The findings suggest that performance outcomes are impacted by different variables at different stages of skill acquisition.

The outcomes reported by Renshaw and Wiggins (2017) suggest that performance at the earlier stages of skill development is related to processes that are domain-independent, while performance at the later stages are likely to engage processes that capitalise on domain-specific experiences. Therefore, cue acquisition, as a domain-independent process, is likely to be predictive of performance only at the earlier stages of skill development. Given that participants in the present study were skilled, qualified pilots, the absence of an association between cue acquisition and flight performance may reflect the relatively diminished impact of cue acquisition on later stages of skill acquisition.

During performance in the flight simulator, there were significant increases in blood pressure, heart rate and levels of oxygen consumption in the prefrontal cortex following the critical event. These observations suggest that the engine failure resulted in an increase in cognitive load. However, no association was evident between individual differences in cue acquisition and changes in cognitive load.

To test the proposition that cue acquisition is only predictive of performance at the early stages of skill acquisition, a longitudinal study is necessary with individual differences in cue acquisition assessed at the outset, and prior to the acquisition of domain-specific cues. Further, a longitudinal study will establish whether there are changes in the influence of cue acquisition as operators gain proficiency in the domain-specific task performance.

In the context of general aviation, cue acquisition can be assessed prior to the commencement of a flight training program. Subsequently, measures of cognitive load and flight performance, comparing performance in a flight simulator and real-world flight data, can be taken periodically at different stages of the training program. Changes in the influence of cue acquisition on flight performance and cognitive load can then be tracked across the different stages of skill development to determine its predictive effect.

Finally, it is important to evaluate the findings of the present study considering the small sample size for regression analyses. Increasing sample size may improve statistical power to detect the effects of cue acquisition on measures of cognitive load and flight performance. A power analysis, assuming small effect size, indicated a total sample size of 79 is recommended to provide a robust test for the relationships between cue acquisition and cognitive load and flight performance.

The present study remedied the issue of noisy data in Study 4 by investigating the association between cue acquisition and performance in a more controlled, simulated setting. Further, the concern regarding insufficient variability in performance was addressed by recruiting licensed pilots across a range of flight experiences. The outcomes of the present findings corroborated the findings of Study 4, demonstrating a lack of support for the predictive influence of individual differences in cue acquisition on subsequent domain-specific performance outcome.

References

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*(7), 623-636. doi:10.1016/j.ergon.2006.04.002

Al-abood, S. A., Davids, K., & Bennett, S. J. (2001). Specificity of task constraints and effects of visual demonstrations and verbal instructions in directing learners' search during skill acquisition. *Journal of Motor Behavior, 33*(3), 295-305. doi: 10.1080/00222890109601915

Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention, 23*(4), 309-316. doi:10.1016/0001-4575(91)90008-s

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention, 42*(3), 898-903. doi:10.1016/j.aap.2009.06.001

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics, 60*(11), 1500-1515. doi:10.1080/00140139.2017.1330494

Brouwers, S., Wiggins, M. W., Helton, W. S., O'Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. *Frontiers in Psychology, 7*, 435. doi:10.3389/fpsyg.2016.00435

Bruder, C., Eißfeldt, H., Maschke, P., & Hasse, C. (2013). Differences in monitoring between experts and novices. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 295-298. doi:10.1177/1541931213571065

Causse, M., Chua, Z., Peysakhovich, V., Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific Reports, 7*(1): 5222. doi:10.1038/s41598-017-05378-x

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81. doi:10.1016/0010-0285(73)90004-2

Crane, M. F., Brouwers, S., Wiggins, M. W., Loveday, T., Forrest, K., Tan, S. G. M., & Cyna, A. M. (2018). "Experience isn't everything": How emotion affects the relationship between experience and cue utilization. *Human Factors, 60*(5), 685–698. doi:10.1177/0018720818765800

Crundall, D. E., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics, 41*(4), 448-458. doi:10.1080/001401398186937

Ekkekakis, P. (2009). Illuminating the black box: Investigating prefrontal cortical hemodynamics during exercise with near-infrared spectroscopy. *Journal of Sport and Exercise Psychology, 31*, 505–553. doi:10.1123/jsep.31.4.505

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic medicine: Journal of the Association of American Medical Colleges, 79*(10 Suppl), 70-81. doi:10.1097/00001888-200410001-00022

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747. doi:10.1037/0003-066X.49.8.725

Evans, J. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459. doi:10.1016/j.tics.2003.08.012

Evans, J. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59*(1), 255-278. doi:10.1146/annurev.psych.59.103006.093629

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., & Dehais, F. (2015). Real-time state estimation in a flight simulator using fnirs. *PLoS ONE, 10*(3). doi:10.1371/journal.pone.0121279

Gilbey, A., & Hill, S. (2012). Confirmation bias in general aviation lost procedures. *Applied Cognitive Psychology, 26*(5), 785-795. doi:10.1002/acp.2860

Goh, J., & Wiegmann, D. A. (2001). An investigation of the factors that contribute to pilots' decisions to continue visual flight rules flight into adverse weather. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(2), 26-29. doi:10.1177/154193120104500205

Gratton, G., & Fabiani, M. (2006). Optical imaging of brain function in R. Parasuraman, M. Rizzo (Eds.), Neuroergonomics: The Brain At Work, Oxford University. In *Oxford Series in Human-Technology Interactions* (pp. 65–81). Cambridge, MA.

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45*, 34-45. doi:10.1016/j.intell.2013.04.001

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience, 7*, 935. doi:10.3389/fnhum.2013.00935

Hirshfield, L. M., Solovey, E., Girouard, A., Kebinger, J., Jacob, R., Sassaroli, A., & Fantini, S. (2009). Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy.

In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2185-2194. doi: 10.1145/1518701.1519035

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. doi:10.1177/154193120104500411

Johnson, C. M., & Wiegmann, D. A. (2016). Vfr into imc: Using simulation to improve weather-related decision-making. *The International Journal of Aviation Psychology, 25*(2), 63-76. doi:10.1080/10508414.2015.1026672

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Klein, G. A. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society Annual Meeting, 30*(6), 576-580. doi:10.1177/154193128603000616

Lehmann, A. C., & Ericsson, A. K. (1997). Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students. *Psychomusicology: A Journal of Research in Music Cognition, 16*(1-2), 40. doi:10.1037/h0094068

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2012). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 125-137. doi:10.1177/0018720812448475

Mansikka, H., Simola, P., Virtanen, K., Harris, D., & Oksama, L. (2016). Fighter pilots' heart rate, heart rate variation and performance during instrument approaches. *Ergonomics*, 1-9. doi:10.1080/00140139.2015.1136699

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record, 2138*(1), 6-12. doi:10.3141/2138-02

Meinz, E. J., & Hambrick, D. Z. (2009). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill. *Psychological Science, 21*(7), 914-919. doi:10.1177/0956797610373933

Mueller, A. S., & Trick, L. M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis & Prevention, 48*, 472-479. doi:10.1016/j.aap.2012.03.003

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 492-502. doi:0096-I515/80/0605-0492S00.75

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics, 54*(10), 932-942. doi:10.1080/00140139.2011.604431

Renshaw, P. F., & Wiggins, M. W. (2017). The predictive utility of cue utilization and spatial aptitude in small visual line-of-sight rotary-wing remotely piloted aircraft operations. *International Journal of Industrial Ergonomics, 61*, 47-61. doi:10.1016/j.ergon.2017.05.014

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica, 127*(2), 258-276. doi:10.1016/j.actpsy.2007.05.004

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology, 53*(1), 61-86. doi:10.1111/j.1464-0597.2004.00161.x

Sassaroli, A., Zheng, F., Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J. K., & Fantini, S. (2008). Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences, 01*(02), 227-237. doi:10.1142/s1793545808000224

Sauvet, F., Jouanin, J., Langrume, C., Beers, P., Papelier, Y., & Dussault, C. (2009). Heart rate variability in novice pilots during and after a multi-leg cross-country flight. *Aviation, Space, and Environmental Medicine, 80*(10), 862-869. doi:10.3357/asem.2531.2009

Sturman, D., Wiggins, M. W., Auton, J. C., & Loft, S. (2019). Cue utilization differentiates resource allocation during sustained attention simulated rail control tasks. *Journal of Experimental Psychology*. doi:10.1037/xap0000204

Todd, M. A., & Thomas, M. J. W. (2012). Flight hours and flight crew performance in commercial aviation. *Aviation, Space, and Environmental Medicine, 83*(8), 776. doi:10.3357/ASEM.3271.2012

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39*(3), 358-381. doi: 10.1080/00140139608964470

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 243-268). John Wiley & Sons, Inc. doi: 10.1002/0470048204

Tsunashima, H., & Yanagisawa, K. (2009). Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS). *Computational Intelligence and Neuroscience, 2009*, 1-12. doi:10.1155/2009/164958

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of Human Factors and Ergonomic Society, 50*(3), 449-455. doi:10.1518/001872008X288394

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human factors, 44*(2), 189-197. doi:10.1518/0018720024497871

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35*(2), 263-281. doi:10.1177/001872089303500205

Wiggins, M. W., Whincup, E., & Auton, J. C. (2018). Cue utilisation reduces effort but increases arousal during a process control task. *Applied Ergonomics, 69*, 120-127. doi:10.1016/j.apergo.2018.01.012

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18. doi:10.1207/s15327108ijap1201_2

CHAPTER NINE

General Discussion

In high-risk operations, where poor performance can often result in significant negative consequences to operators, organisations, and the general public, skilled performance is critical in minimising undesirable outcomes. Skilled performance, characterised by rapid and accurate responses (Beilock, Bertenthal, McCoy, & Carr, 2004), is reliant upon precise assessments of the presenting situations and/or problems (Engström, Gregersen, Hernetkoski, Keskinen, & Nyberg, 2003; Goh & Wiegmann, 2001a, 2001b; Mueller & Trick, 2012; Wiegmann, Goh, & O'Hare, 2002). These assessments are constructed through the process of situation assessment, which involves the derivation and integration of information in the environment from information in memory (Horrey & Wickens, 2001).

Effective situation assessment is achieved through the application of domain-specific feature-event/object associations in memory, also referred to as cues (Klein, 2008; Klein, Calderwood, & Clinton-Cirocco, 1986; Wiggins, 2012). During situation assessment, perceived features in the environment are matched against representations of features in memory (Klein, 2008). When matches are perceived, the relevant information associated with the perceived features is acquired from memory and acts as a template on which assessments of the presenting situations and/or problems are based.

The application of cues to support the construction of accurate and rapid assessments and responses is referred to as cue utilisation (Wiggins, 2012). Higher cue utilisation has been associated with improved domain-specific task performance (Abernethy, 1990; Loveday, Wiggins, Harris, O'Hare, & Smith, 2012; Loveday, Wiggins, Searle, Festa, & Schell, 2012; Vansteenkiste, Vaeyens, Zeuwts, Philippaerts, & Lenoir, 2014; Watkinson, Bristow, Auton, McMahon, & Wiggins, 2018). Furthermore, cue utilisation within a specific domain has been associated with performance in an unrelated, novel domain (Brouwers, Wiggins, Griffin, Helton, & O'Hare, 2017; Renshaw & Wiggins, 2017), suggesting an underlying domain-general trait that governs the acquisition of feature-event/object associations.

In contrast to cue utilisation, which is reliant upon the availability of domain-specific cues in memory, cue acquisition may operate in the absence of domain-specific experiences. Individual differences should, therefore, be observable or measurable during the initial exposure to domain-specific environment where previous domain-specific experiences and cues have not yet been acquired.

Experience allows the exposure of operators to the co-occurrence of features and events or objects within domain-specific environments, where the opportunity to establish meaningful associations results in their correspondence in the form of cues (Kim, Seitz, Feenstra, & Shams, 2009; Reber, Kassin, Lewis, & Cantor, 1980). However, given that similar experience in a domain is not always associated with similar levels of cue utilisation (Crane et al., 2018; Hambrick et al., 2014; Todd & Thomas, 2012), there are likely to be individual differences in the rate at which cues are acquired from similar experiences. Therefore, skilled performance is influenced by the effectiveness of the situation assessment process, which in turn, is impacted by cue utilisation, and subsequently, individual differences in cue acquisition.

Despite the inference that skilled performance is presumably linked to individual differences in cue acquisition through cue utilisation, it remains unclear whether cue acquisition, in itself, has a direct impact on later operational performance. Such an association can serve as a basis for the application of cue acquisition as a predictive measure for selection for later skilled performance. The present programme of research was designed to investigate the predictive influence of individual differences in cue acquisition on future performance.

Study 1 (Chapter 4) was designed to develop and test an instrument to measure individual differences in cue acquisition. The acquisition or learning of implicit associations has previously been assessed using a number of different methods of assessment (Kellman &

Kaiser, 1994; Kirkham, Slemmer, & Johnson, 2002; Meulemans & der Linden, 1997; Reber, 1969). However, this research has been directed predominantly towards establishing the rate at which associations are acquired, rather than whether there are individual differences in cue acquisition. Furthermore, despite the accuracy in demonstrating a universal capacity for the acquisition of implicit associations (Reber, 1969), these measures tended to be applied over an extended period of time which is impractical during the selection process of potential job candidates.

Five approaches to the assessment of cue acquisition were developed that varied by exposure period (either restricted or unrestricted), and saliency of characteristics (either less salient or more salient). Performance on the five tasks was assessed against performance on a novel, rail control task that was intended to reflect cue acquisition at the initial stages of exposure to a novel environment. Of the five tasks, the Timed-Search Task (TST) was the only measure that was associated with performance on the novel rail control task. The attributes of the TST included a restricted period of exposure to the stimuli, and less salient characteristics that comprise the feature-event/object associations.

These attributes are consistent with the attributes of assessment methods that are used to demonstrate a capacity to identify patterns of associations in a novel environment (Reber, 1969; Reber et al., 1980). Consequently, variations in TST scores strongly suggest that while cue acquisition is a capacity that prevails in the majority of individuals, there are individual differences in the rate at which cues are acquired.

The outcomes of Study 1 (Chapter 4) demonstrated that greater scores on the TST were associated with more accurate and faster response latency on the novel rail control task. Therefore, the TST appeared to be a plausible instrument to assess the predictive influence of individual differences in cue acquisition, and one which could be administered over a relatively short period and in the absence of domain-specific experience.

To establish the construct validity of the TST, Study 2 (Chapter 5) was designed to determine the association between cue acquisition and cue utilisation. Based on the proposition that individual differences in cue acquisition underlie the development of cue utilisation across any domain, it was expected that TST scores would be positively associated with cue utilisation after controlling for exposure. Pilot trainees, enrolled in one of three university-administered flight programmes, completed the TST and a general aviation cue utilisation measure, consisting of five tasks administered using the online platform EXPERTise 2.0.

The results indicated that TST scores were only associated with two of the five cue utilisation tasks, namely the Feature Identification Task and Feature Recognition Task. Greater TST scores were associated with faster response latency and greater accuracy in recognising critical features, supporting the proposition that individual differences in cue acquisition involves the ability to identify and recognise meaningful feature-event/object associations in the environment.

Study 3 (Chapter 6) was designed to further establish the construct validity of the TST by testing the relationship between cue acquisition and cue utilisation in the context of driving a motor vehicle. However, the associations between cue acquisition and components of cue utilisation evident in Study 3 (Chapter 6) were not consistent with the observed associations in Study 2 (Chapter 5) with cue acquisition only associated with performance on the Feature Discrimination Task. While this contradictory finding may call into question the validity of the TST and the previously proposed relationship between cue acquisition and cue utilisation, it is important to note that participants in the two studies differed in their experience within their respective domains. Participants in Study 2 (Chapter 5) were general aviation pilot trainees with an average of 112 flight hours, while participants in Study 3 (Chapter 6) had, on average, three years of experience within the domain of driving a motor vehicle. This difference in the extent to which operators were exposed to domain-specific environment at the point of testing may impact the association evident between cue acquisition and cue utilisation.

Consistent with previous research (Watkinson et al., 2018; Wiggins, Griffin, & Brouwers, 2019), the outcomes of Study 3 (Chapter 6) demonstrated that participants with higher cue utilisation recorded fewer driving errors and less variation in eye movement behaviours compared to participants with lower cue utilisation, accounting for individual differences in cue acquisition. These observations suggest that higher cue utilisation is associated with improved driving performance. However, cue acquisition was not associated with driving performance within the driving simulator suggesting that cue acquisition does not predict the performance of experienced motor vehicle drivers.

Studies 4 (Chapter 7) and 5 (Chapter 8) were designed to investigate the predictive influence of individual differences in cue acquisition on domain-specific performance outcomes following exposure to a domain-specific environment. The relationship between cue acquisition and performance was evaluated within the context of general aviation given that the association between cue acquisition and aspects of cue utilisation was demonstrated within the domain of general aviation in Study 2 (Chapter 5).

Consistent with Study 2 (Chapter 5), participants in Study 4 (Chapter 7) were pilot trainees enrolled in a university-administered flight training programme. Flight performance was derived from actual flight data, and operationalised as flight control performance and instructors' subjective ratings of participants' performance during a selected flight lesson, Lesson 11. At the time of completing the TST, participants were in their first, second or third year of flight training.

Accounting for flight experience, no significant association was evident between cue acquisition and flight performance. This suggests that individual differences in cue acquisition are not sensitive to differences in performance following exposure to domain-specific environment.

Extending the outcomes of Study 4 (Chapter 7), Study 5 (Chapter 8) was conducted to introduce greater experimental control over the stimuli to which participants were exposed. The relationship between cue acquisition and flight performance was examined within a controlled, simulated setting. A critical event involving engine failure was introduced as a trigger, and the performance of pilots was compared subsequently.

When faced with complex, high-stress situations, such as an engine failure, operators tend to experience an increase in cognitive load (Reimer & Mehler, 2011; Stokes & Raby, 1989). Consistent with this expectation, participants in Study 5 (Chapter 8) recorded a significant increase in cognitive load following an engine failure during a simulated flight. This was reflected in psychophysiological assessments of blood pressure, heart rate and the level of oxygen consumed. However, the results corroborated the outcomes of Study 4 (Chapter 7) whereby no association was evident between cue acquisition and flight control performance nor landing success, accounting for flight experience. Further, cue acquisition was not associated with measures of cognitive load. Therefore, individual differences in cue acquisition do not appear to be associated with later performance outcomes and cognitive load during task performance, accounting for differences in exposure.

Overall, the findings from the five studies suggest that individuals vary in the rate at which they acquire meaningful feature-event/object associations in a novel environment. This individual difference in cue acquisition is likely to be associated with the identification and recognition components of cue utilisation. However, cue acquisition appears only to be sensitive in predicting task performance at the initial stages of skill development, where individuals have no previous experience with a task. Following initial exposure, cue acquisition, as measured by the TST, appears to lack sensitivity in predicting performance.

Instead, cue utilisation was more predictive of performance where individuals had acquired further skills.

Theoretical Implications

The present programme of research established a number of findings relevant to cue acquisition as a construct, and its relationship to cue utilisation and operational performance. There were three key theoretical contributions, including: (1) the nature of individual differences in cue acquisition, (2) the cognitive aspects of cue acquisition, and (3) the differential sensitivity of cue acquisition in predicting performance.

Individual Differences in Cue Acquisition

Firstly, individual differences in cue acquisition have been observed in the present studies. The capacity to extract meaningful associations between features and events or objects has previously been established as an inherent capability (Reber, 1989). However, the outcomes of the present research appear to demonstrate that there are individual differences in the rate at which cues are acquired.

Cue acquisition, arguably, engages implicit learning, where the operator acquires abstract knowledge through a passive, non-conscious process (Reber, Walkenfeld, & Hernstadt, 1991). The process is presumed to be almost entirely bottom-up, driven by the perceptual stimuli in the environment such that the operators themselves do not intentionally seek to identify and acquire knowledge (Whittlesea & Wright, 1997). Nevertheless, the specific, underlying cognitive mechanisms by which implicit learning operates in this case, remains uncertain.

Propositions of the underlying mechanisms of implicit learning, such as chunking (Meulemans & der Linden, 1997; Servan-Schreiber & Anderson, 1990), are built on the same fundamental assumption. Implicit learning is assumed to function on the basis of individual sensitivity to patterns of associations that exist within situations and the environments that operators encounter.

Reber (1989) argued that implicit learning is evolutionary, whereby individuals' sensitivity to patterns of associations are drawn from the natural world. This sensitivity facilitates the identification of patterns of associations from random noise, minimising conscious and effortful processes to derive relevant associations, which are then integrated into existing schemas in memory. These schemas, in turn, support System 1 processing during decision-making to efficiently and effectively react and respond to problem situations (Kaempf, Klein, Thordsen, & Wolf, 1996).

The adaptive advantage afforded by implicit learning, and which underlies cue acquisition, would necessitate that it is a universal capacity, which supports task performance. While it is evident that sensitivity to patterns of associations is fundamental to cue acquisition and skill development (Reber, 1989), the question remains as to whether there are individual differences in cue acquisition which would signify greater and lesser propensity to acquire relevant cues.

If every individual has a similar sensitivity to patterns of associations in noisy environments, it follows that exposure to similar experiences would result in the acquisition of similar cues, which are then integrated into existing schemas. Subsequently, these schemas should support System 1 processing to promote similar levels of efficiency and effectiveness in subsequent responses.

Previous research has demonstrated that the length of experience in performing a domain-specific task is not, in itself, predictive of task-specific performance (Crane et al., 2018; Hambrick et al., 2014). The observed dissociation between experience and task performance suggests that there may be variations in individual sensitivity to patterns of association. Consistent with this proposition, the outcomes of the present study demonstrated that this capacity is evidently not a binary trait, but rather, is likely to prevail as variations in the effectiveness in acquiring and deriving patterns of associations.

Aspects of Cue Acquisition

The second theoretical contribution associated with the present program of research relates to the nature of cue acquisition. Arguably, individual differences in cue acquisition generate variations in the rates at which relevant cues are acquired from similar experiences. This variation in the availability of relevant cues in memory promotes differences in individuals' cue utilisation, and in turn, leads to variations in situation assessment and task performance. Therefore, individual differences in cue acquisition should relate to one or more aspects of cue utilisation.

The outcomes of Study 2 (Chapter 5) in particular indicated that individual differences in cue acquisition were associated with the identification and recognition aspects of cue utilisation. Performance on the Feature Identification and Feature Recognition measures of cue utilisation are behavioural responses that reflect the capacity to identify and recognise critical cues during situation assessment. Therefore, the association between cue acquisition and identification and recognition characteristics of cue utilisation substantiated the proposition that cue acquisition functions using implicit learning, which constitutes the ability to identify and recognise implicit patterns of associations in noisy environment.

The relationship between cue acquisition and cue utilisation, however, appeared to be influenced by previous domain-specific experience. The outcomes of Study 3 (Chapter 6) demonstrated an association between cue acquisition and the discrimination characteristics of cue utilisation, but not the identification and recognition characteristics. However, participants in Study 2 (Chapter 5) were assessed within the domain of general aviation, whereas participants in Study 3 (Chapter 6) were assessed within the domain of driving a motor vehicle. Consequently, there may be idiosyncratic aspects of each domain, which may have influenced the relationship between cue acquisition and cue utilisation.

Varying Sensitivity in the Predictive Influence of Cue Acquisition on Performance

The third theoretical contribution arising from the present research pertains to the predictive influence of cue acquisition on task performance at different stages of skill development. The outcomes of the present series of studies suggest that individual differences in cue acquisition are more sensitive in predicting performance only at the initial stages of skill acquisition and become less sensitive in predicting performance at the later stages of skill development.

For operators to undertake any skilled performance, such as driving a car or flying an aircraft, they must first be exposed to domain-specific stimuli, including the relevant features, events or objects, and their associations, to allow operators to acquire the relevant feature-event/object associations or cues in memory. The role of this initial period of exposure was established in Study 1 (Chapter 4) by presenting participants with a novel rail task. This initial exposure emulated the synthetic language task administered in previous studies on implicit learning (Buchner, 1994; Reber, 1969).

Under these unfamiliar conditions, cue utilisation, which functions as part of System 1 processing, cannot operate in the absence of relevant feature-event/object associations in memory to support situation assessment. Instead, participants must rely on System 2 processing to analyse and evaluate the novel stimuli to construct assessments of the presenting situations or problems. However, the novel rail task imposed time constraint on task performance, which necessitated the adoption of System 1 processing. While all participants were required to engage System 2 processing initially, a higher capacity for cue acquisition would result in the acquisition of relevant feature-event/object associations more quickly, and therefore, enable participants to engage System 1 processing at an earlier stage of skill acquisition. Consequently, superior performance on the novel rail task should reflect a faster rate of cue acquisition.

The outcomes of Study 1 (Chapter 4) demonstrated that higher cue acquisition was associated with greater accuracy and faster response latency on the novel rail task. This observation suggests that, at the initial stages of exposure, where relevant cues in memory are absent, individual differences in cue acquisition appeared to demonstrate sensitivity in predicting performance on a novel task.

Regardless of the length of initial exposure, the findings from Studies 3 through 5 (Chapters 6 – 8) failed to reveal an association between cue acquisition and task performance within the domains of general aviation and driving a motor vehicle. The consistent observations across all four studies suggest that individual differences in cue acquisition appear to lack sensitivity in predicting task performance following the initial exposure to domain-relevant operations.

Given that participants in Study 4 (Chapter 7) constituted pilot trainees, most of whom were in their initial phases of flight training, the absence of an association between cue acquisition and flight performance amongst the participants suggests that individual differences in cue acquisition are likely to be sensitive in predicting performance at the very early stages of skill development, prior to the acquisition of task-relevant cues. Therefore, where relevant cues are available in memory, individual differences in cue acquisition may lack sensitivity in predicting skilled performance.

The underlying processes that support skilled performance amongst expert operators are described in the Recognition-Primed Decision (RPD) model (Klein et al., 1986; Lipshitz, Klein, Orasanu, & Salas, 2001). The RPD model presupposes that skilled performance is built on effective situation assessment, which is achieved through the application of feature-event/object associations within existing schemas in memory (Klein, 2008). Features in the environment are matched against representations of features in memory to derive associated information that act as assessment templates of the presenting situations.

While the RPD model accounts for the swift and accurate skilled performance observed amongst expert operators, it does not present descriptions pertaining to the underlying feature-event/object associations that support effective situation assessment. In the context of the RPD model, the outcomes from the present series of studies suggests that individual differences in cue acquisition may contribute to differences in performance at the initial stages of skill acquisition. However, following exposure to domain-specific experiences, the availability of relevant feature-event/object associations integrated into domain-specific schemas in memory allow expert operators to effectively utilise cues to support situation assessment, and in turn, skilled performance. However, the point at which different operators achieve similar levels of skilled performance is evidently different given differences in the rate at which relevant feature-event/object associations are acquired to support cue utilisation.

In sum, cue acquisition is likely to constitute a trait on which individuals vary, resulting in differences in the rate at which cues are acquired initially. Individual differences in cue acquisition appeared to be sensitive in predicting task performance during the initial stages of skill development, but not performance outcomes following exposure to domain-specific environment. The apparent changes in the influence of cue acquisition on performance across time and experience remain unclear and represents an avenue for future research.

Limitations and Future Research

While the present programme of research has established a number of theoretical contributions pertaining to the construct of cue acquisition, there are a number of limitations that require consideration when drawing conclusions. The prevailing limitation concerns the cross-sectional nature of the studies.

The conclusions derived from the present studies are reliant on inferences procured across multiple studies. For instance, the conclusion pertaining to the varying sensitivity of cue acquisition in predicting task performance at different stages of skill development was deduced from the results of five independent studies. In this instance, the change in the sensitivity of cue acquisition in predicting task performance across time and exposure was not based on a single participant sample pool. Consequently, there may be distinctive variations in participant characteristics and domain-specific tasks across studies, which may have contributed to the association between cue acquisition and task performance.

To extend and substantiate the conclusions arising from the present studies, a longitudinal study, integrating the key variables identified in the present programme of studies, could be undertaken. When analysing the change in the sensitivity of cue acquisition in predicting task performance, the TST could be administered prior to the participants being introduced into the domain-specific environment. Subsequently, the participants would be tested throughout their training programme, and task performance would be measured at different points.

The associations between individual differences in cue acquisition and performance at different stages of exposure to domain-specific environment should also be examined. This would enable an evaluation of the sensitivity of cue acquisition in predicting task performance at different stages of skill development. Arguably, the findings of the present programme of research support the prediction that the association between cue acquisition and performance should diminish with increased exposure, reflecting a reduction in sensitivity with increases in experience.

To demonstrate a clear trajectory of change in sensitivity, performance on a simple domain-specific task, incorporating patterns of associations, must be assessed prior to exposure to a controlled training programme in a specified domain. This first task would be akin to performance on the novel rail task in Study 1 (Chapter 4) of the present programme of research, where participants' performance was independent of domain-specific experiences. The association between cue acquisition and performance on this novel task would indicate the sensitivity of cue acquisition in predicting learning performance at the initial stages of skill acquisition.

Following the initial stages of skill acquisition, task performance would be measured at fixed intervals to assess changes in the sensitivity of cue acquisition in predicting skilled performance. Performance on these subsequent tasks would correspond to assessments of performance on the various domain-specific tasks in the present Studies 3 through 5. The associations between individual differences in cue acquisition prior to the commencement of the training programme and performance at different stages of skill acquisition can then be examined. The trajectory of change in the sensitivity of cue acquisition in predicting performance with increased experience can then be plotted.

The limitation pertaining to cross-sectional studies also relates to the association between cue acquisition and cue utilisation. The contradictory findings of Study 2 (Chapter 5) and 3 (Chapter 6) may be due, in part, to differences in domain-specific experiences of the two distinct groups of participants. The issue of comparing outcomes of cross-sectional studies, in this instance, relates to a lack of direct, equivalent comparison between the two domains.

Arguably, the process of skill development that promotes skilled performance in different domains may emphasise different aspects of skill acquisition depending upon the nature of the domain-specific tasks. Skilled performance in some domains may not necessarily be heavily reliant on individual differences in cue acquisition, but instead, may be better acquired through other processes, such as repetition of task performance when acquiring motor skills (Ofen-Noy, Dudai, & Karni, 2003; Stefanidis et al., 2006). Therefore, comparing findings from cross-sectional studies may not be analogous due to the range and complexity of tasks, varying by experience, performed in operational settings, some of which may be more dependent on cue utilisation and cue acquisition than others (Ericsson, Whyte, & Ward, 2007). Consequently, it remains unclear whether the relationship between cue acquisition and cue utilisation varies across domains or the contradictory findings of Studies 2 (Chapter 5) and 3 (Chapter 6) are caveats of the differences in the nature of domain-specific tasks.

To ascertain the relationship between cue acquisition and cue utilisation across domains, longitudinal analyses should be undertaken in two distinct domains. The trajectory of change in the association between cue acquisition and aspects of cue utilisation can be tested against increases in experience/exposure. Where the associations between cue acquisition and aspects of cue utilisation are consistent in both domains, the range and complexity of tasks at the given stage of skill development can then be determined. This would allow for a matched comparison between domains with greater control over the nature of domain-specific tasks.

Finally, the TST may be insufficient in assessing cue acquisition. While the findings of Study 1 (Chapter 4) demonstrated the construct validity of the TST against performance in a novel task, the findings of Studies 2 (Chapter 5) and 3 (Chapter 6) pertaining to the relationship between TST scores and cue utilisation appeared to be inconsistent. In combination, these outcomes suggest that the TST may be suitable in capturing the aspect of cue acquisition pertaining to performance in a novel task, but not aspects relevant to the development of relevant feature-event/object associations that underlie cue utilisation.

In addition to conducting a longitudinal study comparing between domains, future research may further develop the TST into a battery of tasks that would better capture the different aspects of cue acquisition relevant to the development of appropriate feature-event/object associations in supporting cue utilisation. The four tasks developed alongside the TST in Study 1 (Chapter 4), including the Search Task, Explicit Association-Learning Task, Problem Solving Task and Categorisation Task, might be amended to incorporate both restricted period of exposure and reduced saliency of characteristics to reflect the nature of cue acquisition. These tasks can then be validated against performance in a novel task and cue utilisation in varying domains to establish a battery of tasks that may be relevant to cue utilisation.

Practical Implications

The present programme of studies was designed to evaluate the predictive influence of individual differences in cue acquisition on skilled performance. The practical goal was to examine whether individual differences in cue acquisition constitute a useful variable during the selection process. When selecting for a role where previous domain-specific experience may not be readily available, such as in the police force, military, and firefighting, cue acquisition, as a domain-general variable, might be utilised as part of the battery of measures in identifying candidates who may acquire task-related skills at a relatively faster rate, and perform at a higher level during a shorter period. Therefore, the conclusions pertaining to the association between individual differences in cue acquisition and task performance can serve to inform the design of both selection procedures and training schemes.

The outcomes from the present programme of studies suggest that individual differences in cue acquisition were positively associated with task performance at the initial stages of skill acquisition, and two characteristics of cue utilisation. Based on these observations, it might be proposed that the measure of cue acquisition could eventually be a useful instrument to identify candidates who are more likely to learn faster at the initial stages of exposure to a domain-specific environment. In acquiring relevant cues at a faster rate, operators are able to engage cue utilisation to support situation assessment, and subsequently, achieve skilled performance at an earlier stage in the process of skill development.

A faster rate of skill acquisition would benefit the organisation given that operators would achieve skilled performance more quickly, thereby reducing training costs. Achieving appropriate or expected standards of skilled performance at a faster rate translates into reduced costs associated with less effective performance, including human errors and less efficient productivity, and improved safety outcomes.

Conclusion

In high-risk domains, such as aviation, firefighting and military, effective situation assessment is crucial for the selection of appropriate action plans and the swift execution of responses to problems. The construction of swift and accurate assessments is supported by higher cue utilisation, which involves the application of cues in memory to derive templates on which assessments are based (Wiggins, 2012).

While cue utilisation constitutes a reliable and predictive measure of domain-specific task performance (Loveday, Wiggins, Searle, et al., 2012; McCormack, Wiggins, Loveday, & Festa, 2014; Watkinson et al., 2018), it is a process that relies on the availability of domain-specific cues in memory. Consequently, the measure of cue utilisation cannot provide meaningful predictions of task performance when selecting candidates for jobs where previous domain-specific experiences may not be readily available. In this instance, a domain-general measure of cue acquisition may be more appropriate.

The construct of cue acquisition was posited as an underlying trait that pertains to individual differences in the acquisition of relevant feature-event/object associations. These associations are integrated into existing schemas in memory to be utilised as cues during situation assessment. A measure of cue acquisition was designed and evaluated in Study 1 (Chapter 4).

The results indicated that performance on the Timed-Search Task (TST) was associated with greater accuracy and faster response latency on a novel rail control task, which incorporated implicit associations. This observation established TST as a measure that corresponded with individuals' ability to derive implicitly embedded patterns of associations in complex novel stimuli. Further, the findings suggest that individual differences in cue acquisition are sensitive in predicting task performance where individuals have no previous experience.

Study 2 (Chapter 5) extended Study 1 (Chapter 4) in establishing the construct validity of the TST by examining the association between cue acquisition and cue utilisation. In the context of general aviation, individual differences in cue acquisition were associated with the identification and recognition aspects of cue utilisation. The findings support the proposition that cue acquisition involves the ability to identify and recognise meaningful associations in complex stimuli.

In Study 3 (Chapter 6), the predictive influence of cue acquisition on domain-specific task performance was evaluated as a covariate to the relationship between cue utilisation and performance. In the domain of driving a motor vehicle, individual differences in cue acquisition were not associated with driving errors and visual search behaviours. However, higher cue utilisation was associated with fewer driving errors, fixations and saccades compared to lower cue utilisation. The outcomes suggest that individual differences in cue acquisition are not sensitive in predicting task performance following the initial acquisition of relevant cues.

Studies 4 (Chapter 7) and 5 (Chapter 8) were designed to examine the association between cue acquisition and domain-specific task performance in the context of general aviation. Study 4 (Chapter 7) assessed flight performance based on actual flight data from a selected flight lesson, whereas Study 5 (Chapter 8) assessed flight performance within a flight simulator. The outcomes of both Studies 4 (Chapter 7) and 5 (Chapter 8) failed to demonstrate an association between cue acquisition and flight control performance.

In combination, the present programme of research constituted a preliminary assessment of cue acquisition, and the findings resulted in three broad theoretical contributions. These contributions included: (1) evidence to support the role of individual differences in cue acquisition, (2) evidence to support the nature of cue acquisition as a cognitive process of identifying and recognising patterns of association, and (3) evidence to support the differential sensitivity of cue acquisition in predicting task performance at different stages of exposure.

References

Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences, 8*(1), 17-34. doi:10.1080/02640419008732128

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11*(2), 373-379. doi:10.3758/bf03196585

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics, 60*(11), 1500-1515. doi:10.1080/00140139.2017.1330494

Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of experimental psychology: learning, 20*(3), 550-566. doi:10.1037/0278-7393.20.3.550

Crane, M. F., Brouwers, S., Wiggins, M. W., Loveday, T., Forrest, K., Tan, S. G. M., & Cyna, A. M. (2018). “Experience isn't everything”: How emotion affects the relationship between experience and cue utilization. *Human Factors, 60*(5), 685–698. doi:10.1177/0018720818765800

Engström, I., Gregersen, N. P., Hernetkoski, K., Keskinen, E., & Nyberg, A. (2003). *Young novice drivers, driver education and training: Literature review*. Linköping, Sweden: Swedish National Road and Transport Research Institute.

Ericsson, K. A., Whyte, J. I. V., & Ward, P. (2007). Expert performance in nursing. *Advances in Nursing Science, 30*(1), E58-E71. doi:10.1097/00012272-200701000-00014

Goh, J., & Wiegmann, D. A. (2001a). An investigation of the factors that contribute to pilots' decisions to continue visual flight rules flight into adverse weather. *Proceedings of* *the Human Factors and Ergonomics Society Annual Meeting, 45*(2), 26-29. doi:10.1177/154193120104500205

Goh, J., & Wiegmann, D. A. (2001b). Visual flight rules flight into instrument meteorological conditions: An empirical investigation of the possible causes. *The International Journal of Aviation Psychology, 11*(4), 359-379. doi:10.1207/s15327108ijap1104_3

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45*, 34-45. doi:10.1016/j.intell.2013.04.001

Horrey, W. J., & Wickens, C. D. (2001). Supporting situation assessment through attention guidance: A cost-benefit and depth of processing analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(4), 316-320. Rdoi:10.1177/154193120104500411

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 220-231. doi:10.1518/001872096779047986

Kellman, P. J., & Kaiser, M. K. (1994). Perceptual learning modules in flight training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(18), 1183-1187. doi:10.1177/154193129403801808

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145-149. doi:10.1016/j.neulet.2009.06.030

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*(2). doi:10.1016/S0010-0277(02)00004-5

Klein, G. A. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456-460. doi:10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society Annual Meeting, 30*(6), 576-580. doi:10.1177/154193128603000616

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331-352. doi:10.1002/bdm.381

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2012). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 90-107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2012). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 125-137. doi:10.1177/0018720812448475

McCormack, C., Wiggins, M. W., Loveday, T., & Festa, M. (2014). Expert and competent non-expert visual cues during simulated diagnosis in intensive care. *Frontiers in Psychology, 5*, 949. doi:10.3389/fpsyg.2014.00949

Meulemans, T., & der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 1007. doi:10.1037/0278-7393.23.4.1007

Mueller, A. S., & Trick, L. M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis & Prevention, 48*, 472-479. doi:10.1016/j.aap.2012.03.003

Ofen-Noy, N., Dudai, Y., & Karni, A. (2003). Skill learning in mirror reading: how repetition determines acquisition. *Cognitive Brain Research, 17*(2), 507-521. doi:10.1016/S0926-6410(03)00166-6

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*(1), 115. doi:10.1037/h0027454

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-235. doi:10.1037/0096-3445.118.3.219

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 492-502. doi:0096-I515/80/0605-0492S00.75

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(5), 888-896. doi:10.1037/0278-7393.17.5.888

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics, 54*(10), 932-942. doi:10.1080/00140139.2011.604431

Renshaw, P. F., & Wiggins, M. W. (2017). The predictive utility of cue utilization and spatial aptitude in small Visual Line-Of-Sight rotary-wing Remotely Piloted Aircraft operations. *International Journal of Industrial Ergonomics, 61*, 47-61. doi:10.1016/j.ergon.2017.05.014

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 592-608. doi:10.1037/0278-7393.16.4.592

Stefanidis, D., Korndorffer, J. R., Black, W. F., Dunne, B. J., Sierra, R., Touchard, C. L., . . . Scott, D. J. (2006). Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery, 140*(2), 252-262. doi:10.1016/j.surg.2006.04.002

Stokes, A. F., & Raby, M. (1989). Stress and cognitive performance in trainee pilots. *Proceedings of the Human Factors Society 33rd Annual Meeting, Vol 2, 33*(14), 883-887. doi:10.1177/154193128903301404

Todd, M. A., & Thomas, M. J. W. (2012). Flight hours and flight crew performance in commercial aviation. *Aviation, Space, and Environmental Medicine, 83*(8), 776. doi:10.3357/ASEM.3271.2012

Vansteenkiste, P., Vaeyens, R., Zeuwts, L., Philippaerts, R., & Lenoir, M. (2014). Cue usage in volleyball: a time course comparison of elite, intermediate and novice female players. *Biology of Sport, 31*(4), 295-302. doi:10.5604/20831862.1127288

Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology*, 1-7. doi:10.1080/14992027.2018.1476782

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(1), 181-200. doi:10.1037/0278-7393.23.1.181

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human factors, 44*(2), 189-197. doi:10.1518/0018720024497871

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science, 15*(3), 283-292. doi:10.1080/1463922X.2012.724725

Wiggins, M. W., Griffin, B., & Brouwers, S. (2019). The potential role of context-related exposure in explaining differences in water safety cue utilization. *Human Factors*, 1-14. doi:10.1177/0018720818814299

Appendix A-D of this thesis have been removed as they may contain sensitive/confidential content