



Assessing the learning and transfer of gaze behaviours in immersive virtual reality

David J. Harris¹ · Kyle J. Hardcastle² · Mark R. Wilson¹ · Samuel J. Vine¹

Received: 3 July 2020 / Accepted: 13 January 2021
© The Author(s) 2021

Abstract

Virtual reality (VR) has clear potential for improving simulation training in many industries. Yet, methods for testing the fidelity, validity and training efficacy of VR environments are, in general, lagging behind their adoption. There is limited understanding of how readily skills learned in VR will transfer, and what features of training design will facilitate effective transfer. Two potentially important elements are the *psychological fidelity* of the environment, and the *stimulus correspondence* with the transfer context. In this study, we examined the effectiveness of VR for training police room searching procedures, and assessed the corresponding development of perceptual-cognitive skill through eye-tracking indices of search efficiency. Participants ($n = 54$) were assigned to a VR rule-learning and search training task (FTG), a search only training task (SG) or a no-practice control group (CG). Both FTG and SG developed more efficient search behaviours during the training task, as indexed by increases in saccade size and reductions in search rate. The FTG performed marginally better than the CG on a novel VR transfer test, but no better than the SG. More efficient gaze behaviours learned during training were not, however, evident during the transfer test. These findings demonstrate how VR can be used to develop perceptual-cognitive skills, but also highlight the challenges of achieving transfer of training.

Keywords Fidelity · Police · Policing · Training · Validity · VR

1 Introduction

Virtual reality (VR) has the potential to revolutionise training in a number of high-pressure and safety-critical environments. VR training of perceptual-cognitive¹ and motor skills—such as in sport (Bird 2019), surgery (Moglia et al. 2016), aviation (Oberhauser and Dreyer 2017) or defence and security (Lele 2013)—has received particular interest.

Ultimately, the success of any form of simulation training is determined by whether practice in the simulation transfers to improvements on a corresponding real-world task (Barnett and Ceci 2002; Gray 2019). While there is wider debate in Psychology about how readily skills learned in one context can be transferred to new tasks (Barnett and Ceci 2002; Sala and Gobet 2017), the ecological validity of VR (Parsons 2015) suggests that the potential for transfer of learning should be high. Yet, in order to train perceptual-cognitive and motor skills in VR, it is necessary to establish that these kinds of skills can actually be developed in VR, and whether they will transfer between contexts, as the unusual perceptual characteristics of virtual environments may affect how skills are learned (Harris et al. 2019a, 2020b). Few studies have addressed this issue by assessing changes in perceptual-cognitive abilities during virtual training (Gray 2019; Tirp et al. 2015). In the present study we aimed to explore the development of perceptual-cognitive skill and transfer of training in VR in the context of police room searching skills.

✉ David J. Harris
D.J.Harris@exeter.ac.uk

Kyle J. Hardcastle
Kyle.J.Hardcastle@met.police.uk

Mark R. Wilson
Mark.Wilson@exeter.ac.uk

Samuel J. Vine
S.J.Vine@exeter.ac.uk

¹ School of Sport and Health Sciences, University of Exeter, St Luke's Campus, Exeter EX1 2LU, UK

² Counter Terrorism Protective Security Operations, Metropolitan Police Service, Lambeth HQ, London SE1 7LP, UK

¹ Perceptual-cognitive skills include use of visual attention, anticipation and decision-making (Marteniuk 1976).

Effective simulations, for skills training or psychological experimentation purposes, must have the requisite levels of *fidelity* to create realistic behaviour (Gray 2019; Harris et al. 2020a; Slater 2009). Fidelity describes the extent to which the simulation recreates elements of the real task, such as the visual, auditory or haptic information available. Ultimately, if the perceptual input available to the user is the same as (or functionally indistinguishable from) the real skill, then transfer of training is more likely. Sub-types of fidelity include physical (visual and auditory realism), psychological (realistic perceptual and cognitive demands) and affective fidelity (realistic emotional responses). In reviewing the use of VR for training perceptual-cognitive skills—abilities such as attention, anticipation and prediction—Gray (2019) identifies *psychological fidelity* as particularly relevant for transfer of training. Psychological fidelity is the extent to which the VR environment demands the same perceptual information for controlling actions, elicits the same gaze behaviour and places the same cognitive demands on the user as the real environment. Despite evidence that VR training of perceptual-motor skills can transfer to real-world skills (Gray 2017; Michalski et al. 2019), there has been limited examination of how perceptual-cognitive skills develop in VR. In particular, there has (to our knowledge) been no study examining the training and transfer of visual search behaviour within a virtual environment. One study that has examined perceptual-cognitive skill learning in VR, by Tirp et al. (2015), found a development of the gaze behaviour *quiet eye* (Vickers 2007) to occur during VR dart throwing practice, as well as transfer to improved performance in real dart throwing. This finding is encouraging, but if VR training applications cannot create a high level of psychological fidelity, then transfer of training will be impaired by the lack of perceptual-cognitive skill development (Gray 2019; Harris et al. 2019b).

A number of wider theoretical debates exist in psychology about how readily new learning can transfer (Rosalie and Müller 2012). Transfer is said to occur when prior experiences in one context can be adapted to similar or dissimilar contexts (Barnett and Ceci 2002); yet real transfer can be hard to achieve. For instance, the extensive cognitive training (or ‘brain training’) literature has shown that while cognitive skills can be developed through practice, performance improvements on untrained tasks generally only extend to those that are closely matched to the training task (Lintern and Boot 2019; Sala and Gobet 2017). In order for transfer to occur, there may need to be a correspondence of certain key features between the training task and the target task (Barnett and Ceci 2002; Braun et al. 2010). Even though VR enables learning environments that, superficially, closely resemble the real task, it is unclear exactly which features of a learning environment will enable effective transfer. A recent theoretical development—the modified perceptual training framework (Hadlow et al. 2018)—aims to describe

the conditions of training that will enable the transfer of perceptual-cognitive skills. A theoretically driven approach, such as this, could enable better design and development of VR learning environments, but requires empirical testing.

The modified perceptual training (MPT) framework outlines how training tools for perceptual-cognitive skills, such as computer-based apps, sports vision training and VR training, are more likely to transfer between contexts when they satisfy three conditions. Firstly, positive transfer is more likely when higher-order, task-specific perceptual-cognitive skills (e.g. visual search) are targeted, as opposed to low-level visual skills (e.g. visual acuity). Secondly, the mode of response should match the natural execution of the skill; hence, training with a button press response is only likely to transfer to other button pressing tasks (i.e. *response correspondence*). VR supports response correspondence in many ways by enabling realistic movements, but the need for hand-held controllers does place limits on the realism of some actions. Finally, training is likely to be more effective when the training stimuli match those of the test stimuli (i.e. *stimulus correspondence*). For instance, learning to track geometric shapes will be less beneficial than learning to track moving people when transfer to sport is the goal. In the present work, we aimed to examine the predictions of the MPT framework and assess the importance of stimulus correspondence in the transfer of perceptual-cognitive skills.

In the present study, we examined the development of a perceptual-cognitive skill (visual search) and transfer of training in the context of police room searching. Officers searching a house must find and identify evidence pertinent to the current investigation as well as be attentive to indicators of other criminal activity. Currently, non-specialist officers are instructed how to recognise indicators of terrorist-related activity using classroom-based exercises. Officers are taught the kind of items that might be indicators of terrorist activity in a mostly didactic manner and then further discuss what to look for. They are also instructed how to conduct a thorough search of the room, but get only limited physical practice because of practical constraints on training. While this kind of approach promotes explicit learning about early indicators of terrorism, it does not allow the more active development of procedural knowledge and perceptual-cognitive skill that can be achieved through realistic searching practice (Freeman et al. 2014; Trninic 2018). VR affords the possibility to practice skills like searching in a more ecologically valid way, when real-world practice is not possible/practical. An example of the application of VR simulation to policing is a training device known as AUGGMED (Automated Serious Game Scenario Generator for Mixed Reality Training). AUGGMED allows police officers and other first responders to practice decision making under pressure during a multi-user VR simulation of a terrorist attack. A comparison study between live exercises and AUGGMED

training has suggested similar learning benefits between live and VR training (Saunders et al. 2019). Following either live exercises or AUGGMED training, police officers were assessed on their knowledge of the terrorist scenario using multiple-choice questions. However, no transfer of learning was assessed, nor were actual decision making skills or perceptual-cognitive skills, like visual search.

The ability to search for a target item amongst other distracting stimuli has been extensively studied in laboratory-based, computerised tests (e.g. see Wolfe 2010). Typically, these tasks take the form of searching for a simple shape hidden within an array of similar shapes: a letter ‘T’ hidden amongst letter ‘L’s. This work has shown that visual search ability can be trained, as participants improve their speed and accuracy over repeated trials, and that expert searchers (e.g. doctors inspecting a scan for a tumour) show advanced search behaviours (Eckstein 2011; Savelsbergh et al. 2002; Wood et al. 2014). Police searches rely on similar principles, such as identifying targets amongst distractions, but are rather more complex and ambiguous as it is not always clear what a target might be. The searches also have to be carried out over a larger area, a whole house rather than just a screen or image. Over extended practice, searchers typically show changes in eye movements that serve to make their search more efficient, such as making a larger initial saccade after trial onset or making fewer fixations but of longer duration (Najemnik and Geisler 2005). While studies have examined visual search in real-world scenes (Torralba et al. 2006), there has been little study of search behaviours in immersive virtual environments. More effective ways to train search skills in an ecologically valid way, such as in VR, would be beneficial, not just for policing, but other settings like airport security or radiology (Biggs et al. 2018).

1.1 The present study

We developed a simple search training task aimed at developing the perceptual-cognitive skill of visual search: the type of tool that could potentially be used to train police officers for house searching. In order to gain a greater understanding of perceptual-cognitive skill learning and transfer in VR, we aimed to test:

1. If search performance and perceptual-cognitive skill could be developed in VR (i.e. establish psychological fidelity of the training task)
2. If performance and perceptual-cognitive skill could transfer between contexts, from a simple to a more complex search task
3. If transfer was more likely when there was greater stimulus correspondence, as predicted by the MPT framework (Hadlow et al. 2018), and when knowledge had been acquired through more active trial-and-error learning.

During the novel training task, the participant is required to search through a number of household items placed on a table, then find and select the target as quickly as possible. The task was designed to develop search skills in an immersive, but relatively simple, task. We assessed whether, through repeated practice, participants learned more efficient visual search behaviours, and whether perceptual-cognitive skill development could transfer to a more complex searching task (moving around and searching a whole room). To examine the importance of stimulus correspondence, as outlined in the MPT framework, one training group were provided with stimuli that were identical across training and transfer contexts, while one group used generic stimuli that differed from the transfer context. In order to focus on the mechanisms of transfer, we aimed to assess a modest level of transfer, from a simple task (searching a small space with no locomotion) to a more complex one (searching a large space requiring locomotion with objects in unknown locations), but still in the medium of VR. Clearly, VR training applications need to transfer to real-world skills, but here we aimed to examine transfer between similar VR tasks to enable more experimental control and an understanding of perceptual-cognitive skills that can inform effective real-world interventions.

Hypotheses It was predicted that actively practicing visual search in VR would lead to a development of visual search skills (changes in eye movement behaviours) as well as positive transfer to a more complex and realistic VR task. Based on the predictions of the MPT framework, it was predicted that transfer to the more complex scenario would be enhanced when stimulus correspondence was greater and learning was more active (through trial and error). Specifically, it was hypothesised that:

- H₁ Participants trained in the full training group (FTG, i.e. search practice with stimulus correspondence) will show better performance on both the training task and a novel transfer task, when compared to search practice without stimulus correspondence (search group; SG) or explicit instruction only (control group; CG). Participants trained on generic search (SG) will in turn show better search performance than those given explicit instruction only (CG).
- H₂ Participants given VR searching practice (FTG and SG) will show a greater development of visual search skill, measured through eye movement metrics (e.g. search rate, fixation duration, entropy), when compared to controls (CG).
- H₃ Visual search skills learned with high stimulus correspondence (FTG) will transfer more readily to the more

Table 1 Demographic information (mean and standard deviation)

	Mean age (years)	% Male	Previous VR use (%)	Video game experience (max 5)	3D vision (arc)
Training group	22.4 (3.6)	28	76	3.3 (1.5)	79.1 (65.5)
Search group	23.6 (4.7)	44	72	4.0 (1.0)	57.4 (41.1)
Control group	21.7 (2.3)	29	73	3.9 (0.9)	93.7 (103.4)

complex VR search task, than those with low stimulus correspondence (SG).

2 Methods

2.1 Participants

In total, 54 participants (33 female, mean age = 22.6 years, SD = 3.9) were recruited via convenience sampling from an undergraduate student population who had no prior police training (see Table 1 for demographics). Participants were randomly allocated to a training group using computerised randomisation, which resulted in 18 participants per group. Sample size estimation was calculated using the ‘SIMR’ package for R (Green and MacLeod 2016). As time to completion on the transfer task was the main effect of interest, the experiment was powered to find a smallest effect of interest ($a \sim 15$ -s difference in completion time). Monte Carlo simulations ($n = 1000$) of a series of linear mixed-effects models with participant as a random factor (and $\beta = 15.0$) were run under scenarios of increasing sample size using SIMR to generate a power curve. Given 10 trials per participant, 95% power was reached for a sample size of 50. (The R code and the power curve for the analysis is available in the supplementary materials: <https://osf.io/5n9j3/>.)

Participants were provided with details of the study and gave written consent before testing began. Ethical approval was obtained from the university departmental ethics committee prior to data collection. The only exclusion criterion was severe vision problems which would prevent participants from completing the task (but spectacles would still be worn within the headset). To ensure homogeneity across the training groups, a number of baseline demographics were collected (Table 1). In addition to age and gender, participants were asked whether they had used VR before (Yes/No), how frequently they played video games (ranging from 1 [very often] to 5 [never]) and had their vision checked for stereoacuity (Randot test).

2.2 Design

The training intervention (FTG) was compared to a passive control group (CG) and a training group that received search practice, but with generic stimuli (SG). The SG

served as an active control for testing the effect of stimulus correspondence, but still received the active search practice. This design was intended to allow conclusions about the mechanism of effect, i.e. whether any benefit for the FTG was just a result of generic searching practice, or if searching practice was more effective when relevant objects were used (i.e. stimulus correspondence) and task-specific knowledge could be developed. The relevant causal contrast for the training intervention is training as usual (Karlsson and Bergmark 2015), which in this case is classroom-based instruction. The type of explicit instruction received by police officers in the classroom (i.e. what to search for) was replicated in the passive control group (CG) and the search only group (SG), whereas the FTG learned the items to search for through trial and error. At the end of training, the experimenter checked that all participants in the FTG had identified the correct search items, to ensure that the only difference at post-test was the manner in which the knowledge was acquired (explicit instruction or trial and error) and the nature of the search practice they had received (stimulus specificity). As a result, a mixed design was used, with group (FTG, SG, CG) as a between-subject factor and test (pre- versus post-intervention) as a within-subject factor. Outcome measures were performance on the training task and two transfer tasks, plus visual search behaviours.

2.3 Tasks and materials

2.3.1 Virtual reality tasks

The virtual environment was programmed using the gaming engine Unity (2018.2.10; Unity technologies, CA) and C#. The VR task was presented using the HTC-Vive (HTC Inc., Taoyuan City, Taiwan), a 6 degree-of-freedom VR-system that consists of one head-mounted display and two hand-held controllers. The headset was run on an HP EliteDesk PC with an i7 processor and GeForce GTX Titan V graphics card (NVIDIA Inc., Santa Clara, CA). Gaze behaviour in VR was recorded using the built-in Tobii eye tracker in the HTC Vive. The eye tracker uses binocular dark pupil tracking and samples at 120 Hz across the full 110° field of view, to an accuracy of 0.5°. Eye tracking data were accessed using the Tobii Pro software development kit.



Fig. 1 The full training task (**a**, left panel) and active control (**b**, right panel). The participant must search for and select the target items as quickly as possible and then end the trial by selecting the button on the wall

2.3.2 Full training task

The full training task (Fig. 1a) was based on traditional tests of visual search (Eckstein 2011; Wolfe 2010) but adapted into a realistic search activity. FTG participants were placed in a simulation of a small basement style room (see Fig. 1). In the room was a large table on which were between 9 and 14 household items, such as a can of beans, a cup, a mobile phone and a roll of tape. On each iteration of the task (i.e. each trial), the items were randomly selected and spawned at set locations on the table. Participants were required to locate the target item(s) and select them by making a visual fixation on the desired item while simultaneously squeezing the trigger on the Vive controller handle. Participants could select as many items as they wished on each trial. To end each trial, the participant was required to select a button on the wall in front of them using the same method. When the trial ended, participants were presented with on-screen feedback about the duration of the trial, how many of the selected items were incorrect, and how many targets they missed.

Five of the 18 total search items were correct targets and 13 were distracters, but the identity of these were initially unknown by the participant who had to deduce them through trial and error. On each trial, the number of targets to place in the scene was selected pseudo-randomly from 33 counterbalanced combinations that ensured all possible combinations appeared (so all participants received a similar ‘training dose’). Nine of the possible 13 distracters were also randomly spawned on and placed randomly across the set locations.

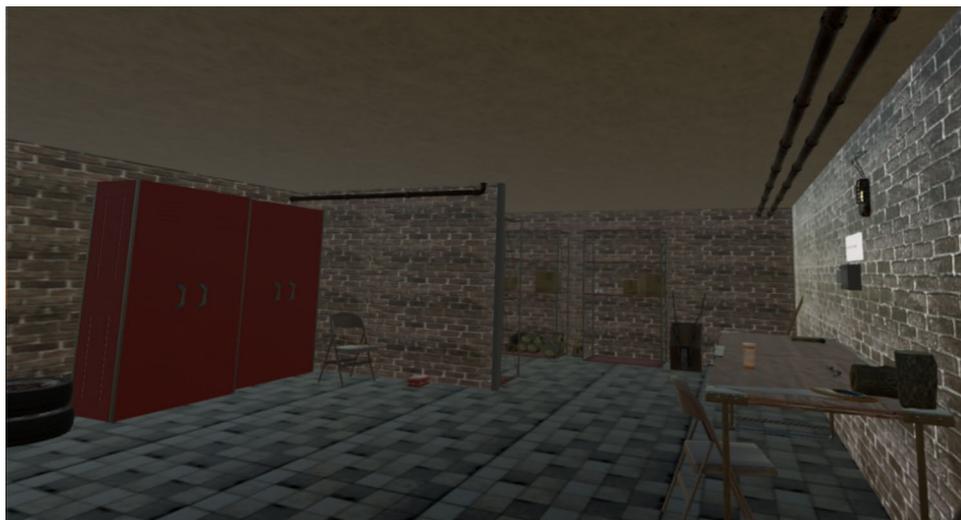
2.3.3 Search only training task

The active control task (Fig. 1b) was designed to provide similar searching practice as the full training tool, but without the use of specific items. Hence, the search only task replaced the household items with cubes of three different sizes and three different materials. Participants had to find and select only the metal boxes. This task placed similar demands on visual search, but without the stimulus correspondence present in the FTG, or any requirement to learn the correct items. Participants received on-screen feedback after each trial in the same way as in the full training task.

2.3.4 2D visual search transfer task

To test whether the VR search training transferred to improvements on a traditional test of visual search, participants also completed a computerised visual search task (Eckstein 2011; Wolfe 2010). This 2D visual search task required participants to identify a target letter ‘T’ in orange colour and regular upright position, presented amongst other letter ‘T’s of incorrect orientation or colour. Participants were required to press the space bar if the target object appeared, or to withhold if the target object was not present. There were 5, 10, 15 or 20 items in each display array, and participants completed 50 arrays (lasting approximately 5 min). Performance was indexed through number of mistakes (failure to find the target or failure to withhold the response) and response time. The visual search task was programmed and completed online using psytoolkit.org (Stoet 2010, 2017).

Fig. 2 The transfer task. Participants were required to search for items located around the garage. The potential locations of the items were unknown (unlike the training task), hence the participant had to complete a thorough search of the room to ensure no items were missed



2.3.5 Transfer task

The transfer task required a much more active search of a large room, more akin to a real police house search. The transfer task (Fig. 2) placed the household items at varying locations around a room so that the participant had to actively move around the room to search for items. Participants navigated the room using a combination of the teleport function and their own locomotion using room-scale VR. Crucially, while all items in the training task were in a known location (on the table), in the transfer task they could be anywhere in the room. Hence, participants had to perform an exhaustive search of the room, making the task more complex and more similar to a real police search. Participants were instructed to be thorough but swift ('Please try to perform a thorough search, but do so as quickly as you can'), and chose to end each trial when they had conducted a sufficiently thorough search, as real officers would. Items were not placed in the most obvious locations but were not actively obscured behind other objects. Videos of all three tasks are available online (<https://osf.io/5n9j3/>).

2.4 Measures

2.4.1 The Simulation Task Load Index (SIM-TLX)

To assess the level of cognitive demand experienced during the full training task, participants completed the SIM-TLX (Harris et al. 2019b). The SIM-TLX is an adaptation of the NASA-TLX (Hart and Staveland 1988) for measuring workload during VR simulations. Following completion of the baseline trials on the training task, participants rated the level of demand they had experienced using nine bipolar workload scales: mental demands; physical demands; temporal demands; frustration; task complexity; situational

stress; distractions; perceptual strain; and task control. For this study, we only used the ratings from the 21-point Likert scale, and not the pair-wised comparisons.

2.4.2 Performance

Performance on the training task, search only training task and transfer test was assessed based on the number of errors (combined score of incorrectly selected items and missed items) and the time taken to complete the trial.

2.4.3 Eye tracking metrics

Gaze data (a timestamp and x, y, z coordinates of point of gaze) were recorded into a csv file which was analysed using an automated approach in MATLAB. Gaze coordinates were converted into fixations using the EYEMMV toolbox for MATLAB (Krassanakis et al. 2014), which identifies fixations using a spatial dispersion algorithm. Fixation parameters were set to a minimum duration criterion of 100 ms and spatial dispersion of 1° of visual angle (as recommended in Salvucci and Goldberg 2000). The search metrics were subsequently derived from this time-stamped record of fixations.

Saccade size. Previous examinations of visual search have found that participants develop more efficient visual search behaviours as a result of practice. One such improvement is an increase in the size of saccades. As expertise is acquired, participants learn to more readily identify targets using peripheral vision and become more proficient at making saccades which are larger and land closer to the target (Eckstein 2011; Mannan et al. 2010). Saccade size was identified from the distance between successive fixation points.

Search rate. Search rate represents the way in which information is extracted from the environment. It is calculated

from the number of fixations made, divided by the mean fixation duration (Murray and Janelle 2003). Search rate has been identified as a characteristic of expertise in tasks like sport (Mann et al. 2007; Savelsbergh et al. 2002) and medical image reading (Gegenfurtner et al. 2019). Search behaviour that uses fewer fixations of longer duration is indicative of a knowledge-driven, top-down, less random search strategy (Gegenfurtner et al. 2019). Search rate was calculated relative to time ((search rate/time)*60) to account for differences in trial duration (Bright et al. 2014; Vine et al. 2014).

Entropy. In general terms, entropy refers to the uncertainty within a system. When applied to gaze behaviour, it can indicate variability or unpredictability of gaze location (Allsop and Gray 2014; Moore et al. 2019; Vine et al. 2015). We adopted a simple measure of entropy (Shannon entropy) to indicate the variability with which gaze was spread across the visual workspace (the visual scene was split into 16 areas of interest and the probability of fixating each was calculated). Shannon entropy (Shannon 1948) expresses the information contained within a probability distribution in ‘bits’. It is calculated from the state space of the system (all possible outcomes) and the relative probabilities of all elements in that state space. For instance, if gaze location was random and distributed fairly evenly across all possible locations, entropy would be high. If it was targeted to particular locations in a predictable manner, entropy would be low. The entropy value was calculated as the sum of the logarithm of all probabilities in the given state space, $H_{(x)} = - \sum_{i=1}^n P(x_i) \log_b P(x_i)$ (as in Shannon 1948). The probability of fixating each location was calculated for each trial, before applying the above formula to those probabilities.

2.5 Procedure

Participants attended the laboratory on two occasions lasting up to an hour each. On the first visit, they completed the informed consent form and a stereoacuity vision screening test (Stereo Optical Inc., Chicago, IL, United States). Participants then completed the computerised 2D visual search task, followed by a baseline assessment on the full training task (one familiarisation trial plus 10 test trials). Participants were then randomised to one of the three training conditions. Participants in the FTG and SG then completed three blocks of 33 trials on their assigned training. Participants returned for the second visit after a 3–4-day interval. The two training groups (FTG and SG) then completed a further three blocks of 33 trials on their assigned training (a total of 198 trials). The training volume was based on previous examinations of perceptual learning in visual search (Sireteanu and Rettenbach 1995). All three groups then completed a post-test (10 trials) of the full training task as a manipulation check and

repeated the computerised 2D visual search transfer task. Finally, participants completed 10 repetitions on the novel transfer task (see Fig. 3). All participants were instructed to find the target items by conducting a thorough but fast search. The CG and SG received explicit instruction about which items to search for in both the post-test training task and the transfer test, but the FTG did not require this instruction as they had acquired the same knowledge through trial-and-error practice with the test items. Pilot testing indicated that most FTG participants had deduced the correct targets half-way through their 6 practice blocks.

2.6 Data analysis

Data analysis was performed in RStudio v1.0.143 (R Core Team 2017). Four participants did not return for the second visit (two in the FTG and two in the SG), but were not excluded from the analysis as their baseline data could still be informative for baseline estimates within the mixed-effects model. Outlying data points (individual trials) more than 3 standard deviations from the mean (Tabachnick and Fidell 1996) were Winsorized by replacing the extreme raw score with the next most extreme score (Signorell et al. 2018). Data were also checked for skewness and kurtosis. Completion time for the training task and completion time and errors for the transfer task were found to be positively skewed so a transform was applied for analyses. Linear mixed-effects models were run to examine the effect of training group (FTG, SG, CG) and trial (pre vs post) on performance and eye tracking measures using the lme4 package for R (Bates et al. 2014). In order to determine the best fitting model in each instance, we initially fitted a near maximal model (Barr et al. 2013)—with random effects for group and/or test within the random factor of participant—and then simplified the random effects structure using principal components analysis, as described by Bates et al. (2014). The best fitting model in each instance was chosen by simplifying the structure in line with the number of principal components that contribute to explaining additional variance. Successive models were compared using the Akaike information criterion and likelihood ratio tests. All raw data and analysis code are freely available from <https://osf.io/5n9j3/>.

3 Results

3.1 SIM-TLX

As in previous work (Harris et al. 2020b), we first evaluated the VR environment with a descriptive analysis

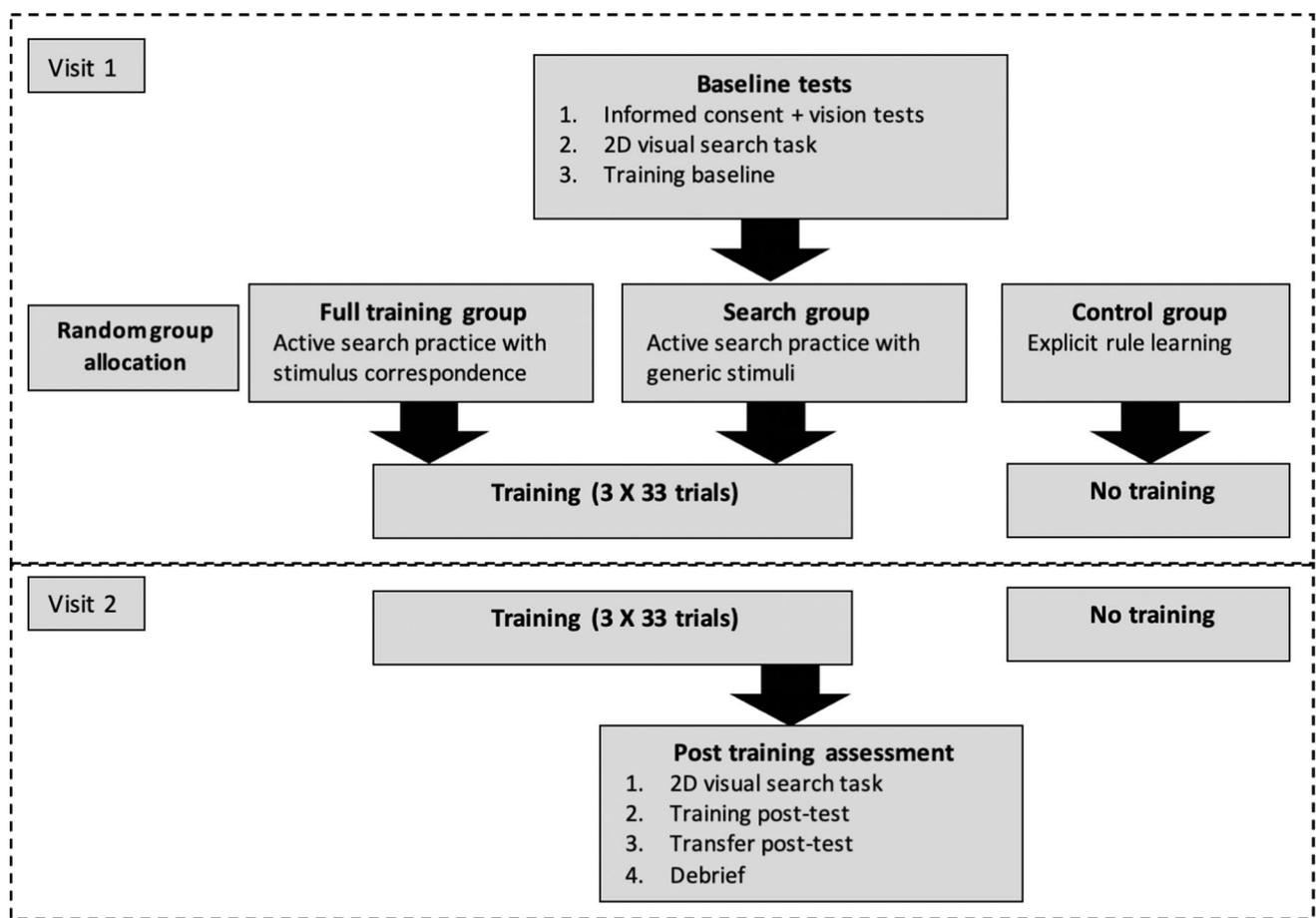
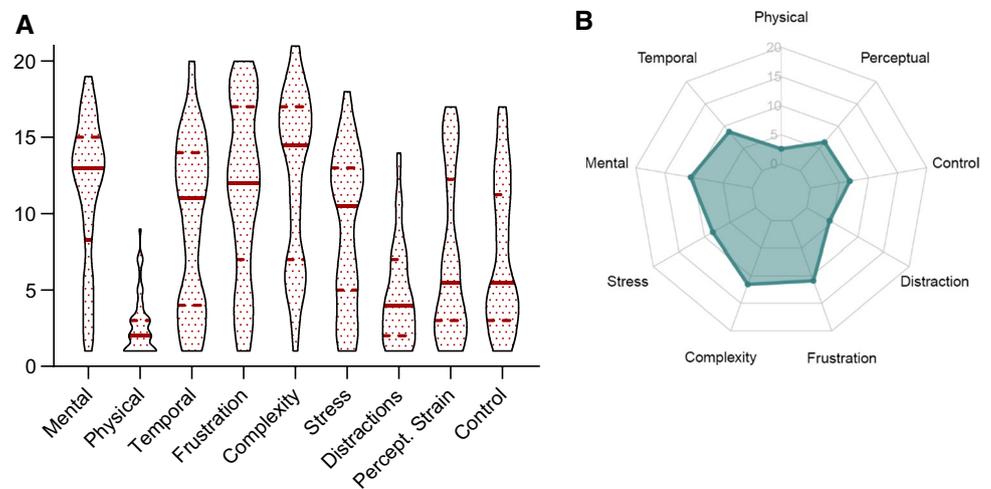


Fig. 3 Schematic representation of the flow of participants through the study

Fig. 4 Violin plots with median (solid line) and inter-quartile range (dashed lines) (a, left) and radar chart (b, right) of scores on the SIM-TLX



of user's workload. The results (Fig. 4) illustrated that the mental demands and the complexity and frustration arising from the task were moderate.² As expected, the

physical demands of the task were low. Participants also reported that there was little perceptual strain induced by the simulation and that it was easy to control, suggesting that although it was cognitively effortful, the simulation

² In the context of the scale maximum of 21.

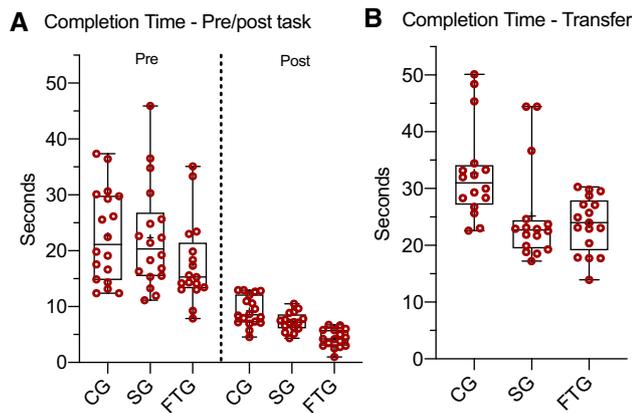


Fig. 5 Box and whisker plot (with data points) showing performance (completion time) on the post-test (a, left) and novel transfer task (b, right)

did not impose additional load as a result of difficulties interacting with it.

3.2 Pre-/post-test task

To assess the effect of training group on performance in the full training task (i.e. a manipulation check), mixed-effects models were run on completion time (seconds) and errors, with participant as a random factor and allowing random effects of test. The overall model predicting time to completion had a total explanatory power (conditional R^2) of 67.06%, in which the fixed effects explained 45.52% of the variance. The model's intercept is at 3.15 (SE = 0.11, 95% CI [2.93, 3.37]). Within this model the effect of group, $F(2,55) = 13.73$, $p < .001$, $n_p^2 = .025$, and the effect of test (pre v post), $F(1,52) = 170.11$, $p < .001$, $n_p^2 = .156$, were both significant. The interaction effect was also significant, $F(2,53) = 5.54$, $p = .007$, $n_p^2 = .010$. Pairwise tests indicated that there were no group differences at baseline ($ps > .18$). At post-test, both FTG ($p < .001$) and SG ($p = .01$) outperformed CG. Participants in the FTG also searched more quickly than SG ($p = .01$) (see Fig. 5a).

The overall model predicting errors had a total explanatory power of 84.17%, in which the fixed effects explain 68.27% of the variance. The model's intercept is at 0.49 (SE = 0.12, 95% CI [0.26, 0.72]). Within this model, the effect of test was significant, $F(1,51) = 365.46$, $p < .001$, $n_p^2 = .293$, with lower errors at post-test. However, the effect of group, $F(2,52) = 1.66$, $p = .20$, $n_p^2 = .004$, and the interaction effect, $F(2,54) = 1.29$, $p = .28$, $n_p^2 = .003$, were not significant.

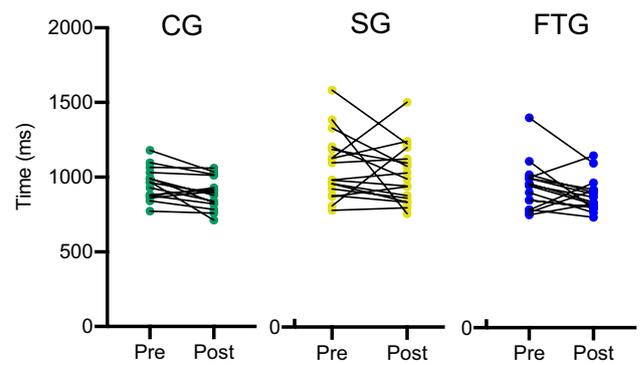


Fig. 6 Pre- and post-row scores (search time) on computerised 2D visual search transfer task

3.3 2D visual search transfer test

To test for transfer of training to the 2D visual search task, mixed-effects models were run on search time (seconds) and number of errors made, with participant as a random factor. The overall model predicting search rate had a total explanatory power (conditional R^2) of 51.62%, in which the fixed effects explain 12.47% of the variance. The model's intercept is at 928.36 (SE = 40.66, 95% CI [850.27, 1006.94]). Within this model, the effect of group, $F(2,47) = 3.75$, $p = .03$, $n_p^2 = .101$, and the effect of test, $F(1,45) = 6.91$, $p = .01$, $n_p^2 = .093$, are both significant. The interaction was not significant, $F(2,45) = 0.04$, $p = .96$, $n_p^2 = .001$. Pairwise tests indicated no significant improvement in any of the three groups, although there was a trend towards an improvement in the FTG ($p = .07$), and SG ($p = .08$), but not CG ($p = .41$) (see Fig. 6).

The overall model predicting search errors had a total explanatory power (conditional R^2) of 61.91%, in which the fixed effects explain 2.55% of the variance. The model's intercept is at 0.83 (SE = 0.22, 95% CI [0.41, 1.24]). Within this model, the effect of group was not significant $F(2,51) = 0.22$, $p = .81$, $n_p^2 = .007$, the effect of test was not significant, $F(1,49) = 0.02$, $p = .90$, $n_p^2 = .000$, and the interaction was not significant, $F(2,49) = 2.49$, $p = .09$, $n_p^2 = .076$.

3.4 Novel transfer test

To compare effect of training group on performance (time and errors) on the novel transfer test, mixed-effects models with participant and trial as random factors were run. The overall model predicting completion time had a total explanatory power of 71.44%, in which the fixed effects explain 11.97% of the variance. The model's intercept is at 4.89 (SE = 0.11, 95% CI [4.67, 5.11]). Within this model, the effect of group was significant, $F(2,44) = 4.96$, $p = .01$, $n_p^2 = .026$. Bonferroni–Holm corrected comparisons

indicated that there was no significant difference between FTG and SG ($p = .27$). SG performed significantly better than CG ($p = .01$), and the difference between FTG and CG was marginal ($p = .058$) (see Fig. 5b).

The overall model predicting errors had total explanatory power of 52.60%, in which the fixed effects explain just 0.66% of the variance. The model's intercept is at 0.39 (SE = 0.13, 95% CI [0.14, 0.63]). Within this model, the effect of group was not significant, $F(2,44) = 0.27$, $p = .76$, $n_p^2 = .001$.

3.5 Gaze parameters

Pre-/post-test. To examine changes in gaze behaviour across training, mixed-effects models were run on saccade size, search rate and gaze entropy. The overall model predicting saccade size had a total explanatory power of 15.38%, in which the fixed effects explain 5.43% of the variance (marginal R²). The model's intercept is at 0.23 (SE = 0.017, 95% CI [0.19, 0.26]). Within this model, the effect of group was not significant, $F(2,54) = 1.75$, $p = .18$, $n_p^2 = .004$. However, the effect of test, $F(1,49) = 26.07$, $p < .001$, $n_p^2 = .027$, and the interaction, $F(2,50) = 4.71$, $p = .01$, $n_p^2 = .010$, were both significant. Pairwise comparisons showed that there were significant increases in saccade size in FTG ($p = .01$) and SG ($p < .001$), but not CG ($p = .43$).

The overall model predicting search rate had a total explanatory power of 43.13%, in which the fixed effects explain 21.80% of the variance. The model's intercept is at 0.85 (SE = 0.046, 95% CI [0.76, 0.94]). Within this model, the effect of test was significant, $F(1,50) = 100.28$, $p < .001$, $n_p^2 = .101$, with a reduction in search rate from pre to post, but the main effect of group, $F(2,58) = 1.33$, $p = .27$, $n_p^2 = .003$, and the interaction were not significant, $F(2,52) = 2.00$, $p = .15$, $n_p^2 = .004$.

The overall model predicting entropy had a total explanatory power of 55.20%, in which the fixed effects explain 7.94% of the variance. The model's intercept is at 1.94 (SE = 0.085, 95% CI [1.78, 2.11]). Within this model, the effect of group was not significant, $F(2,56) = 2.19$, $p = .122$, $n_p^2 = .005$, but the effect of test, $F(1,49) = 6.19$, $p = .02$, $n_p^2 = .007$, and the interaction, $F(2,51) = 4.75$, $p = .01$, $n_p^2 = .011$, were both significant. Pairwise contrasts indicated no change in entropy in CG ($p = .40$) or FTG ($p = .50$), but a significant reduction in SG ($p = .001$).³

Transfer. To compare the gaze behaviour of the training groups during the novel transfer test, a mixed-effects model

was run. The overall model predicting saccade size had a total explanatory power of 13.71%, in which the fixed effects explain 0.33% of the variance (marginal R²). The model's intercept is at 0.44 (SE = 0.053, 95% CI [0.33, 0.54]). Within this model, the effect of group was not significant, $F(2,48) = 0.44$, $p = .65$, $n_p^2 = .011$.

The overall model predicting search rate had a total explanatory power of 50.81%, in which the fixed effects explain 4.33% of the variance. The model's intercept is at 0.55 (SE = 0.031, 95% CI [0.49, 0.61]). Within this model, the effect of group was not significant, $F(2,44) = 1.95$, $p = .15$, $n_p^2 = .010$.

The overall model predicting entropy had a total explanatory power of 35.93%, in which the fixed effects explain 1.55% of the variance. The model's intercept is at 2.42 (SE = 0.050, 95% CI [2.32, 2.52]). Within this model, the effect of group was not significant, $F(2,48) = 0.95$, $p = .39$.

4 Discussion

The increasing accessibility, and decreasing cost, of commercial VR hardware and software has led to considerable interest in VR as training tool, but the effectiveness of VR training of perceptual-cognitive skills is not well understood. In this study, we aimed to test the psychological fidelity of a VR visual search training device, to explore its potential for training house searching in policing contexts. We aimed to establish whether perceptual-cognitive expertise in room searching could be developed in VR, and whether training transferred to a more complex search task. These findings are relevant for developing a better understanding of perceptual-cognitive skill learning in VR, as well as for practical applications of VR.

Our primary prediction that active search practice would lead to improved performance on the transfer task was largely supported, but there was no benefit for the FTG over SG. The SG significantly outperformed CG in the novel transfer test ($p = .01$), but the difference between CG and FTG was only marginal ($p = .058$). This finding suggests that active search practice did transfer to the novel task in a small way. There was, however, no benefit to the FTG over the SG. The FTG had two potential advantages over the SG: trial-and-error learning of the correct search items (as opposed to explicit instruction) and stimulus correspondence between the training and test items. It appears that neither of these had a notable benefit for training. Instead, the findings suggest that a benefit was obtained from both forms of active searching, and that generic stimuli (i.e. no stimulus correspondence) were just as effective as specific stimuli. An alternative possibility is that the familiarity of all participants with the items (common household objects) might have negated the difference in specificity of the training.

³ To validate the use of these metrics, mixed-effects models were run to examine the relationships between gaze metrics and performance. Saccade size, search rate and entropy were all significant predictors of completion time during training ($ps < .02$), and entropy and search rate were significant predictors of completion time in the transfer test ($ps < .001$).

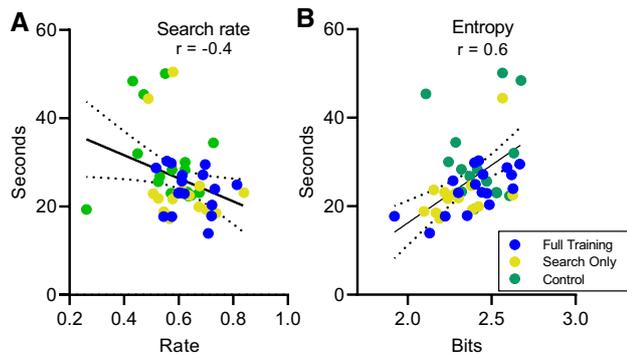


Fig. 7 Correlations of search rate (a, left) and entropy (b, right) with time to completion in novel transfer test

However, participants would not normally be actively searching for these items under time pressure, even if they are familiar. Hence, it seems that specificity of the training items was less important than receiving some form of active search practice in this context.

By contrast, there was no difference between groups in terms of errors committed. Participants made very few errors in the transfer task (all groups averaged less than 0.5 errors per trial), suggesting that there was a ceiling effect in this measure. In real-world policing, conducting a search that is both thorough and fast is important, but the transfer task may not have been sufficiently challenging to observe a difference in errors as well as completion time.

As predicted, active search practice (both FTG and SG) led to changes in gaze behaviour indicative of more efficient visual search (Eckstein 2011; Mann et al. 2007; Wood et al. 2014). Following training, all groups showed a decrease in search rate, but only the two active searching groups (FTG and SG) showed an increase in saccade length and only SG showed a decrease in entropy. Few studies have investigated the development of perceptual-cognitive skill in VR, and these findings suggest that, despite potential differences between real and VR environments, visual search behaviours can be learned in VR. This suggests that the VR environment used in this study had a good degree of psychological fidelity (Harris et al. 2020a).

Despite the aforementioned group differences in performance on the transfer task, accompanying differences in gaze behaviour were not evident, so our final hypothesis—that stimulus correspondence would facilitate transfer of perceptual-cognitive skill—was not supported. This is despite search rate and entropy both being related to better performance in the transfer task (see Fig. 7a, b). Additionally, there were no group differences present in the computerised visual search task either, suggesting that no generalised improvement in visual search ability had occurred. This finding illustrates the challenge of transferring perceptual-cognitive skills to new tasks. Based on the predictions of the

MPT framework (Hadlow et al. 2018), transfer should be likely in this context; there was both stimulus *and* response correspondence, and we aimed to train a higher-level, task-specific perceptual-cognitive skill. But despite satisfying all three conditions, the perceptual learning did not transfer. Specificity is an important feature of expertise, and the differences between the training and transfer tasks, such as the need to move around the room, may have prevented the transfer of gaze behaviours (Barnett and Ceci 2002; Hadlow et al. 2018; Sala and Gobet 2017). Additionally, the level of stimulus correspondence may have just not been high enough to observe the predicted effects. The MPT does not outline exactly what would count as ‘high correspondence’ so the differences between the training and transfer environments—such as new stimuli in the transfer test, in addition to the familiar ones—may have reduced the level of correspondence.

In summary the more gamified, trial-and-error learning method used by the FTG did not show significant benefits over the SG. Gamified training has received support in the literature for both its motivational and learning benefits (Subhash and Cudney 2018). Anecdotally, participants in the FTG often reported the task to be interesting and challenging, which was not the case for the SG. However, this did not transfer into learning benefits in this case. Increases in enjoyment and motivation may be more likely to have long-term impacts on learning through promoting continued participation. In contrast, the active searching element of the training was supported. The educational literature has espoused the benefits of more active learning (Freeman et al. 2014; Trninic 2018), and here, we found it to be beneficial for developing perceptual-cognitive skills. Indeed, such perceptual-cognitive skills are unlikely to develop in training without realistic practice in representative environments (Brunswick 1956; Hadlow et al. 2018).

4.1 Limitations

A limitation of this study is that transfer of training was only assessed in another VR task. The fact that the simple task transferred to a full room search is a promising finding, but transfer of training outside of the virtual world will pose more of a challenge (Barnett and Ceci 2002). Nonetheless, the primary aim of this study was to demonstrate the potential for developing search skills in VR (which was supported by the eye movement metrics) and examining factors that influence transfer. Further work should see to establish whether the training will transfer to a real-world scene, and how visual search behaviour can be trained so as to more readily transfer between contexts. Additionally, as the FTG showed no benefits over SG, it cannot be ruled out that both groups simply benefitted from additional time spent in VR. It seems unlikely that this would account for the training

effects as the task was mostly a perceptual-cognitive one, and required little skill in controlling the VR environment, but future studies should look to additional control groups that perform more distinct control tasks.

5 Conclusions

If VR is to fulfil its promise and become a productive training methodology in environments like sport, rehabilitation, and defence and security, environments with high levels of psychological fidelity are required. That is, VR training environments must be capable of developing the cognitive and visuomotor abilities that underpin performance in these domains. In this study, we demonstrated that gaze behaviours indicative of visual search expertise (search rate, saccade size and entropy) could be trained in a virtual environment, but that transfer of these skills to new tasks remains a challenge; even in immersive VR. Nonetheless transfer of training to a more complex and ecologically valid task was achieved. The enhanced realism and immersion provided by VR (Parsons 2015; Slater 2009), aligned to the motivational (Allcoat and Mühlénen 2018; Salar et al. 2020) and practical benefits, means that VR is set to play a major role in human skills training. Continued testing of psychological fidelity and transfer of training is needed to ensure that adoption of VR training is evidence based.

Author contributions All authors contributed to the design of the studies and preparation of the manuscript.

Funding This work was supported by a Royal Academy of Engineering UKIC Postdoctoral Fellowship awarded to D. Harris.

Availability of data and materials All data are available online from the Open Science Framework [<https://osf.io/5n9j3/>].

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

Ethical approval Ethical approval for this research was obtained from the departmental ethics board, and written consent was given by all participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allcoat, D., & Mühlénen, A. von. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26(0). Retrieved December 19, 2018, from <https://journal.alt.ac.uk/index.php/rlt/article/view/2140>
- Allsop J, Gray R (2014) Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *J Appl Res Mem Cognit* 3(2):63–71
- Barnett SM, Ceci SJ (2002) When and where do we apply what we learn?: A taxonomy for far transfer. *Psychol Bull* 128(4):612–637
- Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang* 68(3):255–278
- Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. [ArXiv:1406.5823](https://arxiv.org/abs/1406.5823) [stat]. Retrieved April 29, 2018, from [ArXiv:1406.5823](https://arxiv.org/abs/1406.5823)
- Biggs AT, Kramer MR, Mitroff SR (2018) Using cognitive psychology research to inform professional visual search operations. *J Appl Res Mem Cognit* 7(2):189–198
- Bird, J. M. (2019). The use of virtual reality head-mounted displays within applied sport psychology. *J Sport Psychol Action*, pp 1–14
- Braun DA, Mehring C, Wolpert DM (2010) Structure learning in action. *Behav Brain Res* 206(2):157–165
- Bright E, Vine SJ, Dutton T, Wilson MR, McGrath JS (2014) Visual control strategies of surgeons: a novel method of establishing the construct validity of a transurethral resection of the prostate surgical simulator. *J Surg Educ* 71(3):434–439
- Brunswick E (1956) Perception and the representative design of psychological experiments, vol 2. University of California Press, Berkeley
- Eckstein MP (2011) Visual search: a retrospective. *J Vis* 11(5):14–14
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014) Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci* 111(23):8410–8415
- Gegenfurtner A, Boucheix JM, Gruber H, Hauser F, Lehtinen E, Lowe RK (2019) The gaze relational index as a measure of visual expertise. *J Expert*, 3(1).
- Gray R (2017) Transfer of training from virtual to real baseball batting. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2017.02183/full>
- Gray R (2019) Virtual environments and their role in developing perceptual-cognitive skills in sports. In: Williams AM, Jackson RC (eds) *Anticipation and Decision Making in Sport*. Taylor and Francis, Routledge. Doi: <https://doi.org/10.4324/9781315146270-19>
- Green P, MacLeod CJ (2016) SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol* 7(4):493–498
- Hadlow SM, Panchuk D, Mann DL, Portus MR, Abernethy B (2018) Modified perceptual training in sport: a new classification framework. *J Sci Med Sport* 21(9):950–958
- Harris DJ, Buckingham G, Wilson MR, Vine SJ (2019a) Virtually the same? How impaired sensory information in virtual reality may disrupt vision for action. *Exp Brain Res* 237(11):2761–2766
- Harris D, Wilson M, Vine S (2019b) Development and validation of a simulation workload measure: the simulation task load index (SIM-TLX). *Virtual Real*. <https://doi.org/10.1007/s10055-019-00422-9>

- Harris DJ, Bird JM, Smart AP, Wilson MR, Vine SJ (2020a) A framework for the testing and validation of simulated environments in experimentation and training. *Front Psychol* 11:605
- Harris DJ, Buckingham G, Wilson MR, Brookes J, Mushtaq F, Mon-Williams M, Vine SJ (2020b) The effect of a virtual reality environment on gaze behaviour and motor skill learning. *Psychol Sport Exerc*, p 101721
- Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv Psychol* (Vol 52, pp 139–183). Elsevier, Amsterdam Retrieved September 3, 2018, from <http://linkinghub.elsevier.com/retrieve/pii/S0166411508623869>
- Karlsson P, Bergmark A (2015) Compared with what? An analysis of control-group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substance use disorders. *Addiction* (Abingdon, England) 110(3):420–428
- Krassanakis V, Filippakopoulou V, Nakos B (2014) EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *J Eye Mov Res*, 7(1). Retrieved December 21, 2018, from <https://bop.unibe.ch/JEMR/article/view/2370>
- Lele A (2013) Virtual reality and its military utility. *J Ambient Intell Hum Comput* 4(1):17–26
- Lintern G, Boot WR (2019) Cognitive training: transfer beyond the laboratory?. *Human Factors*, 0018720819879814.
- Mann DT, Williams AM, Ward P, Janelle CM (2007) Perceptual-cognitive expertise in sport: a meta-analysis. *J Sport Exerc Psychol* 29(4):457–478
- Mannan SK, Pambakian AL, Kennard C (2010) Compensatory strategies following visual search training in patients with homonymous hemianopia: An eye movement study. *J Neurol* 257(11):1812–1821
- Marteniuk RG (1976) *Information processing in motor skills*. Holt, Rinehart and Winston, New York
- Michalski SC, Szpak A, Saredakis D, Ross TJ, Billingham M, Loetscher T (2019) Getting your game on: Using virtual reality to improve real table tennis skills. *PLoS ONE* 14(9):e0222351
- Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A (2016) A systematic review of virtual reality simulators for robot-assisted surgery. *Eur Urol* 69(6):1065–1080
- Moore LJ, Harris DJ, Sharpe BT, Vine SJ, Wilson MR (2019) Perceptual-cognitive expertise when refereeing the scrum in rugby union. *J Sports Sci* 37(15):1778–1786
- Murray NP, Janelle CM (2003) Anxiety and performance: a visual search examination of the processing efficiency theory. *J Sport Exerc Psychol* 25(2):171–187
- Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. *Nature* 434(7031):387–391
- Oberhauser M, Dreyer D (2017) A virtual reality flight simulator for human factors engineering. *Cogn Technol Work* 19(2):263–277
- Parsons TD (2015) Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front Hum Neurosci*. <https://doi.org/10.3389/fnhum.2015.00660/full>
- R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rosalie SM, Müller S (2012) A model for the transfer of perceptual-motor skill learning in human behaviors. *Res Q Exerc Sport* 83(3):413–421
- Sala G, Gobet F (2017) Does far transfer exist? Negative evidence from chess, music, and working memory training. *Curr Dir Psychol Sci* 26(6):515–520
- Salar R, Arici F, Caliklar S, Yilmaz RM (2020) A model for augmented reality immersion experiences of university students studying in science education. *J Sci Educ Technol*. <https://doi.org/10.1007/s10956-019-09810-x>
- Salvucci DD, Goldberg JH (2000) Identifying fixations and saccades in eye-tracking protocols. In: *Proceedings of the symposium on Eye tracking research and applications—ETRA '00* (pp. 71–78). Presented at the symposium, Palm Beach Gardens, Florida, United States: ACM Press. Retrieved February 4, 2019, from <http://portal.acm.org/citation.cfm?doi=355017.355028>
- Saunders J, Davey S, Bayerl PS, Lohrmann P (2019) Validating virtual reality as an effective training medium in the security domain. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR) (pp. 1908–1911). Presented at the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)
- Savelsbergh GJP, Williams AM, Kamp JVD, Ward P (2002) Visual search, anticipation and expertise in soccer goalkeepers. *J Sports Sci* 20(3):279–287
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Signorell A, Aho K, Anderegg N, Aragon T, Arppe A, Baddeley A, Chessel D (2018) DescTools: tools for descriptive statistics. 2015. R package version 0.99, 24. R package version 0.99, 24
- Sireteanu R, Rettenbach R (1995) Perceptual learning in visual search: fast, enduring, but non-specific. *Vision Res* 35(14):2037–2043
- Slater M (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc B Biol Sci* 364(1535):3549–3557
- Stoet G (2010) PsyToolkit: a software package for programming psychological experiments using linux. *Behav Res Methods* 42(4):1096–1104
- Stoet G (2017) PsyToolkit: a novel web-based method for running online questionnaires and reaction-time experiments. *Teach Psychol* 44(1):24–31
- Subhash S, Cudney EA (2018) Gamified learning in higher education: A systematic review of the literature. *Comput Hum Behav* 87:192–206
- Tabachnick BG, Fidell LS (1996) *Using multivariate statistics*. Harper Collins, Northridge, Cal.
- Tirp J, Steingröver C, Wattie N, Baker J, Schorer J (2015) Virtual realities as optimal learning environments in sport—A transfer study of virtual and real dart throwing. *Psychol Test Assess Model* 57(1):13
- Torralba A, Oliva A, Castelano MS, Henderson JM (2006) Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol Rev* 113(4):766–786
- Trninc D (2018) Instruction, repetition, discovery: Restoring the historical educational role of practice. *Instr Sci* 46(1):133–153
- Vickers JN (2007) Perception, cognition, and decision training: The quiet eye in action. *Human Kinetics*
- Vine SJ, McGrath JS, Bright E, Dutton T, Clark J, Wilson MR (2014) Assessing visual control during simulated and live operations: Gathering evidence for the content validity of simulation using eye movement metrics. *Surg Endosc* 28(6):1788–1793
- Vine SJ, Uiga L, Lavric A, Moore LJ, Tsaneva-Atanasova K, Wilson MR (2015) Individual reactions to stress predict performance during a critical aviation incident. *Anxiety Stress Coping* 28(4):467–477
- Wolfe JM (2010) Visual search. *Curr Biol* 20(8):R346–R349
- Wood G, Batt J, Appelboam A, Harris A, Wilson MR (2014) Exploring the impact of expertise, clinical history, and visual search on electrocardiogram interpretation. *Med Decis Making* 34(1):75–83