BRILL

Multisensory Research 37 (2024) 89–124

MULTISENSORY RESEARCH

brill.com/msr

Addressing the Association Between Action Video Game Playing Experience and Visual Search in Naturalistic Multisensory Scenes

Mohammad Hamzeloo [1,2,*,**], **Daria Kvasova** [2] **and Salvador Soto-Faraco** [2,3]

[1] Department of Psychology, University of Hildesheim, 31141 Hildesheim, Germany
[2] Center for Brain and Cognition, University of Pompeu Fabra, 08005 Barcelona, Spain
[3] ICREA — Catalan Institution for Research and Advanced Studies, 08010 Barcelona, Spain
[*]Corresponding author; e-mail: hamzeloo@uni-hildesheim.de
[**]Work performed while at University of Pompeu Fabra.
ORCID iDs: Hamzeloo: 0000-0002-2955-490X; Kvasova: 0000-0002-5220-1379; Soto-Faraco: 0000-0002-4799-3762

Received 6 January 2023; accepted 2 February 2024; published online 13 February 2024; published in print 3 May 2024

Abstract

Prior studies investigating the effects of routine action video game play have demonstrated improvements in a variety of cognitive processes, including improvements in attentional tasks. However, there is little evidence indicating that the cognitive benefits of playing action video games generalize from simplified unisensory stimuli to multisensory scenes — a fundamental characteristic of natural, everyday life environments. The present study addressed if video game experience has an impact on crossmodal congruency effects when searching through such multisensory scenes. We compared the performance of action video game players (AVGPs) and non-video game players (NVGPs) on a visual search task for objects embedded in video clips of realistic scenes. We conducted two identical online experiments with gender-balanced samples, for a total of $N = 130$. Overall, the data replicated previous findings reporting search benefits when visual targets were accompanied by semantically congruent auditory events, compared to neutral or incongruent ones. However, according to the results, AVGPs did not consistently outperform NVGPs in the overall search task, nor did they use multisensory cues more efficiently than NVGPs. Exploratory analyses with self-reported gender as a variable revealed a potential difference in response strategy between experienced male and female AVGPs when dealing with crossmodal cues. These findings suggest that the generalization of the advantage of AVG experience to realistic, crossmodal situations should be made with caution and considering gender-related issues.

Published with license by Koninklijke Brill BV | DOI: 10.1163/22134808-bja10118
 | ISSN: 2213-4794 (print) 2213-4808 (online)

Keywords
action video game, plasticity, attention, visual search, real-world, multisensory

1. Introduction

The rapid development of technology has led to an explosion in video games, which has brought video game playing and its consequences to the forefront of scientific attention. According to a growing body of evidence, video game playing is associated with improvements in a variety of cognitive and perceptual skills (Hisam *et al*., 2018; Oei and Patterson, 2015; Powers *et al*., 2013; Strobach *et al*., 2012). It has been suggested that at least some of the differences in basic cognitive and perceptual performance between action video game players (AVGPs) and non-video game players (NVGPs) are due to baseline individual differences (Boot *et al*., 2008), but a significant body of work in this domain has established that a causal relationship exists between playing action video games and improvements in attentional control (Bavelier and Green, 2019). It has been proposed that attentional control improvement due to playing action video games can be attributed to a training-related transfer between different perceptual and cognitive abilities (Ducrocq *et al*., 2016; Green *et al*., 2016). However, it remains largely unknown whether the improvement observed in AVGPs generalizes beyond simplified laboratory protocols to tasks involving the complexity of naturalistic, multisensory conditions of real-life scenarios. We set out to address this question.

In the literature, action video games are defined as a type of game that requires the processing of a large amount of visual information, presented rapidly over a wide field of view, and often requires the simultaneous tracking of multiple targets under high attentional demands (Green and Bavelier, 2006). These games include the so-called first- or third-person shooter or adventure games: *Medal of Honor* (Electronic Arts Inc.), *Call of Duty* (Activision Publishing Inc.) and *Grand Theft Auto* (Rockstar Games) are examples of popular videogame series of this genre, at the time of writing this article. In all cases, successful performance in this type of games requires extremely efficient attentional control.

Attentional control, a key mechanism for adaptive behavior, refers to our ability to stay focused on task-relevant information while resisting distraction from salient but irrelevant events, and to efficiently respond to changes in the environment that require re-orienting to new sources of relevant information (Engle, 2002). Thus, attentional control enables not only selective attention (focusing on spatial locations or objects that are goal-relevant while minimizing task-irrelevant information), but also the capacity to shift between selective and divided attention regimes in order to consistently monitor changes in the

environment (Corbetta and Shulman, 2002). One of the first scientific studies to address visual attention in AVGPs, compared to NVGPs, reported that they had lower costs associated with targets appearing at low-probability locations in a stimulus detection paradigm (Greenfield *et al*., 1994). Based on this finding, Greenfield *et al*. (1994) suggested that playing action video games may enhance the ability to allocate and divide attention, thereby improving visual search performance.

In the two decades following Greenfield *et al*.'s (1994) study, researchers have become interested in the visual search advantages of AVGPs. For example, AVGPs outperform NVGPs in traditional visual search tasks (Castel *et al*., 2005; Hubert-Wallander *et al*., 2011), flanker/load tasks (Dye *et al*., 2009a; Green and Bavelier, 2003, 2006; Irons *et al*., 2011; Xuemin and Bin, 2010), distraction-based tasks (Chisholm *et al*., 2010; Rupp *et al*., 2016), and a change detection task (Clark *et al*., 2011; Durlach *et al*., 2009). These findings suggest that action video game experience enhances various aspects of top-down attention, and these effects have been found in both cross-sectional and in intervention studies (Bediou *et al*., 2018; Chisholm and Kingstone, 2015a; Schubert *et al*., 2015). Although it has often been proposed that these improvements are the result of changes in selective attention, more recently, enhancement in various additional aspects of top-down attention have been considered, for example, changes in attentional control or in the capacity to swiftly switch between attentional modes according to task demands (Bavelier and Green, 2019).

The vast majority of studies in this area have compared AVGP and NVGP performance on traditional laboratory protocols using simple visual stimuli. Furthermore, only a handful of studies have considered scenarios including sensory modalities other than vision. In a study measuring auditory decision-making, AVGPs were found to be faster than NVGPs at indicating the ear to which the sound was presented, especially at low signal-to-noise ratios (Green *et al*., 2010). In a subsequent study using a demanding auditory discrimination task, AVGPs were able to detect auditory targets and to discriminate them from auditory non-target standards more accurately than NVGPs (Föcker *et al*., 2019). In a crossmodal study, Donohue *et al*. (2010) compared AVGPs and NVGPs in an audiovisual simultaneity judgment task. The results showed that AVGPs were more accurate in discriminating whether pairs of simple visual and auditory stimuli were synchronous or slightly temporally offset. They also revealed an enhanced ability to judge crossmodal temporal order. Of note, Donohue *et al*.'s trials consisted of a small checker-board (5° by 5° visual stimulus) and a pure tone (1200 Hz auditory stimulus), presented at varying onset asynchronies.

Despite the growing body of research on the benefits of action video game experience, few studies have investigated whether the performance benefits

observed in simplified laboratory conditions can also be seen in more complex contexts with some of the features of real-world scenes. Taking a step toward more complex scenarios would be important to understand how the putative superiority of AVGPs in certain tasks transfers to naturalistic environments, which are complex, multisensory, and often semantically meaningful (Kvasova *et al*., 2021; Soto-Faraco *et al*., 2019). In a study by Gaspar *et al*. (2014), AVGPs and NVGPs were compared in a high-fidelity road-crossing task in a virtual-reality environment in two conditions: no distraction (single task) and a dual-task condition in which participants performed a continuous working-memory task while crossing. They found that AVGPs and NVGPs performed similarly during road crossing in both conditions, and they concluded that the attentional benefits associated with AVG experience do not transfer to complex road-crossing performance involving multitasking. There is also some evidence that AVG experience is associated with better 'real-life' skills, such as higher psychomotor skills in laparoscopic surgery (Jalink *et al*., 2014; Lynch *et al*., 2010), alleviation of amblyopic symptoms (Li *et al*., 2015), or reading skills (Franceschini *et al*., 2013).

Recent studies on crossmodal interactions in attentional orienting have highlighted that in naturalistic scenarios, not only temporal and spatial congruence between stimuli across modalities plays a functional role in the control of attention, but also semantic correspondence can facilitate detection and recognition performance (Kvasova and Soto-Faraco, 2019; Roberts and Hall, 2008; Spagna *et al*., 2015). Crossmodal semantic facilitation has been demonstrated in a variety of tasks, including audiovisual matching tasks (Chen and Spence, 2010; Hein *et al*., 2007; Laurienti *et al*., 2003), visual awareness (Chen *et al*., 2011; Cox and Hong, 2015; Hsiao *et al*., 2012), spatial attention (Kvasova *et al*., 2023; List *et al*., 2014; Mastroberardino *et al*., 2015), and object search in naturalistic scenes (Kvasova *et al*., 2019; Pesquita *et al*., 2013).

Here, we report the results of two studies that used a visual search task in scenes that share some of the characteristics of realistic multisensory scenes, to answer two outstanding questions. First, whether AVGPs outperform NVGPs in a search task in which they need to search for an object within complex, dynamic, multisensory scenes. Because naturalistic environments are multisensory, a second interrelated question we sought to answer is whether AVGPs can benefit more efficiently than NVGPs from crossmodal semantic congruence between visual events and their associated sounds when searching for objects.

In the present visual search task, which was similar to the task used in previous studies (Kvasova *et al*., 2019, 2021), targets consisted of ordinary everyday life objects embedded in video clip fragments of complex, naturalistic scenes. For example, participants were asked to search for a dog in a video-clip featuring a busy city street scene. A characteristic object sound (matching

the search target, matching a distractor object present in the scene, or unrelated to any objects in the scene) was presented at the onset of the video-clip, mixed with ambient noise. Reaction times (RTs) as well as visual search accuracy were measured. It should be noted that these stimuli are not strictly equivalent to the naturalistic real-life experience, but they do share many more of its characteristics than the artificial stimuli typically used in search tasks in most studies. For example, the stimuli used here are dynamic instead of static, they are cluttered, they show familiar objects in contextualized interactions, these objects are embedded in semantically structured relationships, and the scenes are multisensory. In this respect, we speculate that the present study takes the issue of naturalistic generalization of the effects of gaming on attentional performance one step further than past research on AVGPs (see Matusz *et al*., 2019, for an example of how this incremental approach can help tackle the difficult problem of ecological validity). However, it is also fair to note important differences of the present stimuli with respect to a naturalistic situation, such as for example the two-dimensional nature of the visual display, the edited sounds, and the lack of spatial congruence of the sounds with the visual targets/distractors. The first difference is a simple matter of experimental convenience (the same reason that two-dimensional displays have been used in a variety of other studies, e.g., Baldassano *et al*., 2018; Peelen and Kastner, 2011; Spiers and Maguire, 2007). The latter two differences relate to design features to control for otherwise important confounding factors in the interpretation of the results (see Section 2.1. *Methods*).

We hypothesized that, (1) if AVGPs benefit from more efficient attentional control to direct their attention to target-relevant information while minimizing target-irrelevant information in multisensory environments, they would outperform NVGPs in the task of searching for objects in videos of naturalistic scenes. We anticipated that, if present, this advantage would be observed in faster reaction times and/or more accurate overall responses in the task. In addition, our second hypothesis was that, (2) if AVGPs have learned to use multisensory environmental cues more efficiently, then we expect a greater crossmodal advantage for AVGPs. That is, AVGPs would exploit crossmodal congruence to a greater extent than NVGPs. We expect this because previous studies have revealed that as AVGPs possess a greater attentional capacity to spill over task-irrelevant information in high-load perceptual scenarios (Dye *et al*., 2009b; Green and Bavelier, 2003). This is consistent with the finding that NVGPs show a reduction in the magnitude of the flanker effect (Bavelier and Green, 2019). Please note that the present study addresses the association, but not the causality between AVGP and performance differences.

The hypotheses, methods, and analysis pipelines for Experiment 1 were preregistered prior to data analysis, with the exception of the Bayesian statistics which were added posteriori. The pre-registration is available at https://osf.

io/sgy5a/. The results of the planned analyses are presented separately from the exploratory ones in Section 2.3. *Results*, below. Experiment 2 is a close replication of Experiment 1 and was not specifically pre-registered.

2. Experiment 1

2.1. Method

2.1.1. Participants

We used G*Power (Faul *et al*., 2009) to estimate the sample size for the study using a repeated-measures ANOVA with within–between interactions, a medium effect size of $f = 0.25$, and an α level of 0.05. The total sample size for a statistical power of 0.95 ($\beta = 0.95$) was estimated to be $N = 54$ (Tomczak *et al*., 2014). By considering inclusion and exclusion criteria, we enrolled 60 individuals using the online platform Prolific.co. Most of the participants were from European countries, with some others from South American countries, the USA, Canada, and one from South Africa. Based on their responses to a questionnaire asking about their video game-playing habits in different genres during the past 12 months (Bavelier Lab Gaming Questionnaire), we had 30 AVGPs and 30 NVGPs. The mean age was 23.78 years and SD = 3.12. There was no significant age difference between the two groups ($t = 0.115$, df = 28, $p = 0.91$), and the two groups were gender-balanced (15 males and 15 females within each group).

Otherwise, the general inclusion criteria were: (1) having a normal or corrected-to-normal vision and hearing; (2) having a good quality of internet connection and access to the experiment from a laptop or personal computer, as the experiment could only be responded to using a keyboard; (3) being a proficient English language speaker; (4) having a false alarm rate on catch trials (trials in which the search target was not present) of less than 15%; and (5) having an average accuracy in the three target-present experimental conditions of greater than 70%. Accuracy exclusion criteria were used to ensure data quality and to provide a sufficient number of trials to reliably estimate reaction times. Although these criteria may reduce the chances of detecting effects of accuracy, they leave room for variation in performance to be captured.

2.1.2. Design

The design includes two independent variables: video game experience (between subjects) and sound–target relationship (within subjects). The first variable, video game experience, was measured using the Bavelier Lab Gaming Questionnaire, version November 2019. Based on the questionnaire adapted from Green *et al*. (2017), we had two groups of participants according to their experience: AVGPs and NVGPs. An AVGP is a person who has played at least five hours per week of first/third person shooter and/or action

RPG/adventure genre of games in the past year. An NVGP is a person who has played no more than one hour per week of any video game genre of in the past year. This categorization of video game playing habits is based on Li *et al*. (2009).

The other variable in the design, sound–target relationship, manipulated as a within-subjects independent factor, refers to the semantic relationship between the visual search target in the task and the object sounds on each trial. This variable had three levels: target-consistent sound, distractor-consistent sound, and neutral sounds. In the target-consistent condition, the sound was matched to the target object. In the distractor-consistent condition, the sound was matched to the distractor (a non-target object present in the scene), and in the neutral condition, the object sound was not semantically congruent with any object in the scene. To measure false alarm rates and to balance the response types (50% yes, 50% no) in the task, the task included additional catch trials in which targets were not present (hence, the correct response was 'NO'). Within these catch trials, as in the target-present trials, the object sound had three levels: target-consistent sound (here, the sound is consistent with the designated search object that is not present in the video clips), distractor-consistent sounds, and neutral sounds. The dependent variable of our design was visual search object performance, measured by correct response times in target-present trials, and by d', calculated from the hit rate in target-present conditions and the false alarm rate in target-absent (catch) trials.

2.1.3. Stimuli

The materials for the visual search task (target objects, sounds, and video clips) in naturalistic settings were selected from those used in Kvasova *et al*. (2019) (see Fig. 1, for an example). There was a set of 156 different video clips that were extracted from movies, TV shows, and advertisements, or recorded by Kvasova *et al*. from scenes of everyday life. The video clips were recorded/played in color, at 30 frames per second and a resolution of 1024×768 pixels. All video clips were cut into 2-s fragments with no fade-in or fade-out. Twelve videos were used for training trials, 72 videos were used for the three target-present trials (24 video each; see Table 1), and another 72 videos were used for the catch trials in which the search target object was not present (again, 24 videos in each condition). The original sound of the videos was replaced by background noise, created by superimposing various everyday sounds that were adapted to the context of the video. Each video clip in the target-present trials contained at least two visual objects with a familiar characteristic sound (e.g., musical instruments, animals, tools, ...). One of these objects was designated as the search target and the other as the distractor. The selection of the target/distractor objects followed two criteria, which were checked by at least two judges during the selection of the videos: (1)

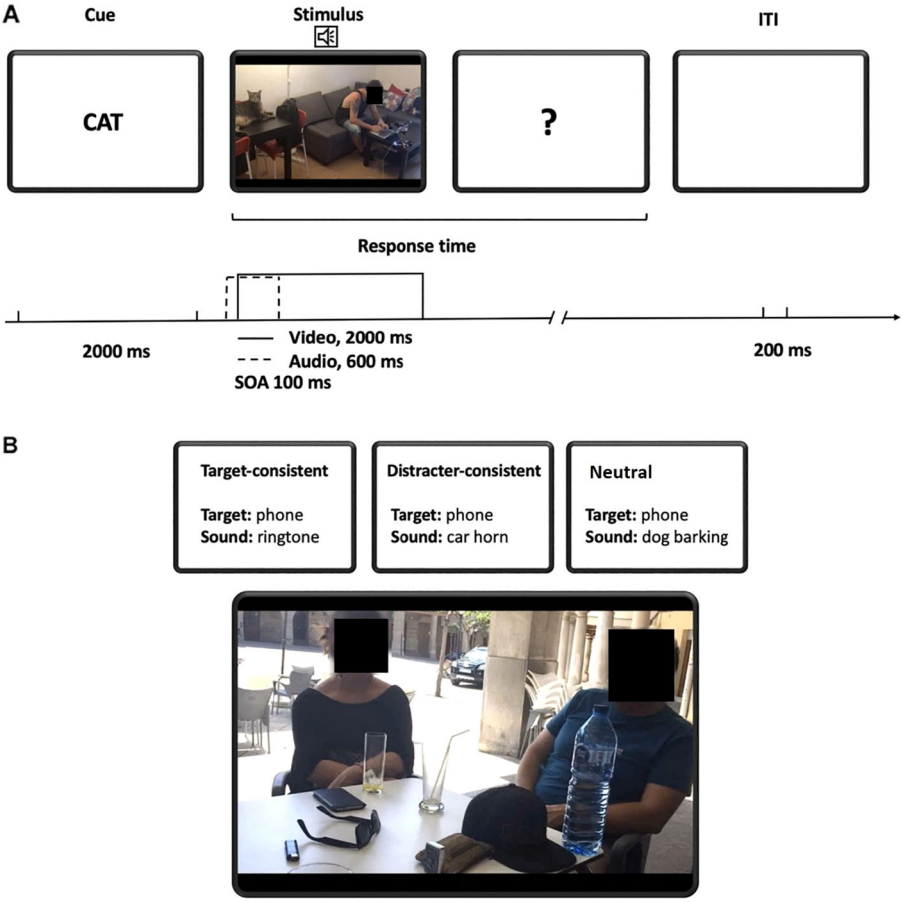

Figure 1. (A) Sequence of events in a typical experimental trial. The trial began with the presentation of the target word for 2000 ms, followed by the auditory cue and the video clip (see times in the figure). There was no time limit on the participant's response. A new target word was presented 200 ms after the participant's response, beginning a new trial. (B) Example of sound conditions. In this stimulus example (shown with a snapshot of the video clip), the possible target (in target-present conditions) is the smartphone. In the target-consistent condition, the sound corresponds to the target (e.g., a ringing tone), in the distractor-consistent condition, the sound corresponds to the distractor object (in this case, the car), and in the catch trial, the sound does not correspond to any object in the scene (e.g., the dog barking). The image is a frame from the video clip filmed by the research group.

they were visible but not a part of the main action in the scene. For instance, if a person is talking on the mobile phone as the main action in the scene, the mobile phone cannot be a target object. (2) Both target and distractor objects were present throughout the video clip.

Table 1.
Distribution of trials and conditions

	Sound condition	Description	Number of trials
Target-present trials	Target-consistent	The target is present and the sound is congruent with the target object	24
	Distractor-consistent	The target is present and the sound is congruent with a distractor object	24
	Neutral	A target is present and the sound is not related to any object in the video	24
Catch trials	Target-consistent	The target is not present and the sound is congruent with the target word	24
	Distractor-consistent	The target is not present and the sound is congruent with a distractor object	24
	Neutral	The target is not present and the sound is not related to any object	24
Total			144

The use of the two designated objects of each video as a target or distractor was counterbalanced across the experimental design to compensate for potential biases related to the specific objects. To reach this goal, we used each video for all three conditions in target-present trials. The three different experimental conditions were created by combining the video with target-consistent sound, distractor-consistent sound, and neutral sound while one of the two target objects was present on the screen before the video, to create six equivalent versions of a target-present trial (two target objects by three conditions). In each video clip in the catch trials, there was at least one visual object with a characteristic sound that was used as a distractor object. We created three equivalent versions of the catch trials from each video, in the same way, so that each catch video appeared in the three conditions in different versions of the experiment. In the end, we had six different versions of the task of equal length including one version of target-present trials and one version of catch trials. Each version was presented to 10 participants with a different random order of videos. Throughout the experiment, each video and possible target/distractor object appeared equally in each condition. This ensured that any variability between video clips could not explain differences between conditions. Characteristic sounds were semantically compatible with the target/distractor object, depending on the particular condition, and were presented centrally, providing no cues to object location. All the sounds were normalized at source to have equivalent sound pressure level (SPL) and were presented for 600 ms.

2.1.4. Procedure

We built the experiment using Builder under the Psychopy package (v. 2020.1.3) and ran it under the Pavlovia.org platform. Psychopy is also the only package with a reaction time precision of less than 4 ms online (Sauter *et al*., 2020). We recruited participants *via* the Prolific.co. platform. Each participant was asked to read the consent form and confirm their agreement to voluntarily participate in the experiment. The study was approved by the UPF Institutional Committee for the Ethical Review of Projects (CIREP-UPF). They could exit the experiment at any time by pressing the 'Escape' key on their keyboard. After consent was given, they filled out a form with demographic data, including age, gender (by selecting M for male, F for female, and O for other), and the video game questionnaire in the first part of the study. We continued running the first part of the study for 119 participants until we had 15 participants in four cells of gender by group in a stratified fashion. If they fell into each group category (AVGP or NVGP), they received an invitation to the second part of the study in less than an hour; if not, they received £ 0.5 as compensation for their participation. By clicking on the invitation link, the instructions for the main task appeared on the screen and they were asked to complete the task indoors with dim lighting and to turn off their phone to avoid distractions during the task. They were also asked to increase the volume on their device up to 80%, or to a level where they could easily hear the sounds. By pressing the space bar, they entered a training block of 12 trials before the beginning of the experiment. During training trials, feedback was provided on the participant's response to ensure that they understood the task. No feedback was given during the experimental blocks.

Each trial began with a cue word designating the search target object, which was displayed in the center of the screen for 2000 ms. This cue was followed by the video clip, with the corresponding sound. Auditory cues (target-consistent, distractor-consistent, or neutral) began slightly ahead of the video, by 100 ms, and lasted for 600 ms. The video was presented for 2000 ms. The participant judged whether the target object was present in the video clip or not. Participants were asked to place their left index finger on the 'Y' key and their right index finger on the 'N' key of their keyboard and to press the Y key if they found the target object in the video or press the N key if the target object was not presented. Participants were instructed to respond as quickly as possible, but there was no time limit for the trial response. A question mark was presented after the video offset, and until a response was made. There was a 200-ms blank screen between trials.

When they finished the task, they were returned to the Prolific website where a successful submission appeared in their account. When their submission was approved by the researcher they were paid £2.5 as compensation for their participation. We collected 85 individual datasets, but 25 (17 NVGPs and

Table 2.
Exclusion of individual subjects

Exclusion criteria	Group	
	AVGPs	NVGPs
False alarm rate in catch trials (higher than 15%)	3	8
Average accuracy in experimental trials (less than 70%)	4	3
Both criteria	1	6
Total	8	17

eight AVGPs) of them were excluded from the analyses due to the accuracy criteria (details provided in Table 2). We continued to run the experiment until we had 60 valid individual datasets to include in the analyses.

2.2. Data Analyses

2.2.1. Preprocessing

To eliminate other data from quick guesses, delayed responses due to subject's inattention, or guesses based on the subject's failure to reach a decision, we applied an outlier filter for RTs ± 2SD around the mean of each condition for each subject: neither RT nor accuracy data were analyzed for these outlier trials. 3.89% of the data was filtered out using the above outlier filter criterion (137 out of 3522 RTs). For RT analyses, incorrect trials from all target-present trials were also filtered out.

2.2.2. Reaction Times

In some of the within-subject analyses, we normalized RT data from the conditions of interest across participants, in order to reduce interindividual differences or expectancy-related differences in performance (demand characteristics) that could be created by filling the Video Game Questionnaire in AVGP and NVGP participants, and to concentrate on the effects of interest. Normalization was performed according to equation (1) for each subject (i) and condition of interest (j):

$$\mathrm{Norm(RT)}_{i,j} = \left(\frac{\mathrm{RT}_{i,j}}{\mathrm{RT}_{i,\mathrm{Neutral}}} - 1 \right) * -1, \tag{1}$$

where $\mathrm{Norm(RT)}_{i,j}$ is the normalized RT for subject i in condition j, $\mathrm{RT}_{i,j}$ is the mean RT for subject i in condition j, and $\mathrm{RT}_{i,\mathrm{Neutral}}$ is the mean RT in the neutral sound condition of subject i. A normalized RT equal to zero indicates that the RT for the condition of interest was equal to the RT for the neutral condition, which established a baseline. Positive normalized RTs indicate the crossmodal facilitation, and negative one's crossmodal distraction.

2.2.3. Performance Data

Signal detection was used to measure the precision of responses from the conditions of interest across participants by calculating d' according to equation (2) for each subject (i) and condition of interest (j):

$$d'_{i,j} = z(\text{hit\_rate}_{i,j}) - z(\text{fa\_rate}_{i,j}), \tag{2}$$

where $d'_{i,j}$ is the d' for subject i in condition j, $\text{hit\_rate}_{i,j}$ is the proportion of yes responses in target-present trials for subject i in condition j, $\text{fa\_rate}_{i,j}$ is the proportion of yes responses in catch trials for subject i in condition j. If hit_rate or fa_rate was 0, the value was replaced by $\frac{1}{\text{total\_response}}$; and if hit_rate or fa_rate was equal to 1, the value was replaced by

$$\frac{(\text{total}_{\text{responses}} - 1)}{\text{total\_response}}.$$

To obtain normalized d' scores from the conditions of interest across participants we applied equation (3) for each subject i and condition of interest j. The purpose of the d' normalization was to investigate whether the congruency effects, if any, led to facilitation or distraction, with respect to the neutral sound condition.

$$\text{Norm}(d')_{i,j} = \left(\frac{d'_{i,j}}{d'_{i,\text{Neutral}}} - 1\right) \tag{3}$$

where $\text{Norm}(d')_{i,j}$ is the normalized d' for subject i in condition j, $d'_{i,j}$ is the mean d' for subject i in condition j, and $d'_{i,\text{Neutral}}$ is the mean d' in the neutral sound condition of subject i. A normalized d' score equal to zero means that the d' for the condition of interest was equal to the d' for the neutral condition, which established a baseline. Positive values indicated facilitation, and negative ones distraction.

2.2.4. Bayesian Analysis

To go beyond the limitations of standard Null Hypothesis Significance Testing (NHST) p-values, such as the inability to provide sufficient evidence for the absence of an effect (for a review see Kubsch *et al.*, 2021), we conducted Bayesian analyses in parallel with each NHST in our study. Given that the goal of this study was to ascertain the presence or absence of the AVGP effect, we recruited Bayes Factor (BF) hypothesis testing for each effect to determine whether the data were more likely under the research hypothesis (supporting the presence of the effect), the null hypothesis (supporting the absence of the effect), or whether there are ambiguous BFs indicating insufficient evidence to adequately support either hypothesis (for a more detailed discussion see, e.g., Dienes, 2014). All Bayesian analyses were implemented in the statistical software package JASP (JASPTeam, 2021).

Table 3.
Descriptive statistics (means and standard deviations) for Experiment 1

Variable	AVGP ($n = 30$)			NVGP ($n = 30$)		
	Male	Female	Total	Male	Female	Total
Age	24.67 (2.92)	23.93 (2.96)	24.30 (2.91)	25.67 (4.05)	23.80 (3.36)	24.23 (3.68)
RTs for TC	1.40 (0.12)	1.42 (0.28)	1.41 (0.21)	1.73 (0.40)	1.58 (0.24)	1.65 (0.33)
RTs for DC	1.46 (0.13)	1.45 (0.46)	1.45 (0.33)	1.70 (0.40)	1.65 (0.32)	1.68 (0.36)
RTs for NC	1.40 (0.13)	1.53 (0.44)	1.47 (0.33)	1.74 (0.39)	1.63 (0.22)	1.69 (0.32)
RTs for TT	1.42 (0.44)	1.51 (0.61	1.46 (0.53)	1.73 (0.59)	1.62 (0.53)	1.68 (0.56)
Norm RTs for TC	0.00 (0.10)	0.05(/11)	0.03 (0.11)	0.00 (0.13)	0.03 (0.08)	0.02 (0.10)
Norm RTs for DC	−0.04 (0.10)	0.06 (0.10)	0.01 (0.11)	0.02 (0.09)	−0.01 (0.10)	0.01 (0.10)
d' for TC	2.52 (0.52)	2.68 (0.33)	2.60 (0.43)	2.72 (0.51)	2.51 (0.47)	2.61 (0.49)
d' for DC	2.79 (0.40)	2.36 (0.40)	2.57 (0.45)	2.59 (0.56)	2.25 (0.60)	2.42 (0.60)
d' for NC	2.71 (0.53)	2.33 (0.59)	2.52 (0.58)	2.72 (0.48)	2.16 (0.40)	2.44 (0.52)
d' Total	2.8 (0.55)	2.52 (0.44)	2.66 (0.51)	2.77 (0.59)	2.38 (0.39)	2.58 (0.53)
Norm d' for TC	−0.05 (0.20)	0.24 (0.45)	0.09 (0.37)	−0.03 (0.27)	0.22 (0.38)	0.12 (0.34)
Norm d' for DC	0.05 (0.17)	0.06 (0.25)	0.05 (0.21)	−0.02 (0.30)	0.05 (0.26)	0.02 (0.28)

DC, distractor-consistent condition; NC, neutral condition; RT, reaction time; TC, target-consistent condition; TT, target-present trials.

2.3. Results

2.3.1. Results of the Pre-Registered Analyses

Descriptive statistics for Experiment 1 are shown in Table 3. To evaluate our first hypothesis, whether the advantage of action video game experience can be transferred to search in complex, multisensory scenes, we addressed whether there is a superiority of AVGPs in speed and/or precision (d'). To test this, we used the intersubject averages per group (AVGP, NVGP) pooled across all other conditions and performed a one-tailed t-test on RTs of correct responses to target-present trials, and on d' scores. The result showed significant differences between the two groups on mean RTs ($t_{58} = 2.61$, $p < 0.01$, Cohen's $d = 0.68$, $BF_{10} = 8.43$) (Fig. 2A), whereas no such difference occurred in the d' total ($t_{58} = 0.63$, $p = 0.265$, Cohen's $d = 0.16$, $BF_{10} = 0.45$) (Fig. 2B). These results confirmed our first hypothesis, suggesting that AVGPs responded faster than NVGPs., albeit with similar precision. In particular, the average advantage of AVGPs was of 220 ms, a 14% variation over the total RT, which can be considered as moderate evidence with a medium effect size. Note that the lack of group differences in precision may be related to the accuracy-based

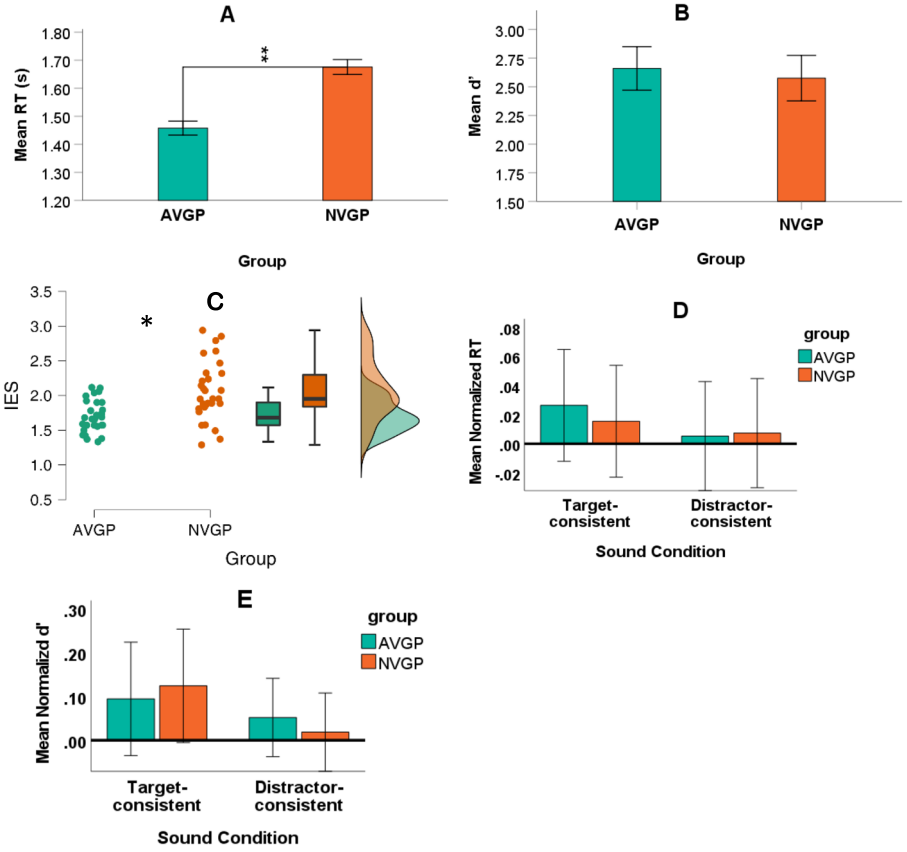

Figure 2. (A) Visual search mean reaction times (RT) for correct responses to target-present trials, plotted separately for action video game players (AVGPs) and non-players (NVGPs). (B) Total visual search accuracy (d') plotted for AVGPs and NVGPs. (C) Inverse efficiency scores (IES) on a raincloud plot with mean and confidence interval (95%) for the two groups. (D) Normalized reaction times in the target-consistent and distractor-consistent conditions plotted for AVGPs and NVGPs. (E) Normalized d' scores in the target-consistent and distractor-consistent conditions plotted for AVGPs and NVGPs. In all plots, error bars indicate the standard error of the mean. Significant differences are indicated by asterisks (*, p-value < 0.05; **, p-value < 0.01).

exclusion criteria, which were introduced precisely to render latency estimates interpretable.

We identified two potential concerns with the interpretation of this analysis using absolute RTs, which may affect the test of Hypothesis 1. A first potential concern that was brought to our attention was that we had no control over the size and quality of the monitors that participants used to perform the task. Since AVGP may have used larger displays and higher quality rendering equipment (an otherwise reasonable assumption), their RT advantage

may have depended on this. We address this concern in Experiment 2 below. A second potential concern with the pre-registered protocol was that we set performance exclusion criteria across the board, which may have tended to select high-performing NVGPs, hence equating performance (i.e., d' scores) across groups. Since, if this were the case, it would indeed contradict the predicted results of group differences in RTs, the significant group difference in RTs strengthens the conclusion. Nevertheless, we decided to further evaluate our first hypothesis in order to compensate for potential speed–accuracy trade-offs between the two groups. In an exploratory analysis we analyzed the Inverse Efficiency Scores (IES) calculated based on equation (4) to combine speed and accuracy into a single metric. The result of a one-tailed t-test on IES showed a significant difference between groups ($t_{58} = 3.01$, $p < 0.005$, Cohen's $d = 0.78$, $\mathrm{BF}_{10} = 9.92$) (Fig. 2C), in the expected direction, and with a slightly larger effect size than the RT effect. Finally, we also repeated the RT and d' analyses including all $N = 85$ subjects tested (i.e., including those that had been initially filtered out by the pre-registered exclusion criteria) and report them in the Supplementary Material. In short, the data were more variable (given the inclusion of low-quality datasets), but the statistical tests yielded equivalent results in RT, and a small but significant d' effect in the expected direction.

$$\mathrm{IES} = \frac{\mathrm{mean(RT)}}{\text{proportion of correct responses for target} - \text{present trials}} \tag{4}$$

To evaluate our second hypothesis, whether AVGPs benefit more from congruent crossmodal cues and/or are more resistant to crossmodal distractors, we entered the normalized RTs into a repeated-measures analysis of variance (ANOVA) with the between-participants factor of group (AVGPs *vs* NVGPs) and the within-participants factor of sound condition (target-congruent and distractor-congruent, normalized by RTs in the neutral condition). Please note that for this test, it is important to compare relative effects (crossmodal facilitation/distraction effects) between the groups, as AVGPs and NVGPs may have different baseline performance (as was indeed shown in the analyses above). The use of normalization allows us to directly test for potential differences in the crossmodal effects without the potential confounding of overall group differences, and intersubject variability. Neither the interaction of group by sound condition ($F_{(1,58)} = 0.218$, $p = 0.64$, ${\eta_p}^2 = 0.001$, $\mathrm{BF}_{10} = 0.06$) (Figure 2D), nor the main effect of sound condition ($F_{(1,58)} = 1.10$, $p = 0.29$, ${\eta_p}^2 = 0.019$, $\mathrm{BF}_{10} = 0.23$) were significant. The effect of group also showed a non-significant result ($F_{(1,58)} = 0.04$, $p = 0.85$, ${\eta_p}^2 = 0.001$, $\mathrm{BF}_{10} = 0.06$), which is not surprising given the normalization.

We performed another repeated-measures ANOVA on the normalized d' scores with the same factors, and similar to the RTs, neither the interaction

of group by sound condition ($F_{(1,58)} = 0.56$, $p = 0.46$, $\eta_p{}^2 = 0.010$, $BF_{10} = 0.02$), nor the main effect of sound conditions ($F_{(1,58)} = 3.09$, $p = 0.08$, $\eta_p{}^2 = 0.05$, $BF_{10} = 0.14$) were significant (Fig. 2E). Group also showed a nonsignificant result ($F_{(1,58)} = 0.01$, $p = 0.98$, $\eta_p{}^2 = 0.000$, $BF_{10} = 0.06$). The analyses then did not confirm our second hypothesis, and in fact Bayesian analyses provided strong evidence for H_0, suggesting that AVGPs do not seem to benefit more from congruent crossmodal cues than NVGPs in terms of speed or accuracy. Nor do they appear to be less distracted by incongruent crossmodal cues than NVGPs.

2.3.2. Reality Check: Crossmodal Congruency Effects

The outcome from the analysis test for potential differences in the crossmodal effect was a null effect. One might wonder whether the possible cause of this null group effect is due to the absence of a crossmodal congruence effect altogether. To address this, we conducted a reality check analysis for crossmodal congruency effects in our data, attempting to replicate congruency effects in other, similar studies. We ran two one-tailed paired-samples t-tests on the data set from Experiment 1 to compare the mean RTs in the target-consistent condition and the distractor-consistent condition with the mean RTs of the neutral condition. The results showed a significant RT advantage of the target-consistent condition over the neutral condition with anecdotal evidence ($t_{59} = 1.85$, $p < 0.05$, Cohen's $d = 0.24$, $BF_{10} = 1.337$). The difference between mean RTs of the distractor-consistent condition and the neutral condition was, however, not significant, with substantial evidence for no effect ($t_{59} = 0.602$, $p = 0.275$, Cohen's $d = 0.08$, $BF_{10} = 0.242$). The results of the t-tests on d' showed a marginal but nonsignificant difference between the target-consistent condition and the neutral condition with anecdotal evidence ($t_{59} = 1.41$, $p = 0.08$, Cohen's $d = 0.18$, $BF_{10} = 0.654$). The difference between the distractor-consistent condition and the neutral condition was also not significant which was supported with substantial evidence for no effect ($t_{59} = 0.266$, $p = 0.396$, Cohen's $d = 0.03$, $BF_{10} = 0.116$).

3. Experiment 2

The results of Experiment 1 showed that AVGPs were faster than NVGPs in the visual search task, without any loss of accuracy (if anything, d' scores went in the same direction as RTs). The conclusion supported by this first study regarding overall search performance is that it extrapolates the search advantage of AVGPs over NVGPs to more complex, dynamic, multisensory scenarios. However, we found no evidence that AVGPs benefit more from crossmodal congruency than NVGPs, a null finding. Experiment 2 was a replication experiment with two purposes, First, in order to confirm the lack of group differences regarding crossmodal advantage. Second, Experiment

2 addressed the concern regarding potential group differences in rendering equipment, raised earlier. Namely, in Experiment 1 we had no control over the size and quality of the monitors that participants used to run the task. It is reasonable to assume that AVGPs tend to have larger, higher quality screens with higher resolution (see Note 1).

3.1. Methods

The design, stimuli, and procedure of Experiment 2 were identical to those of Experiment 1, except for the following aspects. First, we used a new version of the Bavelier Lab Gaming Questionnaire (November, 2022). The new version incorporates two questions for each type of game to specify whether the participant played more than half of the time on a small screen, i.e. 12 inches or smaller, and/or using a touchscreen. These two questions were used to avoid including individuals who play games on smartphones and tablets. We also recorded the resolution, and size of the monitor people used to perform the task.

3.2. Participants

To calculate the sample size, we used the effect size (Cohen's $d = 0.68$) obtained from the analysis of the first hypothesis in Experiment 1. Considering a t-test for comparing two independent means, α level $= 0.05$ and $\beta = 0.95$ in G*Power software (Faul *et al.*, 2009), the total sample size was calculated to be 96. After administering the video game questionnaire, we sent 100 invitations to those who fell into each group category (AVGP or NVGP) based on their gaming habits in the past 12 months. We collected 96 individual records.

Considering the accuracy criteria (hit rate above 85% for catch trials and 70% for target-present trials), we excluded 26 data (14 AVGPs and 12 NVGPs). The final sample consisted of 70 data including 36 AVGPs (16 female, $M = 24.50$, SD $= 4.03$) and 34 NVGPs (18 female, $M = 28.06$, SD $=$ 4.36). Analyses with all 96 data (50 AVGPs and 46 NVGPs) can be found in the Supplementary Material.

3.3. Results

Before conducting the analysis for the main hypotheses, we compared the two groups according to the size of the monitors they used to run the experiment. As expected, AVGPs used larger monitors ($M = 19.39''$, SD $= 5.52$) than NVGPs ($M = 17.64''$, SD $= 3.64$) ($t_{68} = 2.05$, $p < 0.05$). However, screen size is not a good indicator of the quality. Since we recorded the participants' screen resolution in pixels and size in inches, we calculated the pixels per inch (PPI) as a measure of pixel density using equation (5). The results showed that there were no significant differences between the two groups in terms of PPI ($t_{68} = 0.65$, $p = 0.26$), meaning that the quality of the screens in terms of the

Table 4.
Descriptive statistics (means and standard deviations) for Experiment 2

Variable	AVGP ($n = 36$)			NVGP ($n = 34$)		
	Female	Male	Total	Female	Male	Total
Age	25.62 (4.18)	23.60 (3.78)	24.50 (4.03)	27.11 (4.64)	29.12 (3.88)	28.06 (4.36)
RTs for TC	1.74 (0.46)	1.55 (0.28)	1.63 (0.38)	1.73 (0.39)	1.65 (0.36)	1.69 (0.37)
RTs for DC		1.61 (0.31)	1.68 (0.39)	1.72 (0.35)	1.76 (0.41)	1.74 (0.37)
RTs for NC	1.89 (0.50)	1.55 (0.29)	1.70 (0.43)	1.73 (0.38)	1.72 (0.34)	1.72 (0.36)
RTs for TT	1.91 (0.44)	1.74 (0.30)	1.82 (0.37)	1.86 (0.35)	1.88 (0.34)	1.87 (0.34)
Norm RTs for TC	0.07 (0.12)	−0.01 (0.12)	0.03 (0.12)	0.00 (0.08)	0.04 (0.08)	0.02 (0.08)
Norm RTs for DC	0.06 (0.10)	−0.04 (0.09)	0.01 (0.10)	0.00 (0.08)	0.02 (0.09)	−0.01 (0.09)
d' for TC	2.70 (0.55)	2.71 (0.69)	2.71 (0.63)	2.59 (0.66)	2.90 (0.59)	2.74 (0.64)
d' for DC	2.63 (0.61)	2.36 (0.43)	2.48 (0.52)	2.53 (0.69)	2.38 (0.55)	2.46 (0.62)
d' for NC	2.82 (0.55)	2.35 (0.59)	2.56 (0.62)	2.84 (0.39)	2.41 (0.44)	2.64 (0.62)
d' Total	2.63 (0.42)	2.37 (0.37)	2.48 (0.41)	2.56 (0.40)	2.42 (0.37)	2.49 (0.38)
Norm d' for TC	0.03 (0.19)	0.24 (0.53)	0.12 (0.43)	−0.07 (0.28)	0.23 (0.33)	0.07 (0.3)
Norm d' for DC	−0.04 (0.24)	0.06 (0.28)	0.01 (0.26)	−0.10 (0.26)	0.01 (0.29)	−0.05 (0.28)

TC = Target-consistent condition; DC = Distractor-consistent condition; NC = Neutral condition; TT = Target-present Trials.

number of pixels per 1-inch line was the same for both groups. Given that the size of the stimulus was equated as part of the online protocol, the size of the screen could not affect the stimulus size.

$$\text{PPI} = \frac{\sqrt{\text{width}^2 + \text{height}^2}}{\text{size in inches}} \tag{5}$$

RT outlier filtering from all target-present trials (RTs ± 2SD around the mean of each condition for each subject) led to 1.17% of the trials (81 out of 6907 RTs) being removed. Descriptive statistics for Experiment 2 are presented in Table 4.

To test our first hypothesis, we ran one-tailed t-tests on the RTs of correct responses of all target-present trials and on the d' scores. The results showed that there was no significant difference between the two groups in terms of RTs ($t_{68} = 0.59$, $p = 0.72$, Cohen's $d = 0.14$, $\text{BF}_{10} = 0.29$) and d' scores ($t_{68} = 0.05$, $p = 0.52$, Cohen's $d = 0.01$, $\text{BF}_{10} = 0.25$). Although the null result for d' scores is consistent with the result of Experiment 1, the null result for RTs is inconsistent with Experiment 1. The result of a one-tailed t-test on IES also showed a nonsignificant difference between the groups ($t_{68} = 0.74$, $p = 0.77$,

Cohen's $d = 0.18$, $BF_{10} = 0.31$) (Fig. 2C). These results do not support our first hypothesis, as they suggest that AVGPs did not outperform NVGPs in the task of searching for objects in videos of realistic, dynamic complex scenes. As in Experiment 1, we performed the same analyses, but with the full dataset ($N = 96$) without filtering out subjects based on performance criteria. These analyses are reported in the Supplementary Material.

To evaluate our second hypothesis, we entered normalized RTs into a repeated-measures ANOVA with group as the between-subjects factor (AVGPs *vs* NVGPs) and sound condition as the within-subjects factor (target congruent and distractor congruent RTs, normalized using the RTs in the neutral condition). The result of the analysis for the between-subjects factor (group) was not significant ($F_{(1,68)} = 0.35$, $p = 0.55$, ${\eta_p}^2 = 0.005$, $BF_{10} = 0.30$). Sound condition had a significant main effect ($F_{(1,68)} = 5.77$, $p < 0.05$, ${\eta_p}^2 = 0.02$, $BF_{10} = 3.99$), while the interaction of group by sound condition was not significant ($F_{(1,68)} = 0.17$, $p = 0.68$, ${\eta_p}^2 = 0.002$, $BF_{10} = 0.37$). We performed another repeated-measures ANOVA on the normalized d' scores with the same factors. The test for the between-subjects factor (group) was not significant ($F_{(1,68)} = 0.58$, $p = 0.45$, ${\eta_p}^2 = 0.009$, $BF_{10} = 0.26$). The sound condition had a significant main effect ($F_{(1,68)} = 8.94$, $p < 0.01$, ${\eta_p}^2 = 0.121$, $BF_{10} = 4.27$), while the result of the interaction between the group and sound condition was not significant ($F_{(1,68)} = 0.02$, $p = 0.88$, ${\eta_p}^2 = 0.000$, $BF_{10} = 0.26$). Figures 3D and 3E plotted the facilitation effect of the target-consistent sound on both reaction time and the response accuracy for both groups, whereas the distractor-congruent sound showed neither the facilitation nor the distraction effect. These results, consistent with the outcome of Experiment 1, do not support our second hypothesis and suggested that AVGPs and NVGPs benefit equally from crossmodal congruency in speed or accuracy. They also do not seem to be less distracted by incongruent crossmodal cues than NVGPs.

3.4. Reality Check: Crossmodal Congruency Effects

Like in Experiment 1, we performed two one-tailed paired-samples t-tests on the data set from Experiment 1 to compare the mean RTs in the target-consistent and the distractor-consistent conditions with the mean RTs of the neutral condition. The results showed a significant RT advantage of the target-consistent condition over neutral condition with strong evidence ($t_{(69)} = 2.51$, $p < 0.01$, Cohen's $d = 0.30$, $BF_{10} = 4.718$). However, the difference between the mean RTs of the distractor-consistent condition and the neutral condition was not significant, with strong evidence for no effect ($t_{69} = 0.307$, $p = 0.38$, Cohen's $d = 0.04$, $BF_{10} = 0.105$). The results of the t-tests on d' showed a non-significant difference between the target-consistent condition and the neutral condition with anecdotal evidence ($t_{69} = 1.26$, $p = 0.11$, Cohen's

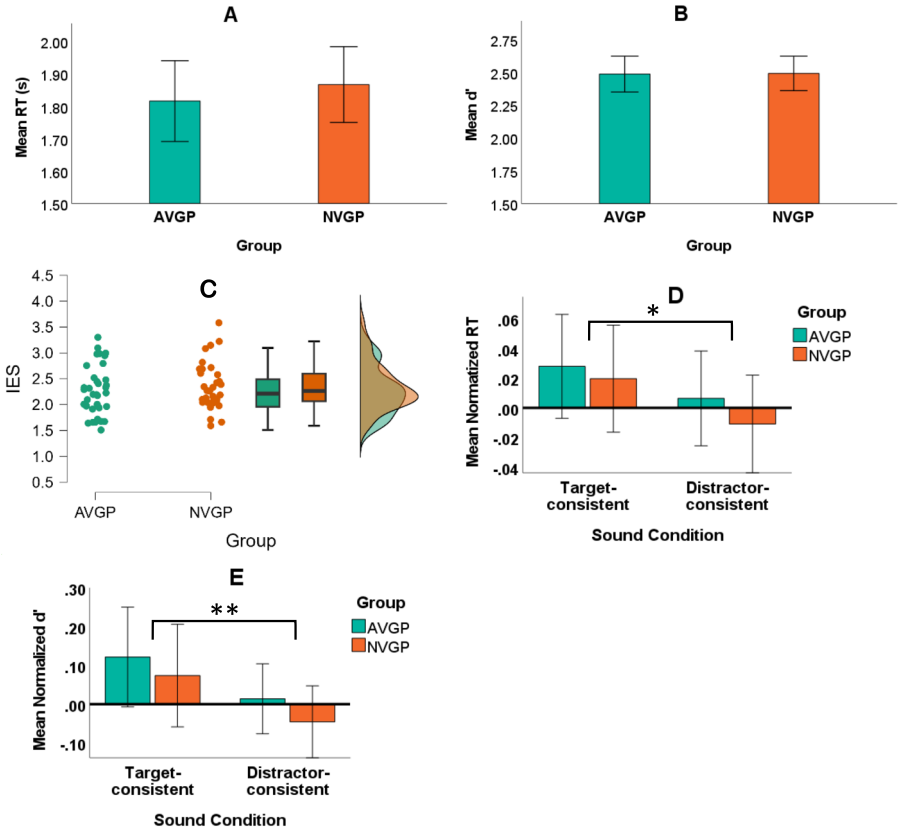

Figure 3. (A) Visual search mean reaction times (RT) for correct responses to target-present trials, plotted separately for action video game players (AVGPs) and non-players (NVGPs). (B) Total visual search accuracy (d') plotted for AVGPs and NVGPs. (C) Inverse efficiency scores on a raincloud plot with mean and confidence interval (95%) for the two groups. (D) Normalized reaction times in the target-consistent and distractor-consistent conditions plotted for AVGPs and NVGPs. (E) Normalized d' scores in the target-consistent and distractor-consistent conditions plotted for AVGPs and NVGPs. In all plots, error bars indicate the standard error of the mean. Significant differences are indicated by asterisks (*, p-value < 0.05; **, p-value < 0.01).

$d = 0.15$, $\mathrm{BF}_{10} = 0.499$). The difference between the distractor-consistent condition and the neutral condition was also not significant which was supported by anecdotal evidence ($t_{69} = 1.55$, $p = 0.063$, Cohen's $d = 0.18$, $\mathrm{BF}_{10} = 0.408$).

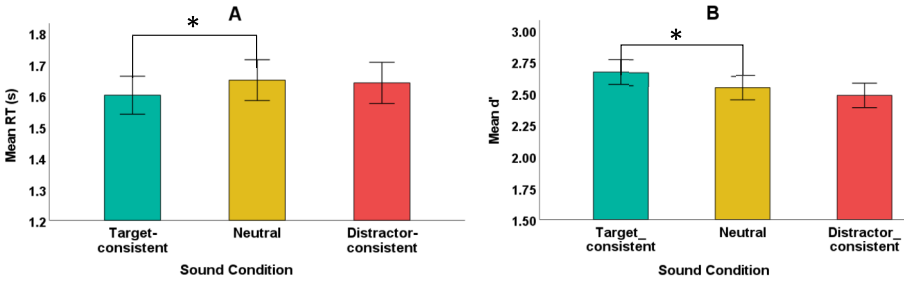

Figure 4. (A) Average reaction times and (B) D' scores in each of the three experimental sound conditions are plotted separately. Error bars indicate the standard error of the mean. Significant differences are indicated by asterisks (*, p-value < 0.05).

4. Exploratory Analyses With Pooled Datasets

In addition to the hypothesis-driven analyses described above, we pooled the two datasets from Experiments 1 and 2 together and performed two additional exploratory analyses that were not directly addressed in our initial hypotheses.

4.1. Reality Check: Overall Crossmodal Effects

In parallel to the reality checks conducted on each experiment individually, here we looked at the comparison between the neutral and target-consistent condition, and between the neutral and distractor-consistent condition across the board. This analysis will shed light on whether, overall, congruent crossmodal cues benefit performance, incongruent crossmodal cues hinder performance, or both. This analysis will also help to situate the present data with respect to previous studies that have found crossmodal benefits in visual search. We conducted two one-tailed, paired-samples t-tests to compare the mean RTs in the target-consistent and distractor-consistent conditions with the mean RTs of the neutral condition. The results showed a significant RT advantage of the target-consistent condition over the neutral condition with strong evidence ($t_{129} = 3.09$, $p < 0.01$, Cohen's $d = 0.27$, $\mathrm{BF}_{10} = 18.105$). The difference between mean RTs of distractor-consistent and neutral conditions was, however, not significant, with substantial evidence for lack of an effect ($t_{129} = 0.623$, $p = 0.267$, Cohen's $d = 0.05$, $\mathrm{BF}_{10} = 0.172$) (Fig. 4A). We performed two more one-tailed paired-samples t-tests, this time to compare d' of the target-consistent and distractor-consistent conditions with the neutral condition. The results showed a significantly higher d' in the target-consistent condition compared to the neutral condition but with low evidential value ($t_{129} = 1.872$, $p = 0.032$, Cohen's $d = 0.16$, $\mathrm{BF}_{10} = 1.026$) (Fig. 4B). Similar to the RT outcome, there was no significant difference between d' of the distractor-consistent and neutral conditions ($t_{129} = 1.101$, $p = 0.136$, Cohen's $d = 0.097$, $\mathrm{BF}_{10} = 0.303$). Taken together, these results suggest that,

relative to the neutral condition, congruent crossmodal cues tend to benefit performance, but incongruent crossmodal cues do not hinder performance. These results are consistent with the reality checks for Experiments 1 and 2, and replicate the recent findings of Kvasova *et al*. (2019) using the same search task in naturalistic scenes, and are consistent with several previous crossmodal findings using more traditional laboratory tasks (Iordanescu *et al*., 2008; Knoeferle *et al*., 2016; Kvasova *et al*., 2023; Laurienti *et al*., 2003).

4.2. Pooled Datasets Analyses and Gender-Dependent Effects

We ran analyses of the main hypotheses using the pooled datasets from Experiments 1 and 2. Since we collected data from a gender-balanced sample in both experiments for both AVGP and NVGP groups, we decided to conduct exploratory analyses of the two main hypotheses, considering gender as a between-subjects variable. Please note that gender was initially balanced to obtain measures from a more representative sample and to avoid selection bias, but not to systematically examine the effects of gender. The participant selection by the online platform included male/female/other options, and therefore our gender variable corresponds to the gender chosen by the participants at the time of the experiment. We only had one data point where gender was selected as 'other', but this dataset was excluded from the analyses due to independent accuracy criteria. For these reasons, we note here that the outcomes of these analyses must be interpreted with caution, as the study was underpowered to examine interactions with the gender variable.

Regarding the first hypothesis (overall advantage of AVGPs over NVGPs), the result of a group (AVGP/NVGP) by gender (male/female) between-subjects ANOVA on mean RTs revealed neither a group by gender interaction nor a main effect of gender ($F_{(1,126)} = 1.43$, $p = 0.23$, ${\eta_p}^2 = 0.011$, $\mathrm{BF}_{10} = 0.15$; $F_{(1,126)} = 0.19$, $p = 0.66$, ${\eta_p}^2 = 0.001$, $\mathrm{BF}_{10} = 0.17$, respectively). The main effect of group was also not significant ($F_{(1,126)} = 3.50$, $p = 0.064$, ${\eta_p}^2 = 0.027$, $\mathrm{BF}_{10} = 0.69$) (Fig. 5A). Thus, the outcome of the analysis using the pooled data sets did not give support to the hypothesis of an AVGP search advantage, in this task. We performed another group by gender between-subjects ANOVA on d' scores. The result indicated no significant effects: the group by gender interaction ($F_{(1,126)} = 0.70$, $p = 0.40$, ${\eta_p}^2 = 0.006$, $\mathrm{BF}_{10} = 0.04$), the main effect of gender ($F_{(1,126)} = 0.295$, $p = 0.59$, ${\eta_p}^2 = 0.002$, $\mathrm{BF}_{10} = 0.15$), and the main effect of group ($F_{(1,126)} = 0{,}17$, $p = 0.68$, ${\eta_p}^2 = 0.001$, $\mathrm{BF}_{10} = 0.14$) (Fig. 5B). Taken together, the RT and d' outcomes using the pooled gender-balanced dataset of 170 participants reveal that AVGPs do not show an advantage over NVGPs in our visual search task, with certain evidential value for the null hypothesis. Similarly, there is no support at all for gender differences in total RTs.

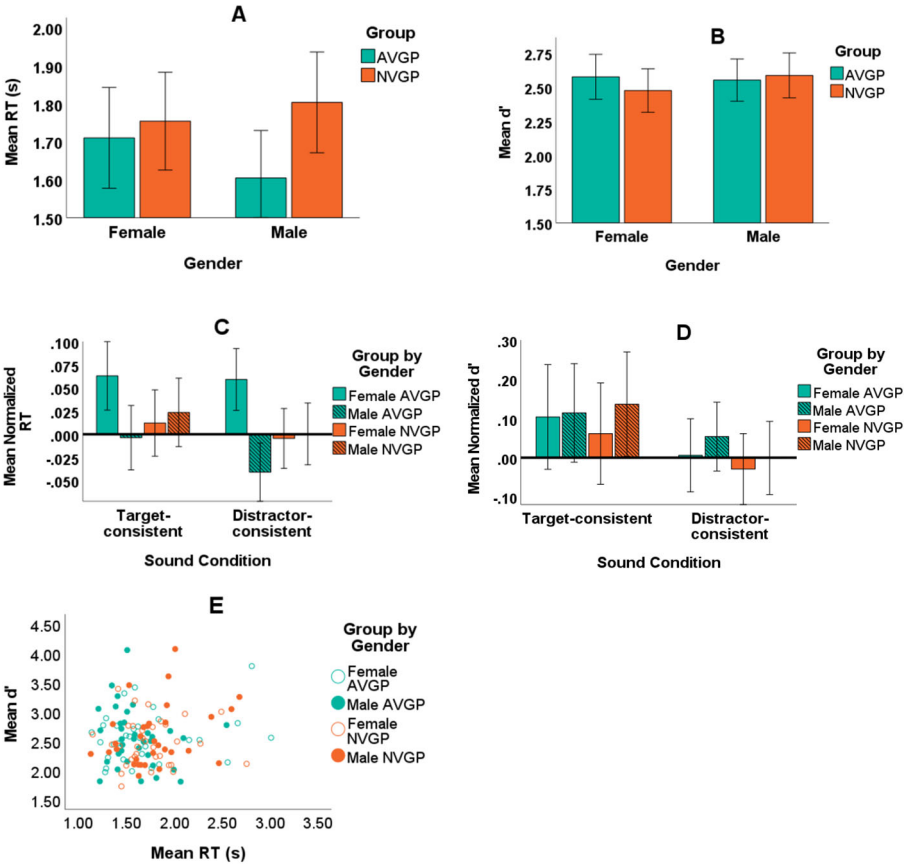

Figure 5. (A) Visual search mean reaction times (RT) for correct responses to target-present trials, plotted separately for male and female action video game players (AVGPs) and non-players (NVGPs). (B) Total visual search accuracy (d') plotted for male and female AVGPs and NVGPs. (C). Normalized reaction times in the target-consistent and distractor-consistent conditions plotted for male and female AVGPs and NVGPs. (D) d' values of the three sound conditions plotted separately for male and female AVGPs and NVGPs. (E) Individual d' scores plotted against average RTs pooled across the four group by gender categories. In all plots, error bars indicate the standard error of the mean.

Regarding the second hypothesis (differential crossmodal advantage for AVGPs), we entered gender and group as between-subjects factors and sound condition as a within-participants variable in a repeated-measures ANOVA on normalized RTs. The test for within-subjects effects revealed a significant main effect for sound condition ($F_{(1,126)} = 5.67$, $p < 0.05$, ${\eta_p}^2 = 0.04$, $\mathrm{BF}_{10} = 1.35$), owing to the crossmodal congruency effect, a significant main effect of gender ($F_{(1,126)} = 6.58$, $p < 0.05$, ${\eta_p}^2 = 0.05$, $\mathrm{BF}_{10} = 7.67$), and no effect of group ($F_{(1,126)} = 0.57$, $p = 0.45$, ${\eta_p}^2 = 0.05$, $\mathrm{BF}_{10} = 1.55$). The

only significant interaction was gender by group ($F_{(1,126)} = 9.55$, $p < 0.01$, ${\eta_p}^2 = 0.07$, $\mathrm{BF}_{10} = 6.71$). The other interactions between sound condition by group ($F_{(1,126)} = 0.008$, $p = 0.93$, ${\eta_p}^2 = 0.00$, $\mathrm{BF}_{10} = 0.24$), sound condition by gender ($F_{(1,126)} = 1.41$, $p = 0.24$, ${\eta_p}^2 = 0.48$, ${\eta_p}^2 = 0.004$, $\mathrm{BF}_{10} = 0.13$) were not significant. To follow up on the significant interaction, *post-hoc* analyses with Bonferroni correction showed that there was a significant slowing for males AVGP compared to females AVGP in the distractor-consistent condition ($M_{\text{difference}} = 0.1$, $p < 0.01$); the female advantage is only marginally significant in the target-consistent condition ($M_{\text{difference}} = 0.067$, $p = 0.052$) (Fig. 5C). These results suggest that crossmodal cues helped to accelerate the RTs of female AVGPs more than those of males AVGPs, possibly due to greater interference from distracting crossmodal cues in males.

We conducted another repeated-measures ANOVA on the normalized d' scores, by entering gender and group as between-subjects factors and sound condition as a within-subjects factor. The test of the within-subjects main effect test revealed a significant main effect for sound condition, $F_{(1,126)} = 11.63$, $p < 0.01$, ${\eta_p}^2 = 0.08$, $\mathrm{BF}_{10} = 0.86$, confirming the crossmodal congruency effect. The main effects of the between-subjects factors showed that neither gender ($F_{(1,126)} = 0.69$, $p = 0.41$, ${\eta_p}^2 = 0.005$, $\mathrm{BF}_{10} = 0.071$) nor group ($F_{(1,126)} = 0.32$, $p = 0.57$, ${\eta_p}^2 = 0.003$, $\mathrm{BF}_{10} = 0.062$) were significant. None of the interactions in this analysis were significant: sound condition by group ($F_{(1,126)} = 0.37$, $p = 0.54$, ${\eta_p}^2 = 0.003$, $\mathrm{BF}_{10} = 0.016$), sound condition by gender ($F_{(1,126)} = 0.007$, $p = 0.93$, ${\eta_p}^2 = 0.000$, $\mathrm{BF}_{10} = 0.019$), gender by group was significant ($F_{(1,126)} = 0.05$, $p = 0.82$, ${\eta_p}^2 = 0.001$, $\mathrm{BF}_{10} = 0.017$), sound condition by group by gender ($F_{(1,126)} = 0.55$, $p = 0.46$, ${\eta_p}^2 = 0.004$, $\mathrm{BF}_{10} = 0.001$) (Fig. 5D). These results confirm the crossmodal congruency effect in their response accuracy, but suggest that this effect was not different for AVGPs *vs* NVGPs, males *vs* females, or any combination thereof (Fig. 5E).

5. General Discussion

The first aim of the present study was to investigate whether the visual search advantage demonstrated by AVGPs over NVGPs with simple stimuli in typical laboratory attention tasks would generalize to a multisensory search display with naturalistic scenes. Using a visual search task in video clips of naturalistic scenes, with audiovisual congruent/incongruent crossmodal cues, Experiment 1 initially suggested that the AVGPs advantage extrapolates to naturalistic scenarios in the form of faster reaction times without a speed-accuracy trade-off. However, this advantage of the AVGP was not replicated in Experiment 2, or when the two datasets were pooled in exploratory analyses. Given that Experiment 2 was conducted to better control for potential differences in the screen

size and resolution, one would have to conclude that rendering equipment differences between groups may have been at the root of the AVGP *vs* NVGP differences seen in Experiment 1. Could there still be a very small effect with inconsistent outcomes given the sample size? We believe that the analyses of pooled datasets ($N = 130$), which yielded null differences with evidential value for absence of an effect based on Bayesian statistics may alleviate this potential concern. Although it may be criticized that one limitation of the current study in testing the transfer of the AVGP advantage to the visual search task in naturalistic displays was due to the participants exclusion criteria, the outcomes of our analyses remained unchanged even when including data from all subjects tested regardless of the performance criteria (see Supplementary Material). These results, therefore, cast reasonable doubt on the idea that the advantage of extensive action video game experience can transfer to search in naturalistic scenes under multisensory conditions.

The literature on transferring the advantage of action video game experience to crossmodal and complex tasks is equivocal. While some previous studies on visual and multisensory tasks have shown an advantage of AVGP experience (Chisholm and Kingstone, 2012, 2015a; Di Luzio *et al*., 2021; Donohue *et al*., 2010; Stewart *et al*., 2020), others have failed to show an advantage on crossmodal and complex tasks (Delmas *et al*., 2022; Gaspar *et al*., 2014; Prevratil *et al*., 2022; Stewart *et al*., 2020). These results suggest that the AVGP advantage may be modality-specific, i.e., the advantage can be transferred to visually familiar or game-like tasks, but cannot be generalized to more realistic multisensory environments.

Given the online, cross-sectional design of the current study, it provides evidence that the advantages associated with action video game experience do not extend beyond typical laboratory stimuli to more complex scenes that incorporate some of the properties of real-world environments. As with many other studies, we cannot draw strong conclusions about the causal relationship between action video game experience and any effect (or lack thereof) from this study, as no training intervention was included in the protocol. Therefore, future work assessing performance within multisensory paradigms pre- and post-action video game training will be important to further evaluate the relation between action video game experience and cognitive abilities in real life.

The second aim of the present study was to assess whether there is a crossmodal advantage for AVGPs over NVGPs. We used normalized data as a synthesis of net crossmodal effects independent of group differences, necessary to test this hypothesis. The results suggested that AVGPs do not benefit more from crossmodal cues than NVGPs (this conclusion can be interpreted within the statistical power of the present study, which is sensitive to medium to large effect sizes). These results are consistent with a previous study (Gao

et al., 2018), which showed that AVGPs do not outperform NVGPs in their ability to integrate audiovisual stimuli. However, AVGPs display greater performance in some crossmodal tasks, such as simultaneity and temporal order judgments for visual and auditory stimuli (Donohue *et al*., 2010). Please note that in contrast to our task, temporal-order judgments (TOJs) are required for effective segregation (rather than integration) of crossmodal events. This is because processing of the meaning of a complex naturalistic sound may require temporal integration due to the nature of the information (for a similar procedure, see Kvasova *et al*., 2019; and for a review, see Vatakis and Spence, 2010). In the task used in our study, the sound was presented slightly earlier than the videos to capture attention and allowed us to study object-based congruency/incongruency effects in the integration of semantic information. It should be noted that a crossmodal semantic congruency effect was present, as shown by the results of all three reality check analyses, and the results of Experiment 2, and the pooled analyses of the normalized RT data. However, crossmodal congruency benefited performance across the board, to the same extent in AVGPs and NVGPs. Overall, this pattern of results suggests that AVGPs do not benefit more than NVGPs from semantic audiovisual integration to direct their attention toward the target object. One explanation for this is that AVGPs and NVGPs may not necessarily employ different strategies or have developed different abilities when processing higher-level, semantic information used in the present study. Hence, the AVGP advantage consistently reported in other studies may be based on early, low-level processing. AVGPs have been shown to benefit from low-level spatial and temporal factors in their visual (Green and Bavelier, 2006; Greenfield *et al*., 1994; Schubert *et al*., 2015; Wong and Chang, 2018; Xuemin and Bin, 2010), auditory (Green *et al*., 2010; Stewart *et al*., 2019), or audiovisual (Zhang *et al*., 2017) search tasks. A recent meta-analysis confirmed a small impact of AVG playing on higher cognitive performance such as inhibition ($g = 0.31$, 95% CI [0.07, 0.56], df $= 7.2$, $p = 0.02$) and verbal cognition ($g = 0.30$, 95% CI [0.03, 0.56], df $= 7.7$ $p = 0.033$), while there was no significant impact on problem-solving in cross-sectional studies (Bediou *et al*., 2018). Future studies need to explore whether extensive action video game experience has a positive impact on high-level cognitive functions or is limited to low-level perception, spatial cognition, even if mediated by top-down attention.

From our exploratory analyses, we found that overall semantic congruence between sounds and target objects accelerated search latencies relative to neutral sounds, whereas semantic incongruence in the distractor-consistent condition did not produce an overall disadvantage relative to the baseline (neutral condition). Hence, consistent crossmodal cues benefit performance, whereas incongruent crossmodal cues do not hinder performance to a measurable extent. This result is consistent with other previous findings (Iordanescu

et al., 2010, 2011, 2013; Knoeferle *et al*., 2016; Kvasova *et al*., 2019, 2023) and demonstrates that the semantic congruency effect can generalize beyond typical laboratory protocols and guide attention in a complex, multisensory environment. It also represents an extension of simpler, laboratory protocols (Iordanescu *et al*., 2008; Laurienti *et al*., 2003).

The design of this study allows us to calculate the false alarm rate and d' for the distractor-consistent condition, which was a limitation of the Kvasova *et al*. (2019) study. In line with the conclusions of that study, there was no difference in d' between the distractor-consistent and neutral conditions. In the literature on crossmodal semantic congruency effects in general, the effects of target-consistent sounds are reliable, but the confounding effects of distractor-consistent sounds are often null or weak, if found at all (Iordanescu *et al*., 2008, 2012; Knoeferle *et al*., 2016). This suggests that the crossmodal congruency effects can be more reliably interpreted as an advantage of congruent sounds, than distractor effects, which neither slowed down response time nor produced a more impulsive response than neutral conditions. An open question here is: if crossmodal semantics are clearly being processed, given the facilitation, why does semantic incongruence not hinder performance? There are two possible answers to this question. One is that when visual stimuli are highly informative, a target-inconsistent sounds does not strongly distract participants from responding to the visual target. In this study, when the cue word is presented to participants, it activates a semantic network associated with the target and creates an attentional template that directs participants' attention to other semantically congruent inputs from other modalities while inhibiting unrelated crossmodal information. In previous versions of the same task using the same materials, Kvasova *et al*. (2019) included a 'no sound' condition. The results of this condition did not differ from the neutral condition (and for that matter, from the distractor-consistent condition). From these results, one could conclude that when sounds are incongruent or irrelevant with respect to the visual target, they may be not be processed sufficiently to produce a distractor effect (please see Kvasova *et al*., 2023, for further discussion on this point). The second possibility is that in our study, some auditory stimuli may not be informative enough to direct participants' attention and influence visual search because some of them are physically similar (e.g., the sound of the coin and keys) or semantically similar (e.g., the sound of musical instruments). Therefore, we cannot be sure that they were strong enough to play an effective distracting role over and above a neutral sound. To clarify this point, further studies should be conducted on the effect of highly informative (hard to inhibit) *vs* low-informative (easy to inhibit) crossmodal cues.

Finally, our findings may provide some insight into potential gender differences (or lack thereof). The low prevalence of female AVGPs has typically led to a large gender imbalance in cross-sectional studies comparing AVGPs

with NVGPs, which have been based almost exclusively on a male population (Cohen *et al*., 2007; Green and Bavelier, 2003; Green *et al*., 2010; Oei and Patterson, 2013). Here, we had a first opportunity at examining a relatively large, gender-balanced sample. Therefore, although sample size calculations were not designed to detect possible gender effects, we believe that exploratory analyses using this variable in our pooled datasets ($N = 130$; 49F, 51M) may provide a first impression, not available in previous studies. With regard to the putative general advantage of AVGPs in the search task as a whole (first proposed hypothesis), our results suggested no gender differences. That is, no overall RT difference, and no gender by group (AVGP/NVGP) interactions. Playing AVGs was not associated with improved processing speed not accuracy of responses (d' scores), and this remained unchanged across gender, as far as the sensitivity of our design allows us to tell.

There were some gender differences in the crossmodal advantage results for male and female AVGPs and NVGPs. Please note that these variations must be interpreted in the context of an overall crossmodal congruency advantage, without video game experience group effects. Now, we observed that female AVGPs benefited more from crossmodal congruency in their RTs and were less distracted by crossmodal incongruency, whereas male AVGPs showed more distraction, with significant RT slowing on distractor-consistent trials (see Fig. 5C,D). These results may suggest a potential difference in response strategy between experienced male and female AVGPs when dealing with crossmodal cues, which needs to be confirmed in the future. For example, with due caution, this finding may be consistent with some previous studies suggesting that there are gender differences in selective attention and spatial cognition (Evans and Hampson, 2015; Halpern, 2013; Merritt *et al*., 2007; Posner and Marin, 1985; Stoet, 2017). Another study investigated the gender-specific effects of AVGP experience and showed that playing AVGs reduced the differences in spatial cognition between males and females (Feng *et al*., 2007). Further studies should be performed to follow up on these findings. For now, one conclusion that can be drawn from this study is that there may be potential for studying gender differences in the effects of AVGP, an area that has been neglected to date.

On a different topic, it is worth mentioning that this study was performed online, with the experiment presented in a browser on the participants' personal device (desktop computer or laptop). This contributes to the generalizability of our findings to more ecological situations, but might have a negative impact on the control of some variables. Despite our efforts (see Section 2.1. *Methods*), the intensity of auditory stimuli and luminance of video clips may have varied between participants. We can only assume that this variation was similar between the groups (AVGPs *vs* NVGPs *vs* male *vs* female). There is

reason to assume that such group differences were negligible, as our data replicated the finding of the crossmodal semantic effect, which is very similar to the laboratory-based study by Kvasova *et al*. (2019). Another potential limitation of the study is the possibility that home setups may account for the reaction time differences observed in Experiment 1. It seems plausible that AVGPs performed the experiment using different setups than NVGPs: perhaps they had ergonomic keyboards or used external monitors (as opposed to laptops). They may also have had more comfortable (gaming) chairs and a dedicated room with fewer distractions. Although research has shown that there are no significant differences for RTs in lab- *vs* web-based tasks (Hilbig, 2016), hardware can cause potential delays in time measurement ranging from 0.0006 to 80 ms for keyboards and 0.0005 to 68 ms for monitors (Crocetta and Andrade, 2015). However, further work in other natural contexts (e.g., virtual reality or other situations that do not involve sitting in front of a display) is needed to evaluate the extent to which the observed advantages can be generalized.

6. Conclusions

The results of the present study suggest that the AVGP advantage in attentional tasks does not extend to complex search tasks with naturalistic, multisensory scenes strongly based on semantic and object-based information. We were also unable to demonstrate that AVGP experience confers a specific advantage over NVGPs in exploiting crossmodal cues, or in resisting distractors. Our data, however, shed light on semantic aspects of multisensory integration by generalizing (and confirming) previous laboratory findings on the semantic congruency/incongruency effect on crossmodal interactions to an online setting. Further findings from exploratory analyses with gender suggest that there may be strategic differences in how this advantage is achieved. In order to gain a full understanding of the AVG advantage in males and females, it will be important to include female participants in this area of research and to more systematically compare results between genders.

Note

1. Please note despite we did not control (or record) participants' monitor resolution and refresh rate, our online protocol included the calibration of the size of the visual stimuli (video clips) and frame rates were equated across participants.

Funding

Salvador Soto-Faraco has been funded by Ministerio de Ciencia e Innovación (Ref: PID2022-137277NB-I00 AEI/FEDER), and AGAUR Generalitat de Catalunya (2021 SGR 00911). This project has been co-funded for 50% by the European Regional Development Fund under the framework of the ERFD Operative Programme for Catalunya 2014-2020, with a grant of €1.527.637,88. This article was published with financial support from the University of Hildesheim Foundation.

Data Availability Statement

The datasets of this study are publicly available at https://osf.io/sgy5a/.

Supplementary Material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.25145081

References

Baldassano, C., Hasson, U. and Norman, K. A. (2018). Representation of real-world event schemas during narrative perception, *J. Neurosci.* **38**, 9689–9699. DOI:10.1523/JNEUROSCI.0251-18.2018.

Bavelier, D. and Green, C. S. (2019). Enhancing attentional control: lessons from action video games, *Neuron* **104**, 147–163. DOI:10.1016/j.neuron.2019.09.031.

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S. and Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills, *Psychol. Bull.* **144**, 77–110. DOI:10.1037/bul0000130.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M. and Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control, *Acta Psychol.* **129**(3), 387–398.

Castel, A. D., Pratt, J. and Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search, *Acta Psychol.* **119**, 217–230. DOI:10.1016/j.actpsy.2005.02.004.

Chen, Y.-C. and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures, *Cognition* **114**, 389–404. DOI:10.1016/j.cognition.2009.10.012.

Chen, Y.-C., Yeh, S.-L. and Spence, C. (2011). Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry, *Front. Psychol.* **2**, 212. DOI:10.3389/fpsyg.2011.00212.

Chiappe, D., Conger, M., Liao, J., Caldwell, J. L. and Vu, K.-P. L. (2013). Improving multitasking ability through action videogames, *Appl. Ergon.* **44**, 278–284. DOI:10.1016/j.apergo.2012.08.002.

Chisholm, J. D. and Kingstone, A. (2012). Improved top-down control reduces oculomotor capture: the case of action video game players, *Atten. Percept. Psychophys.* **74**, 257–262. DOI:10.3758/s13414-011-0253-0.

Chisholm, J. D. and Kingstone, A. (2015a). Action video game players' visual search advantage extends to biologically relevant stimuli, *Acta Psychol.* **159**, 93–99. DOI:10.1016/j.actpsy.2015.06.001.

Chisholm, J. D. and Kingstone, A. (2015b). Action video games and improved attentional control: disentangling selection- and response-based processes, *Psychon. Bull. Rev.* **22**, 1430–1436. DOI:10.3758/s13423-015-0818-3.

Chisholm, J. D., Hickey, C., Theeuwes, J. and Kingstone, A. (2010). Reduced attentional capture in action video game players, *Atten. Percept. Psychophys.* **72**, 667–671. DOI:10.3758/APP.72.3.667.

Clark, K., Fleck, M. S. and Mitroff, S. R. (2011). Enhanced change detection performance reveals improved strategy use in avid action video game players, *Acta Psychol.* **136**, 67–72. DOI:10.1016/j.actpsy.2010.10.003.

Cohen, J. E., Green, C. S. and Bavelier, D. (2007). Training visual attention with video games: not all games are created equal, in: *Computer Games and Team and Individual Learning*, H. F. O'Neil, H. O'Neil and R. S. Perez (Eds), pp. 205–227. Elsevier, Amsterdam, The Netherlands.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain, *Nat. Rev. Neurosci.* **3**, 201–215. DOI:10.1038/nrn755.

Cox, D. and Hong, S. W. (2015). Semantic-based cross-modal processing during visual suppression, *Front. Psychol.* **6**, 722. DOI:10.3389/fpsyg.2015.00722.

Crocetta, T. B. and Andrade, A. (2015). The problem of measuring reaction time using software and hardware: a systematic review, *Rev. Iberoan. Psicol. Ejerc. Deporte* **24**, 341–349.

Delmas, M., Caroux, L. and Lemercier, C. (2022). Searching in clutter: visual behavior and performance of expert action video game players, *Appl. Ergon.* **99**, 103628. DOI:10.1016/j.apergo.2021.103628.

Di Luzio, P., Borgomaneri, S., Sanchioni, S., Tessari, A. and Romei, V. (2021). Exposure to first-person shooter videogames is associated with multisensory temporal precision and migraine incidence, *Cortex* **134**, 223–238. DOI:10.1016/j.cortex.2020.10.009.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results, *Front. Psychol.* **5**, 781. DOI:10.3389/fpsyg.2014.00781.

Donohue, S. E., Woldorff, M. G. and Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities, *Atten. Percept. Psychophys.* **72**, 1120–1129. DOI:10.3758/app.72.4.1120.

Ducrocq, E., Wilson, M., Vine, S. and Derakshan, N. (2016). Training attentional control improves cognitive and motor task performance, *J. Sport Exerc. Psychol.* **38**, 521–533. DOI:10.1123/jsep.2016-0052.

Durlach, P. J., Kring, J. P. and Bowens, L. D. (2009). Effects of action video game experience on change detection, *Mil. Psychol.* **21**, 24–39. DOI:10.1080/08995600802565694.

Dye, M. W. G., Green, C. S. and Bavelier, D. (2009a). The development of attention skills in action video game players, *Neuropsychologia* **47**, 1780–1789. DOI:10.1016/j.neuropsychologia.2009.02.002.

Dye, M. W. G., Green, C. S. and Bavelier, D. (2009b). Increasing speed of processing with action video games, *Curr. Dir. Psychol. Sci.* **18**, 321–326. DOI:10.1111/j.1467-8721.2009.01660.x.

Engle, R. W. (2002). Working memory capacity as executive attention, *Curr. Dir. Psychol. Sci.* **11**, 19–23. DOI:10.1111/1467-8721.00160.

Evans, K. L. and Hampson, E. (2015). Sex-dependent effects on tasks assessing reinforcement learning and interference inhibition, *Front. Psychol.* **6**, 1044. DOI:10.3389/fpsyg.2015.01044.

Faul, F., Erdfelder, E., Buchner, A. and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses, *Behav. Res. Methods* **41**, 1149–1160. DOI:10.3758/BRM.41.4.1149.

Feng, J., Spence, I. and Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition, *Psychol. Sci.* **18**, 850–855. DOI:10.1111/j.1467-9280.2007.01990.x.

Föcker, J., Mortazavi, M., Khoe, W., Hillyard, S. A. and Bavelier, D. (2019). Neural correlates of enhanced visual attentional control in action video game players: an event-related potential study, *Cogn. Neurosci.* **31**, 377–389. DOI:10.1162/jocn_a_01230.

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M. and Facoetti, A. (2013). Action video games make dyslexic children read better, *Curr. Biol.* **23**, 462–466. DOI:10.1016/j.cub.2013.01.044.

Gao, Y.-L., Zhang, S.-H., Zhang, Y., Wang, M.-X., Li, Y.-X., Wang, Z.-N. and Ye, C. (2018). Action video games influence on audiovisual integration in visual selective attention condition, paper presented at the International Conference on Medicine Sciences and Bioengineering, Suzhou, Jiangsu, China. DOI:10.12783/dtbh/icmsb2018/25442.

Gaspar, J. G., Neider, M. B., Crowell, J. A., Lutz, A., Kaczmarski, H. and Kramer, A. F. (2014). Are gamers better crossers? An examination of action video game experience and dual task effects in a simulated street crossing task, *Hum. Fact.* **56**, 443–452. DOI:10.1177/0018720813499930.

Green, C. S. and Bavelier, D. (2003). Action video game modifies visual selective attention, *Nature* **423**, 534–537. DOI:10.1038/nature01647.

Green, C. S. and Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention, *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 1465–1478. DOI:10.1037/0096-1523.32.6.1465.

Green, C. S. and Bavelier, D. (2012). Learning, attentional control, and action video games, *Curr. Biol.* **22**, R197–R206. DOI:10.1016/j.cub.2012.02.012.

Green, C. S. and Bavelier, D. (2015). Action video game training for cognitive enhancement, *Curr. Opin. Behav. Sci.* **4**, 103–108. DOI:10.1016/j.cobeha.2015.04.012.

Green, C. S., Pouget, A. and Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games, *Curr. Biol.* **20**, 1573–1579. DOI:10.1016/j.cub.2010.07.040.

Green, C. S., Gorman, T. and Bavelier, D. (2016). Action video-game training and its effects on perception and attentional control, in: *Cognitive Training: an Overview of Features and Applications*, T. Strobach and J. Karbach (Eds), pp. 107–116. Springer International Publishing AG, Cham, Switzerland. DOI:10.1007/978-3-319-42662-4_10.

Green, C. S., Kattner, F., Eichenbaum, A., Bediou, B., Adams, D. M., Mayer, R. E. and Bavelier, D. (2017). Playing some video games but not others is related to cognitive abilities: a critique of Unsworth *et al*. (2015), *Psychol. Sci.* **28**, 679–682. DOI:10.1177/0956797616644837.

Greenfield, P. M., DeWinstanley, P., Kilpatrick, H. and Kaye, D. (1994). Action video games and informal education: effects on strategies for dividing visual attention, *J. Appl. Dev. Psychol.* **15**, 105–123. DOI:10.1016/0193-3973(94)90008-6.

Halpern, D. F. (2013). *Sex Differences in Cognitive Abilities*, 4th edn. Psychology Press, New York, NY, USA.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L. and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas, *J. Neurosci.* **27**, 7881–7887. DOI:10.1523/JNEUROSCI.1740-07.2007.

Heitz, R. P. (2014). The speed–accuracy tradeoff: history, physiology, methodology, and behavior, *Front. Neurosci.* **8**, 150. DOI:10.3389/fnins.2014.00150.

Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: experimental evidence, *Behav. Res. Methods* **48**, 1718–1724. DOI:10.3758/s13428-015-0678-9.

Hisam, A., Mashhadi, S. F., Faheem, M., Sohail, M., Ikhlaq, B. and Iqbal, I. (2018). Does playing video games effect cognitive abilities in Pakistani children?, *Pak. J. Med. Sci.* **34**, 1507–1511. DOI:10.12669/pjms.346.15532.

Hsiao, J.-Y., Chen, Y.-C., Spence, C. and Yeh, S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure, *Consc. Cogn.* **21**, 775–787. DOI:10.1016/j.concog.2012.02.001.

Hubert-Wallander, B., Green, C. S., Sugarman, M. and Bavelier, D. (2011). Changes in search rate but not in the dynamics of exogenous attention in action videogame players, *Atten. Percept. Psychophys.* **73**, 2399–2412. DOI:10.3758/s13414-011-0194-7.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M. and Suzuki, S. (2008). Characteristic sounds facilitate visual search, *Psychon. Bull. Rev.* **15**, 548–554. DOI:10.3758/PBR.15.3.548.

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J. and Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly, *Atten. Percept. Psychophys.* **72**, 1736–1741. DOI:10.3758/APP.72.7.1736.

Iordanescu, L., Grabowecky, M. and Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets, *Acta Psychol.* **137**, 252–259. DOI:10.1016/j.actpsy.2010.07.017.

Iordanescu, L., Grabowecky, M. and Suzuki, S. (2013). Action enhances auditory but not visual temporal sensitivity, *Psychon. Bull. Rev.* **20**, 108–114. DOI:10.3758/s13423-012-0330-y.

Irons, J. L., Remington, R. W. and Mclean, J. P. (2011). Not so fast: rethinking the effects of action video games on attentional capacity, *Austr. J. Psychol.* **63**, 224–231. DOI:10.1111/j.1742-9536.2011.00001.x.

Jalink, M. B., Goris, J., Heineman, E., Pierie, J.-P. E. N. and ten Cate Hoedemaker, H. O. (2014). The effects of video games on laparoscopic simulator skills, *Am. J. Surg.* **208**, 151–156. DOI:10.1016/j.amjsurg.2013.11.006.

JASPTeam (2021). JASP (version 0.14.1) [computer software]. Retrieved from https://jasp-stats.org/.

Jeffreys, H. (2003). *The Theory of Probability*, 3rd edn. Oxford University Press, Oxford, United Kingdom.

Knoeferle, K. M., Knoeferle, P., Velasco, C. and Spence, C. (2016). Multisensory brand search: how the meaning of sounds guides consumers' visual attention, *J. Exp. Psychol. Appl.* **22**, 196–210. DOI:10.1037/xap0000084.

Kubsch, M., Stamer, I., Steiner, M., Neumann, K. and Parchmann, I. (2021). Beyond p-values: using Bayesian data analysis in science education research, *Pract. Assess. Res. Eval.* **26**, 4. DOI:10.7275/vzpw-ng13.

Kvasova, D. and Soto-Faraco, S. (2019). Not so automatic: task relevance and perceptual load modulate cross-modal semantic congruence effects on spatial orienting, *bioRxiv*, 830679. DOI:10.1101/830679.

Kvasova, D., Garcia-Vernet, L. and Soto-Faraco, S. (2019). Characteristic sounds facilitate object search in real-life scenes, *Front. Psychol.* **10**, 2511. DOI:10.3389/fpsyg.2019.02511.

Kvasova, D., Stewart, T. and Soto-Faraco, S. (2021). Cross-modal semantic relationships guide spontaneous orienting in real-life scenes, *bioRxiv*, 470351. DOI:10.1101/2021.11.29.470351.

Kvasova, D., Spence, C. and Soto-Faraco, S. (2023). Not so fast: orienting to cross-modal semantic congruence depends on task relevance and perceptual load, *Psicologica* **44**, e15662.

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E. and Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices, *Hum. Brain Mapp.* **19**, 213–223. DOI:10.1002/hbm.10112.

Li, R., Polat, U., Makous, W. and Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training, *Nat. Neurosci.* **12**, 549–551. DOI:10.1038/nn.2296.

Li, R. W., Ngo, C. V. and Levi, D. M. (2015). Relieving the attentional blink in the amblyopic brain with video games, *Sci. Rep.* **5**, 8483. DOI:10.1038/srep08483.

List, A., Iordanescu, L., Grabowecky, M. and Suzuki, S. (2014). Haptic guidance of overt visual attention, *Atten. Percept. Psychophys.* **76**, 2221–2228. DOI:10.3758/s13414-014-0696-1.

Lynch, J., Aughwane, P. and Hammond, T. M. (2010). Video games and surgical ability: a literature review, *J. Surg. Educ.* **67**, 184–189. DOI:10.1016/j.jsurg.2010.02.010.

Mastroberardino, S., Santangelo, V. and Macaluso, E. (2015). Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks, *Front. Integr. Neurosci.* **9**, 45. DOI:10.3389/fnint.2015.00045.

Matusz, P. J., Dikker, S., Huth, A. G. and Perrodin, C. (2019). Are we ready for real-world neuroscience?, *J. Cogn. Neurosci.* **31**, 327–338. DOI:10.1162/jocn_e_01276.

Merritt, P., Hirshman, E., Wharton, W., Stangl, B., Devlin, J. and Lenz, A. (2007). Evidence for gender differences in visual selective attention, *Pers. Individ. Dif.* **43**, 597–609. DOI:10.1016/j.paid.2007.01.016.

Nuyens, F. M., Kuss, D. J., Lopez-Fernandez, O. and Griffiths, M. D. (2019). The empirical analysis of non-problematic video gaming and cognitive skills: a systematic review, *Int. J. Ment. Health Addict* **17**, 389–414. DOI:10.1007/s11469-018-9946-0.

Oei, A. C. and Patterson, M. D. (2013). Enhancing cognition with video games: a multiple game training study, *PLoS One* **8**, e58546. DOI:10.1371/journal.pone.0058546.

Oei, A. C. and Patterson, M. D. (2015). Enhancing perceptual and attentional skills requires common demands between the action video games and transfer tasks, *Front. Psychol.* **6**, 113. DOI:10.3389/fpsyg.2015.00113.

Pardina-Torner, H. P., Carbonell, X. and Castejón, M. (2019). A comparative analysis of the processing speed between video game players and non-players, *Aloma* **37**, 13–20. DOI:10.51698/aloma.2019.37.1.13-20.

Peelen, M. V. and Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex, *Proc. Natl Acad. Sci. U. S. A.* **108**, 12125–12130. DOI:10.1073/pnas.1101042108.

Pesquita, A., Brennan, A. A., Enns, J. T. and Soto-Faraco, S. (2013). Isolating shape from semantics in haptic-visual priming, *Exp. Brain Res.* **227**, 311–322. DOI:10.1007/s00221-013-3489-1.

Posner, M. I. and Marin, O. S. (1985). *Attention and Performance XI*. Routledge, London, United Kingdom. DOI:10.4324/9781315630236.

Powers, K. L. and Brooks, P. J. (2014). Evaluating the specificity of effects of video game training, in: *Learning by Playing: Video Gaming in Education*, F. C. Blumberg (Ed.), pp. 302–330. Oxford University Press, Oxford, United Kingdom.

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A. and Alfieri, L. (2013). Effects of video-game play on information processing: a meta-analytic investigation, *Psychon. Bull. Rev.* **20**, 1055–1079. DOI:10.3758/s13423-013-0418-z.

Prevratil, M. J., Harwell, K. W., Boot, W. R. and Towne, T. J. (2022). Investigating the behavioral mechanisms of action video game effects in a complex transfer task, *Acta Psychol.* **230**, 103718. DOI:10.1016/j.actpsy.2022.103718.

Roberts, K. L. and Hall, D. A. (2008). Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory Stroop tasks, *J. Cogn. Neurosci.* **20**, 1063–1078. DOI:10.1162/jocn.2008.20074.

Rupp, M. A., McConnell, D. S. and Smither, J. A. (2016). Examining the relationship between action video game experience and performance in a distracted driving task, *Curr. Psychol.* **35**, 527–539. DOI:10.1007/s12144-015-9318-x.

Sauter, M., Draschkow, D. and Mack, W. (2020). Building, hosting and recruiting: a brief introduction to running behavioral experiments online, *Brain Sci.* **10**(4), 251. DOI:10.3390/brainsci10040251.

Schubert, T., Finke, K., Redel, P., Kluckow, S., Müller, H. and Strobach, T. (2015). Video game experience and its influence on visual attention parameters: an investigation using the framework of the Theory of Visual Attention (TVA), *Acta Psychol.* **157**, 200–214. DOI:10.1016/j.actpsy.2015.03.005.

Soto-Faraco, S., Kvasova, D., Biau, E., Ikumi, N., Ruzzoli, M., Morís-Fernández, L. and Torralba, M. (2019). *Multisensory Interactions in the Real World*. Cambridge University Press, Cambridge, United Kingdom. DOI:10.1017/9781108578738.

Spagna, A., Mackie, M.-A. and Fan, J. (2015). Supramodal executive control of attention, *Front. Psychol.* **6**, 65. DOI:10.3389/fpsyg.2015.00065.

Spiers, H. J. and Maguire, E. A. (2007). Decoding human brain activity during real-world experiences, *Trends Cogn. Sci.* **11**, 356–365. DOI:10.1016/j.tics.2007.06.002.

Stewart, H. J., Martinez, J. L., Perdew, A., Green, C. S. and Moore, D. R. (2020). Auditory cognition and perception of action video game players, *Sci. Rep.* **10**, 14410. DOI:10.1038/s41598-020-71235-z.

Stoet, G. (2017). Sex differences in the Simon task help to interpret sex differences in selective attention, *Psychol. Res.* **81**, 571–581. DOI:10.1007/s00426-016-0763-4.

Strobach, T., Frensch, P. A. and Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations, *Acta Psychol.* **140**, 13–24. DOI:10.1016/j.actpsy.2012.02.001.

Tomczak, M., Tomczak, E., Kleka, P. and Lew, R. (2014). Using power analysis to estimate appropriate sample size, *Trends Sport Sci.* **21**, 195–206.

Vatakis, A. and Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli, in: *Multisensory Object Perception in the Primate Brain*, M. J. Naumer and J. Kaiser (Eds), pp. 95–121. Springer Science + Business Media, New York, NY, USA.

Wong, N. H. L. and Chang, D. H. F. (2018). Attentional advantages in video-game experts are not related to perceptual tendencies, *Sci. Rep.* **8**, 5528. DOI:10.1038/s41598-018-23819-z.

Xuemin, Z. and Bin, Y. (2010). Effects of action video game on spatial attention distribution in low and high perceptual load task, *J. Softw.* **5**, 1434–1441. DOI:10.4304/jsw.5.12.1434-1441.

Zhang, Y.-X., Tang, D.-L., Moore, D. R. and Amitay, S. (2017). Supramodal enhancement of auditory perceptual and cognitive learning by video game playing, *Front. Psychol.* **8**, 1086. DOI:10.3389/fpsyg.2017.01086.