Association Between Action Video Game Playing Experience and Visual Search in Real-life Multisensory scenes

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

[1] Department of Psychology, University of Hildesheim, Hildesheim, Germany, work performed while at [2] Center for Brain and Cognition, University of Pompeu Fabra, Barcelona, Spain,

[3] ICREA – Catalan Institution for Research and Advanced Studies, Barcelona Spain

Author Note

Correspondence concerning this article should be addressed to Mohammad Hamzeloo, Department of Psychology, University of Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Germany. e-mail: hamzeloo@uni-hildesheim.de

Abstract

Prior studies investigating the effects of playing action video games on attentional control have demonstrated improvements on a variety of basic psychophysical tasks. However, as of yet, there is little evidence indicating that the cognitive benefits of playing action video games generalize to naturalistic multisensory scenes – a fundamental characteristic of our natural, everyday life environments. The present study addressed the generalization of attentional control enhancement due to AVGP experience to real-life like scenarios by comparing the performance of action video-game players (AVGPs) with non-players (NVGPs) on a visual search task using naturalistic, dynamic audio-visual scenes. To this end, a questionnaire collecting data on gaming habits and sociodemographic data as well as a visual search task was administered online to a gender-balanced sample of 60 participants of age 18 to 30 years. According to the standard hypothesis, AVGPs outperformed NVGPs in the search task overall, showing faster reaction times without sacrificing accuracy. In addition, in replication of previous findings, semantically congruent cross-modal cues benefited performance overall. However, according to our results, despite the overall advantage in search, and the multisensory congruence benefit, AVGPs did not exploit multisensory cues more efficiently than NVGPs. Exploratory analyses with gender as a variable indicated that the advantage of AVG experience to both genders should be done with caution.

Keywords: Action Video Game Experience; Attentional Control; Visual search Task; Real-life Multisensory scenes; Gender Difference.

Introduction

Over the last decade, the rapid development in smart technologies has produced an explosion of mobile gaming, consequently pushing video-game playing to the forefront of scientific attention. Video game playing is associated with improvements in a variety of cognitive and perceptual skills (Hisam et al., 2018; A. C. Oei & Patterson, 2015; Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013; Strobach, Frensch, & Schubert, 2012), and a significant body of work in this domain has established that a causal relation exists between the playing action video games and attentional control (Bavelier & Green, 2019). It is proposed that attentional control improvement due to playing action video games can be considered as training-related transfer between different perceptual and cognitive abilities (Ducrocq, Wilson, Vine, & Derakshan, 2016; C. Green, Gorman, & Bavelier, 2016). However, it remains unknown whether the attentional control improvement in AVGPs generalizes beyond simplified laboratory conditions to tasks with the complexity of realistic, multisensory conditions. We set out to address this question.

The literature defines action video games 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 attention demands (C Shawn Green & Bavelier, 2006). These games include games in the so-called first- or third-person shooter category or adventure games, such as *Medal of Honor* (Electronic Arts Inc.), *Call of Duty* (Activision Publishing Inc.) or *Grand Theft Auto* (Rockstar Games).

Attentional control is a key mechanism of adaptive behavior and can be defined as our ability to stay focused on task-relevant information while resisting distraction, and being responsive to changes in the environment that require efficient re-orienting to new sources of relevant information (Engle, 2002). Thus, attentional control not only enables 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 to allow consistent monitoring of changes in the environment (Corbetta & Shulman, 2002). One of the first scientific studies related to visual attention in Action Video Game Players (AVGPs), compared to Non-Video Game Players (NVGPs), reported that the cost associated to targets appearing at low probability positions in a stimulus detection paradigm, was lower for AVPGs (Greenfield, DeWinstanley, Kilpatrick, & Kaye, 1994). Based on this finding, Greenfield et al. suggested that playing action video games can foster the skills of allocating and dividing attention, thereby improving visual search performance.

Over the past decade, researchers have taken an interest in visual search advantages in AVGPs. For example, AVGPs outperform NVGPs on traditional visual search tasks (Castel, Pratt, & Drummond, 2005; Hubert-Wallander, Green, Sugarman, & Bavelier, 2011), flanker/load tasks (Matthew WG Dye, C Shawn Green, & Daphne Bavelier, 2009; C Shawn Green & Bavelier, 2003; C Shawn Green & Bavelier, 2006; Irons, Remington, & McLean, 2011; Xuemin & Bin, 2010), distraction-based tasks (Joseph D Chisholm, Hickey, Theeuwes, & Kingstone, 2010; Rupp, McConnell, & Smither, 2016), and a change detection task (Clark, Fleck, & Mitroff, 2011; Durlach, Kring, & Bowens, 2009). These findings have suggested that action video game experience enhances various aspects of top-down attention, and the effect can be seen both in cross-sectional and in intervention studies (Bediou et al., 2018; Joseph D. Chisholm & Kingstone, 2015a; Schubert et al., 2015). Although it has often proposed that these improvements are a 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 based on task demands (Bavelier & Green, 2019).

The vast majority of studies in this area have compared AVGP and NVGP performance on traditional laboratory protocols with basic visual stimuli. What is more, only a handful of studies have considered scenarios where other sensory modalities are present in their experimental designs. In a cross-modal study, Donohue, Woldorff, and Mitroff (2010) compared AVGPs and NVGPs in an audio-visual simultaneity judgment task. The results showed that AVGPs were more accurate to distinguish whether simple visual and auditory pairs occurred in synchrony or slightly offset in time. They also revealed an enhanced ability to determine the temporal order of the different modalities in cross-modal stimuli. In a study measuring auditory decision making, AVGPs were found to be faster compared to NVGPs at indicating the ear to which the sound was presented, especially at low signal to noise ratios (C. S. Green, Pouget, & Bavelier, 2010). In a recent study using a highly demanding auditory discrimination task, AVGPs managed to detect auditory targets and to distinguish them from auditory non-target standards more accurately than NVGPs (Föcker, Mortazavi, Khoe, Hillyard, & Bavelier, 2019).

Despite the growing body of research on the benefits of action video game experience, there have been little studies investigating whether the cognitive benefits observed in the laboratory conditions can also be seen in more complex, naturalistic contexts. This step would be important to understand how the putative superiority of AVGPs in certain tasks transfers to real-life environments that are complex, multisensory, and often semantically meaningful (Soto-Faraco et al., 2019). Recent studies on cross-modal interactions in attention orienting have highlighted that in real-life 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 & Soto-Faraco, 2019; Roberts & Hall, 2008; Spagna, Mackie, & Fan, 2015). Cross-modal semantic facilitation has been shown in a variety of tasks, including audio-visual matching task (Chen & Spence, 2010; Hein et al., 2007; Laurienti et al., 2003), visual awareness (Chen, Yeh, & Spence, 2011; Cox & Hong, 2015; Hsiao, Chen, Spence, & Yeh, 2012), spatial attention (List, Iordanescu, Grabowecky, & Suzuki, 2014; Mastroberardino, Santangelo, & Macaluso, 2015; Pesquita, Brennan, Enns, & Soto-Faraco, 2013), and object search in real-life scenes (Kvasova, Garcia-Vernet, & Soto-Faraco, 2019).

Although evidence to support the idea that benefits of action video game experience can be found in real-life performance is very limited, C. Shawn Green and Bavelier (2015) suggest that action video game experience induces a form of ‘learning to learn’. This meta-learning would enable players to learn to suppress sources of noise or distraction efficiently while they extract task-relevant information very fast and more precisely. Based on this hypothesis, one outstanding question is whether the ‘learning to learn’ ability that putatively emerges as a result of action video game experience can be observed in real-life search. That is, for example when looking for an object in complex, dynamic and naturalistic scenes. Because naturalistic environments are multisensory, a second interrelated question is whether AVGPs can benefit more efficiently than NVGPs from cross-modal semantic congruence between visual events and their associated sounds when they search for an object in real-life scenes. To answer the two questions posed above, we conducted a study using a visual search task on realistic multisensory scenes.

In the present visual search task, targets consist of usual everyday life objects embedded in video clip fragments of natural scenes. For example, participants were asked to

look for a dog in a city street scene. A characteristic object sound (consistent with the search target, with a distractor object present in the scene, or unrelated any of the objects in the scene) was presented by stimulus onset, mixed with ambient noise. Reaction times as well as visual search accuracy were measured.

We hypothesized that, 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 objects in real-life scenarios. We assumed that this advantage would be observed in faster reaction times and/or more accurate responses overall in the task. In addition, if AVPGs have learned to use environmental multisensory cues more efficiently, then we expect a larger cross-modal advantage for AVGPs. That is, in AVGPs the improvement in reaction times in the cross-modal congruent condition with respect to the incongruent one will be proportionally larger, compared to that of the NVGPs. We expect this because previous studies revealed that AVGPs can resist distraction more efficiently in high-load perceptual scenarios (M. W. G. Dye, C. S. Green, & D. Bavelier, 2009; C Shawn Green & Bavelier, 2003), whilst NVGPs show a reduction in the magnitude of the flanker effect (Bavelier & Green, 2019).

The hypothesis, methods and analysis pipeline were pre-registered prior to data analysis. The pre-registration can be found at https://osf.io/sgy5a/. The results of the planned analyses are presented separately from the exploratory ones in the results section, below.

Method

Experimental Design

The design includes two independent variables: Video-game experience (between subjects) and sound-target relation (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 C. S. Green et al. (2017), we had two groups of participants according to their experience: AVGPs and NVGPs. An AVGP is a person who plays 5+ hours per week of First/Third-person shooter and/or action-RPG/adventure genre of games. An NVGP is a person who plays at most 1 hour per week of any genre of video games. This categorization of video game playing habits is based on Li, Polat, Makous, and Bavelier (2009).

The other variable in the design, Sound-target relation, manipulated as an independent factor within-subjects, relates to the semantic relationship between the visual search target in the task and the object sounds on each trial. This variable had 3 levels: target-consistent sound, distractor-consistent sound, and neutral sounds. In the target-consistent condition, the sound was matched with the target object. In the distractor-consistent condition, the sound was matched with 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. In order to be able 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 are not present (hence, the correct response was ‘NO’). Within these catch trials, the object sound had 3 levels, like in target-present trials: target-consistent sound (here, the sound is consistent with the designated search object, which is not present in the video clips), distractor-consistent sounds, and neutral sounds. The dependent variable of our design was visual search object performance, that was measured with reaction times of correct responses in target-present

trials, and with d' using the hit rate in target-present conditions and false-alarm rate from target-absent (catch) trials.

Participants

We used G Power to estimate the sample size for the study, taking a repeated-measures ANOVA within-between interactions with a medium effect size f=0.25, and α level= 0.05. The total sample size for a statistical power of 0.95 was estimated at N=54 (Tomczak, Tomczak, Kleka, & Lew, 2014). By considering drop-out and inclusion criteria, we enrolled 60 individuals using the online platform Prolific.co. Most of the participants were from European countries, and some other participants from South American countries, the US, Canada, and one from South Africa. According to their responses to a questionnaire asking about video game play habits of different genres during the previous 12 months (Bavelier Lab Gaming Questionnaire), we had 30 AVGPs and 30 NVGPs. The mean age was 23.78 and SD = 3.12. There was not a 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, 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, (3) the false alarm rate in catch trials (trials in which the search target was not present) was less than 15%, and (4) the average accuracy in the three target-present experimental conditions was more than 70%.

Stimuli

The materials for the visual search task (target objects, sounds, and video clips) in naturalistic environments 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 extracted from movies, TV shows, advertisements, and others recorded by Kvasova et al. from everyday life scenes. Video clips were recorded/played in color, with 30 fps and 1024 * 768 pixels resolution. All video-clips were edited to 2s duration fragments, without fade-in or -out. Twelve videos were used for training trials, 72 videos were used for the three target-present experimental conditions (24 each; see Table 1), and 72 further videos were used for catch trials in which the search target object was not present (again, 24 videos in each condition). The original sounds of the videos had been replaced with background noise created by the superposition of various everyday life sounds matched with the context of the video. Each video clip in target-present trials contained at least two visual objects that have a familiar characteristic sound (such as musical instruments, animals, tools, ...). One of these objects was designated as the search target and the other as the distractor. The choice of target/distractor objects followed two criteria, checked during selection of the videos by at least two judges: (1) They were visible but not a part of the main action in the scene. For instance, if a person is talking on the cellphone as a main action in the scene, the cellphone cannot be a target object. (2) Both target and distractor objects were present throughout the video clip. The contribution 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 specific objects. To reach this goal, we used each video of 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 present in the screen before the video to create six equivalent versions of a target-present trial (2 target objects * 3 conditions). In each video clip in catch trials, there was at least one visual object that had a characteristic sound, that was used as a distractor object. We created 3 equivalent versions of catch trials from each video, in the same way, so that each catch video

appeared in the three conditions across different versions of the experiment. By 6 versions of target-present trials and 3 versions of catch trials, we had six versions of the task with the same length. Each version was presented to 10 participants with different random order of videos. Characteristic sounds were semantically compatible with the target/distractor object, depending on the particular condition, and were presented centrally, providing no cues for object location. All the sounds were normalized at origin to have equivalent SPL and were presented for 600ms.

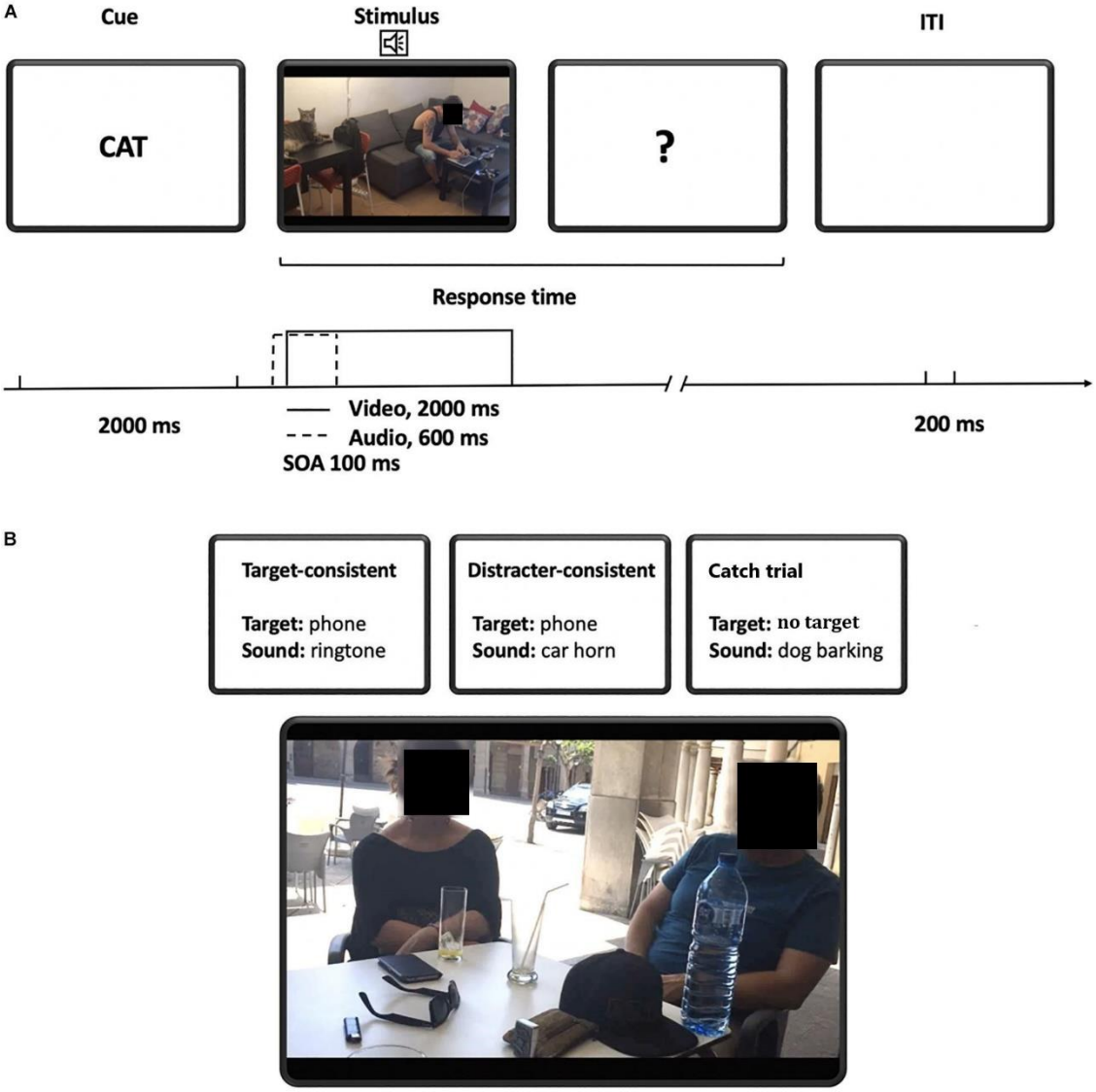

Figure 1. (A) Sequence of events in a typical experimental trial. The trial started with the presentation of target word for 2000 ms, followed by the auditory cue and the video-clip (see times in figure). There was no time limit for the participant response. 200 ms after the participant had responded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

Procedure

The coronavirus outbreak triggered us to run the experiment online. We built the experiment using Builder under Psychopy package (v. 2020.1.3) and ran it under the Pavlovia.org platform. Psychopy also is the only package with reaction time precision under 4 ms online (Sauter, Draschkow, & Mack, 2020). We recruited participants from the Prolific.co. platform. Each participant was asked to read the informed consent and confirm their agreement to participate in the experiment voluntarily. The study approved by the UPF Institutional Committee for Ethical Review of Projects (CIREP-UPF). They were able to exit the experiment at any moment by pressing the 'Escape' key in their keyboard. After consent approval, they filled a form of demographic data including age, gender (by selecting M for males, F for females, and O for others), and the Video Game Questionnaire in the first part of the study. If they felt in each group category (AVGP or NVGP), they have received an invitation for the second part of the study. By clicking on the invitation link, the instruction of the main task appeared on the screen and they were asked to do the task in indoor conditions with dim illumination, turn their phone off in order to avoid distractions during the task. They were also asked to increase the volume of their device up to 80% or to some extent in which they can easily hear the sounds. By pressing the space bar, they entered a 12-trial training block before the beginning of the experiment. Feedback related to the participant's response was provided in the training trials to make sure that participants understand the task. No feedback was provided during the experimental blocks.

Table 1: Distribution of trials and conditions

	Condition (type of trial)	**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 sound	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 sound	The target is not present and the sound is not related to any object	24
Total			144

Each trial started with a cue word designating the search target object, printed in the middle of the screen for 2000ms. This cue was followed by the video clip, with the corresponding sound. Auditory cues (target-consistent, distractor-consistent, or neutral sound) begin slightly ahead of the video, by 100ms and last for 600ms. The video was shown for 2000ms. The participant judged whether or not the target object is present in the video clip. They responded by pressing the Y key as Yes with their left index finger in the case of finding the target object in the video or pressing the N key as No with their right index finger in the case of the target object were not presented. Participants were instructed to respond as fast as possible, but there was no time limitation for the trial response. A question mark was presented after the video offset, and until a response was made. There were a 200ms blank screen between trials.

When they finished the task, they were returned to the Prolific website where a successful submission appeared in their account that waiting for the researcher's approval. If

the accuracy of the data they provided was more than 85% of all catch trials, and 70% of all target-present trials, their submission was approved and they were paid 2.5 £ as a compensation for their participation. We collected 85 individual datasets but 25 (17 NVGPs and 8 AVGPs) of them were excluded from analyses because of the accuracy criteria. We continued to run the experiment until we had 60 valid individual datasets to enter in the analyses.

Data analyses

Data pre-processing

To eliminate other processes outside the interest of the study, including fast lucky guesses, delayed responses due to subject's inattention, or guesses based on the subject's failure to reach a decision, we considered 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. In some of the within-subject analyses, where indicated, we used normalized RTs. Data from neutral trials were used to normalize the data from the conditions of interest across participants, in order to reduce inter-individual differences and concentrate on the effects of interest. The normalization was done according to Equation (1) for each subject (i) and condition of interest (j):

$$(1) \qquad NormRT_j_i = RT_j_i \,/\, NeutralRT_i$$

Where $NormRT_j_i$ is the normalized RT for subject i in condition j, RT_j_i is the mean RTs for subject i in condition j, and $NeutralRT_i$ is the mean RTs in the neutral sound condition of subject i.

Accuracy data

Signal Detection was used to measure 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):

$$(2) \qquad d'_j_i = z(hit_rate_j_i) - z(fa_rate_j_i)$$

where d'_j_i is the d prime for subject i in condition j, $hit_rate_j_i$ is yes responses / total responses for subject i in condition j of target-present trials, $fa_rate_j_i$ is yes responses / total responses for subject i in condition j of catch trials, $z(hit_rate_j_i)$ and $z(fa_rate_j_i)$ is standardized z score for $hit_rate_j_i$ and $fa_rate_j_i$. If hit_rate or fa_rate is equal to 0, it was replaced with 1/total_responses; and if hit_rate or fa_rate is equal to 1 it was replaced with (total_responses-1)/total_responses.

Additionally, data from neutral trials were used to be able to explore whether the congruence effects were seen in the conditions of interest, if any, are due to cross-modal benefit in target detection, cross-modal interference from distractors, or both.

Results

Results of pre-registered analyses

To evaluate our first hypothesis, whether the benefit of action video game experience can transfer to search in real-life scenarios, we addressed whether there is a superiority of AVPGs in RTs and/or precision (*d*'). To test this, we used the inter-subject averages per group (AVPG, NVGP) across all other conditions pooled and performed a one-sided t-test on RTs of correct responses to target-present trials, and on *d'* scores. The result showed

significant differences between groups on mean RTs ($t_{(58)} = 7.10$, $p < 0.001$, *Cohen's d*=0.39) (Figure 2A), whereas no such difference occurred in the *d'* total ($t_{(58)} = 0.68$, $p = 0.245$) (Figure 2B). These results confirmed the first part of our first hypothesis and indicate that AVGPs responded faster, albeit with similar precision than NVGPs. In particular, the average advantage of AVPGs was of 220ms, a 14% variation over the total RT, which can be considered a medium effect size. Please note that the lack of group differences in precision may be related to the accuracy-based inclusion criteria, which were introduced precisely to render latency estimations interpretable.

Table 2: Descriptive statistics (means and standard deviations)

Variable	AVGP (n=30)			NVGP (n=30)			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Age	25(3)	24(3)	24(3)	25(4)	24(3)	24(4)	25(3)	24(3)	24(3)
RTs for TC	1.40(.12)	1.42(.28)	1.41(.21)	1.73(.40)	1.58(.24)	1.65(.33)	1.56(.34)	1.50(.27)	1.53(.30)
RTs for DC	1.46(.13)	1.45(.46)	1.45(.33)	1.70(.40)	1.65(.32)	1.68(.36)	1.58(.32)	1.55(.40)	1.56(.34)
RTs for NC	1.40(.13)	1.53(.44)	1.47(.33)	1.74(.39)	1.63(.22)	1.69(.32)	1.57(.34)	1.58(.34)	1.58(.34)
RTs for TT	1.42(.44)	1.51(.61	1.46(.53)	1.73(.59)	1.62(.53)	1.68(.56)	1.58(.54)	1.56(.58)	1.57(.56)
Norm RTs for TC	1.00(.10)	0.95(/11)	0.97(.11)	1.00(.13)	0.97(.08)	0.99(.11)	1.00(.11)	0.96(.09)	0.98(.11)
Norm RTs for DC	1.05(.10)	0.95(.10)	1.00(.11)	0.98(.09)	1.01(.10)	0.99(.09)	1.01(.10)	0.98(.10)	0.99(.10)
D prime for TC	2.52(.52)	2.68(.33)	2.60(.43)	2.72(.51)	2.51(.47)	2.61(.49)	2.62(.52)	2.60(.41)	2.61(.46)
D prime for DC	2.79(.40)	2.36(.40)	2.57(.45)	2.59(.56)	2.25(.60)	2.42(.60)	2.69(.49)	2.31(.51)	2.50(.53)
D prime for NC	2.71(.53)	2.33(.59)	2.52(.58)	2.72(.48)	2.16(.40)	2.44(.52)	2.72(.49)	2.24(.50)	2.48(.55)
D prime Total	2.8(.55)	2.52(.44)	2.66(.51)	2.77(.59)	2.38(.39)	2.58(.53)	2.78(.56)	2.45(.42)	2.62(.52)

TC= Target_consistent condition; DC= Distractor_consistent condition; NC= Neutral condition; TT= Target-present Trials;

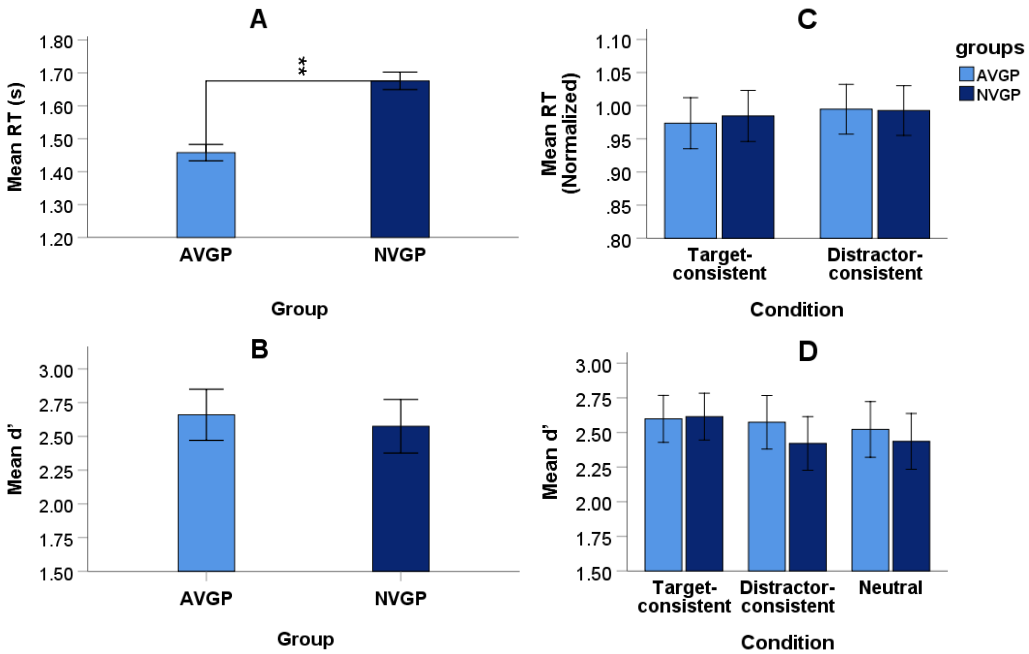

Figure 2. (A) Visual search average reaction times toward targets and error rates plotted separately for Action Video Game Players (AVGP) and Non-Video Game Players (NVGP). In all plots, error bars indicate the standard error of mean, and asterisks indicate significant differences (**p-value<0.01). **(B)** Visual search accuracy (d prime) plotted for the two groups (all three sound conditions pooled; see text for details). **(C)** Normalized reaction times in the target-consistent and distractor-consistent conditions. **(D)** D primes plotted separately in the three sound conditions.

To evaluate our second hypothesis, whether AVGPs can benefit more from cross-modal cues and/or are more resistant to cross-modal distractors, we expected to see an interaction between Group and Sound Condition. We entered 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 Sound Condition (target-congruent and distractor congruent, normalized by RTs in the neutral condition). The interaction of group*condition was not significant ($F_{(1, 58)} = 0.218$, $p = 0.64$) (Figure 2C). We ran another repeated measures ANOVA on *d'* scores with the same factors and the result of the interaction between groups and sound conditions was not significant ($F_{(2, 57)} = 0.526$, $p = 0.59$) (Figure 2D). The analyses did not confirm our second hypothesis and suggest that AVGPs do not seem to benefit more, or to be more interfered from, cross-modal cues than NVGPs in speed or accuracy.

Exploratory analyses

In addition to the hypothesis-driven analyses described above, we performed two further exploratory analyses which are not addressed directly at testing our initial hypotheses, but help characterize the pattern of results and provide some reality checks:

(1) Overall cross-modal effects. We addressed a comparison between the neutral and target-consistent condition, and between the neutral and distractor consistent condition. This analysis shall be informative as to whether overall, congruent cross-modal cues benefit performance, incongruent cross-modal cues hinder performance, or both. Also, will help situate the present data with respect to previous studies finding cross-modal benefits in visual search. We ran two one-sided paired samples t-tests to compare mean RTs of target-consistent condition and distractor consistent condition to mean RTs of neutral condition. The results show a significant speed advantage of target-consistent condition over neutral condition ($t_{(59)} = 1.849$, $p = 0.033$, *Cohen's d*=0.16). The difference between mean RTs of distractor-consistent condition and neutral condition was, however, not significant ($t_{(59)} = 0.602$, $p = 0.275$) (Figure 3A). We ran two more one-sided paired samples t-tests, this time to compare *d'* of target-consistent condition and distractor consistent condition to the neutral condition. The results showed a marginal tendency for higher sensitivity in the target-consistent condition compared the neutral condition ($t_{(59)} = 1.405$, $p = 0.083$, *Cohen's d*=0.26) (Figure 3B). Similar to the RT outcome, there was no significant difference between *d'* of distractor-consistent condition and neutral condition ($t_{(59)} = 0.266$, $p = 0.395$). Altogether, these results indicate that congruent cross-modal cues tend to benefit performance, but incongruent cross-modal cues do not hinder performance, with respect to neutral conditions. These results replicate recent findings from Kvasova & Soto-Faraco (2019) using naturalistic scenes, and are in line with previous cross-modal findings using more traditional laboratory tasks (Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008; Knoeferle, Knoeferle, Velasco, & Spence, 2016; Laurienti et al., 2003).

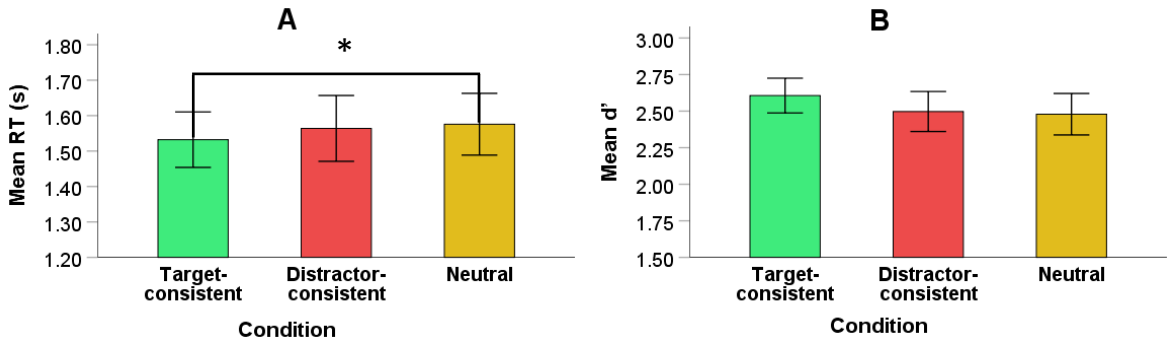

Figure 3. (A) Average reaction times and **(B)** d prime in each of the three experimental conditions separately. Error bars indicate the standard error of the mean. Significant differences are indicated by asterisks (*p-value<0.05).

(2) Gender-dependent effects. Given that we collected data from a gender-balanced sample for both AVGP and NVGP groups, we decided to perform exploratory analyses of the two main hypotheses, considering gender as a between subject variable. Relating to the first hypothesis (overall advantage of AVGPs over NVGPs), the result of a group (AVGP/NVGP) by gender (male/female) between-subject ANOVA on mean RTs revealed no group*gender interaction or main effect of gender ($F_{(1, 56)} = 0.98$, $p = 0.32$; $F_{(1, 56)} = 0.04$, $p = 0.85$ respectively), whilst group had a significant main effect ($F_{(1, 56)} = 6.72$, $p < 0.05$, $\eta 2=0.11$), as it would be expected from the main analysis. We ran another group*gender between-subject ANOVA on *d'* scores. The result indicated no significant group*gender interaction ($F_{(1, 56)} = 0.19$, $p = 0.67$) nor main effect of group ($F_{(1, 56)} = 0.43$, $p = 0.51$), while the main effect of gender was significant ($F_{(1, 56)} = 6.79$, $p < 0.05$, $\eta 2=0.11$). As it can be seen in the descriptive data in table 1, total *d'* scores for males (M = 2.78, SD = .56) was higher than for females (M = 2.45, SD = .42) (see Figure 4).

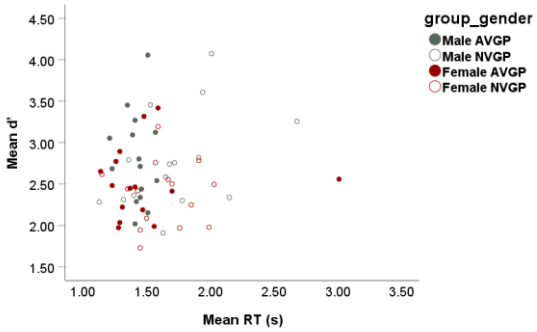

Figure 4. Individual d' scores plotted against average RTs pooled across all three sound conditions for each participant. Male participants in gray, female in red; Full symbols represent AVGPs and empty symbols NVGPs.

Related to our second hypothesis (differential cross-modal advantage for AVGPs), we entered gender and video-game experience as between-participant's factors and condition as a within participant variable in an ANOVA on normalized RTs. This ANOVA rendered a nearly significant interaction for group*condition*gender ($F_{(1, 56)} = 3.07$, $p = 0.085$). Owing to this marginal effect, we broke down the analysis in two separate ANOVAS, one for each gender, seeking for a possible group by condition interaction. Neither the ANOVA for male ($F_{(1, 28)} = 2.58$, $p = 0.12$) (Figure 5A) nor for female ($F_{(1, 28)} = 0.79$, $p = 0.38$) (Figure 5B) were significant in isolation. This is parallel to the result of the main analysis, reported above. We performed some follow up analyses by running a repeated measures ANOVA on normalized RTs just for AVGPs, using gender as a factor. We found a significant interaction between gender and sound condition on normalized RTs ($F_{(1, 57)} = 5.84$, $p = 0.022$) (Figure 5C). Please note that the main effect of gender would be meaningless in this analysis, as the RTs are normalized, but interactions can be interpreted. Post hoc t-tests show a significant difference between male AVGPs and female AVGPs in the distractor-consistent condition ($t_{(28)} = 2.80$, $p = 0.009$, *Cohen's d*=1.0) while there was not such difference between male NVGPs and female NVGPs ($t_{(28)} = 1.47$, $p = 0.152$). The result would suggest that there was an interference effect for male AVGPs compared to female AVGPs, in way that distractor sounds slowed down their reaction times. Given the low statistical power of this exploratory analysis, this result should be taken cautiously.

By entering gender as factor in the repeated measures ANOVA gender*group*condition on *d'* scores, we found a significant interaction for condition and gender ($F_{(2, 55)} = 3.51$, $p = 0.037$) and a significant main effect of gender ($F_{(1, 59)} = 11.34$, $p = 0.001$). Follow up analyses showed significant differences between male AVGPs and female AVGPs in *d*' for the distractor-consistent condition ($t_{(28)} = 2.91$, $p = 0.007$, *Cohen's d*=1.08)

(Figure 5D) and between male NVGPs and female NVGPs in *d'* in the neutral condition ($t_{(28)}$ = 3.488, p = 0.002, *Cohen's d*=1.27) (Figure 5E). When we analyzed all sensitivity data for gender differences, the differences between male and female in d' of distractor-consistent condition ($t_{(58)}$ = 2.966, p = 0.004, *Cohen's d*=0.77) and neutral condition ($t_{(58)}$ = 3.667, p = 0.001, *Cohen's d*=0.97) were significant (Figure 5F). These findings suggest that males in general, provided more precise responses in distractor-consistent and neutral condition than females. Again, these outcomes must be carefully interpreted, given that the study was underpowered in terms of investigating interactions with the gender variable.

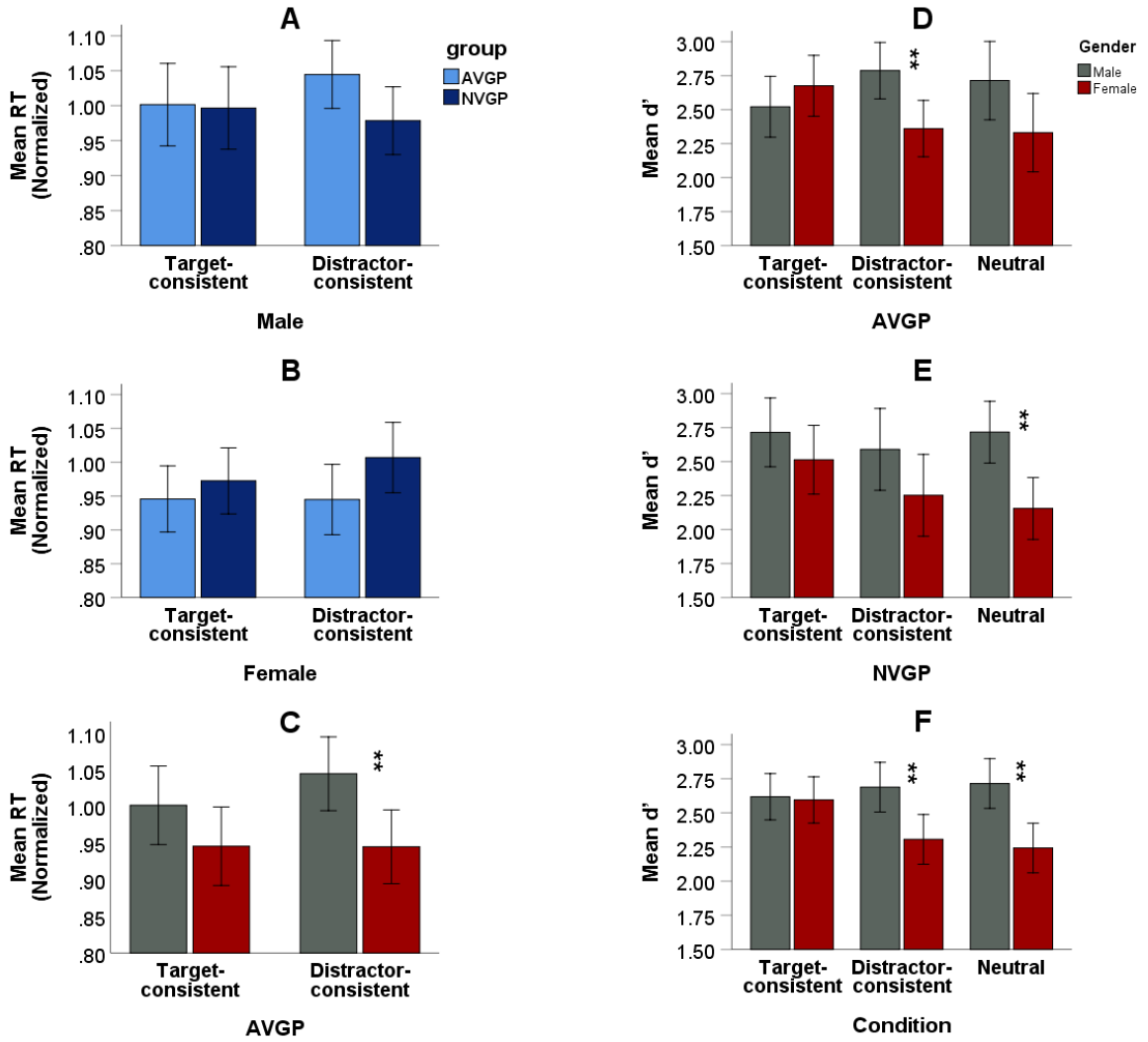

Figure 5. (A) Normalized reaction times separated by group in the target- and distractor-consistent conditions were plotted for males and **(B)** females. **(C)** Normalized reaction times in the target- and distractor-consistent conditions were plotted for males and female AVGPs. **(D)** Visual search accuracy (d prime) for target-consistent, distractor-consistent, and neutral conditions plotted for male and female AVGPs, and **(E)** for male and female NVGPs. **(F)** D primes separately for each gender in each of the three sound conditions. Error bars indicate standard error of the mean, and significant differences are indicated by asterisks (*p-value<0.05, **p-value<0.01).

Discussion

The first aim of the present study was to assess whether the visual search advantage demonstrated by AVGPs in comparison to NVGPs with simple stimuli in classic psychophysical tasks would generalize to a real-life multisensory complex situation. Using a visual search task in real-life scenes, with audio-visual congruent/incongruent cross-modal cues, we demonstrated that AVGPs can extend their advantage to real-life like scenarios by a faster response time while they keeping precision equal to NVGPs. In the line with previous studies on multisensory advantages of AVGP experience (Joseph D. Chisholm & Kingstone, 2012, 2015a; Donohue et al., 2010; Stewart, Martinez, Perdew, Green, & Moore, 2019) our results have revealed that advantage of extensive action video game experience can transfer to naturalistic multisensory situations. While the mechanism of this generalization still

remains debated, one explanation is that AVGPs may have a heightened speed of processing (e.g., M. W. Dye, C. S. Green, & D. Bavelier, 2009). Action video game playing requires rapid processing of audio-visual information and promotes action per unit of time, forcing players to make decisions and execute responses as fast as possible. Each speed processing which ends up with a speed accurate response provides an incentive for players while delays in processing often have severe consequences. Therefore, extensive experience with action video games may lead to more efficient visual processing and speeded RTs across a range of unrelated tasks without sacrificing accuracy. Most past studies have compared AVGPs to NVGPs by measuring RTs have consistently shown that AVGPs are faster overall than NVGP (Castel et al., 2005; Greenfield et al., 1994; Nuyens, Kuss, Lopez-Fernandez, & Griffiths, 2019; Powers et al., 2013; Schubert et al., 2015; Stewart et al., 2019; Torner, Carbonell, & Castejón, 2019). Please note that if playing merely induced faster decision making, without increasing information processing efficiency, it would typically come at the cost of accuracy (Heitz, 2014). This does not seem to be the case in most findings.

Another explanation for the advantage of AVGPs would be greater attentional control capacity, compared to NVGPs (e.g., C Shawn Green & Bavelier, 2006), allowing them to direct their attention toward the spatial position of the visual stimulus more quickly and accurately, which in turn allows them to include other modality inputs when they are consistent with goal-directed visual search and exclude other irrelevant inputs from processing. It has been suggested that attentional control plasticity as a result of action video gameplay is a fundamental building block of cognitive enhancement in AVGPs (Bavelier & Green, 2019). Enhancement of various aspects of top-down attention as an effect of action video gameplay is seen in cross-sectional studies (Bediou et al., 2018; Joseph D. Chisholm & Kingstone, 2012, 2015b). Similarly, intervention studies mirrored the analysis of cross-sectional studies by showing that action video game training can radically alter visual attentional processing. The results of these studies suggest a causal role of action gaming on spatial cognition and top-down attention, since they rule out inherently enhanced attentional control in AVGPs (Bediou et al., 2018; Chiappe, Conger, Liao, Caldwell, & Vu, 2013; C Shawn Green & Bavelier, 2003; C Shawn Green & Bavelier, 2012). Furthermore, training studies make a distinction between action video games and other video game types such as mimetic games, sport games, simulation games or puzzle games because they could not find the same impact for other types of games (Cohen, Green, & Bavelier, 2007; Powers & Brooks, 2014; Powers et al., 2013).

Given the online cross-sectional design of the current study, it provides evidence that the advantages associated with action video game experience extend beyond typical laboratory tasks to more complex naturalistic scenes. Like in many other studies before, a degree of caution must be advised when considering the causal relationship between action video game experience and improved visual search in real-life multisensory context, since a training component was not 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 cross-modal advantage for AVGPs than NVGPs. By entering the normalized data in the analysis of performance in the visual search task in real-life scenarios, the results suggest that AVGPs do not benefit from cross-modal cues more than NVGPs, nor they are more resistant to distractors within the statistical power of the present study, sensitive to medium to large effect sizes. These results are consistent with the previous studies (Gao et al., 2018; Rupp et al., 2016) showing that there is no difference between AVGPs and NVGPs in the ability to

integrate audio-visual stimuli or in the driving performance while distracted in a driving simulator. Yet, AVGPs display greater performance in some cross-modal tasks, such as discrimination (temporal order judgments, TOJs) for visual and auditory stimuli (Donohue et al., 2010). Please note that unlike in our task, TOJs require for effective segregation (rather than integration) of cross modal events. Because processing of the meaning of a complex naturalistic sound can require more time due to the temporal nature of the information (for a similar procedure Kvasova et al., 2019; and for a review see Vatakis & Spence, 2010), in the task used in our study, the sound 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. Overall, this cross-modal semantic congruence benefited performance across the board, but it did so to the same extend in AVPGs and NVPGs. Overall, this pattern of results suggests that although AVGPs are faster in overall visual search, they do not benefit more from semantic audio-visual integration to direct their attention toward the target object than NVGPs. One explanation for this is that AVGPs and NVGPs may employ similar cognitive strategies when they need to process the information at the higher-level, semantic aspects. Hence, the AVGP advantage may be based on early, lower-level processing stages. It has been shown that AVGPs benefit from low-level spatial and temporal factors in their visual (C Shawn Green & Bavelier, 2006; Greenfield et al., 1994; Schubert et al., 2015; Wong & Chang, 2018; Xuemin & Bin, 2010), auditory (C. S. Green et al., 2010; Stewart et al., 2019), or audio-visual (Zhang, Tang, Moore, & Amitay, 2017) search tasks. A recent meta-analysis confirmed a smaller impact on higher cognitive performance like inhibition (g = 0.31, 95% CI [0.07, 0.56], df = 7.2, p = .02 and verbal cognition (g = 0.30, 95% CI [0.03, 0.56], df = 7.7 p = .033), while there was not a 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 cognitive functions or it is limited to low-level perception, spatial cognition, even if under the mediation of top-down attention.

From our exploratory analyses, we could reveal that overall semantic congruency between sounds and target objects speed up search latencies in comparison with neutral sounds while semantic incongruency in distractor-consistent condition did not produce any disadvantage with respect to the baseline (neutral condition). Hence, consistent cross-modal cues benefit performance, whilst incongruent cross-modal cues do not hinder performance to a measurable extent. This finding presents a replication of previous findings (Knoeferle et al., 2016; Kvasova et al., 2019), indicating that semantic congruency effect can generalize beyond typical laboratory protocols and guide attention in a complex, multisensory environment. Also, it provides an extension of simpler, laboratory protocols (Iordanescu et al., 2008; Laurienti et al., 2003).

The design of this study allows us to calculate false alarm rate and d prime for distractor-consistent condition which was a limitation in Kvasova's (2019) study. In keeping with the conclusions of that study, there was no difference in d' between distractor-consistent and neutral conditions. This indicates that the cross-modal congruence effects can be more reliably taken as an advantage of congruent sounds, than distractor effects, which neither slow down response time nor create a more impulsive response than neutral conditions. An open question here is: If cross-modal semantics are clearly being processed, given the facilitation, why semantic incongruency does not hinder performance? There are two possibilities to answer this question. One is that a target-inconsistent sounds do not strongly distract participants from responding to the visual target when visual stimuli are highly informative. In this study, when the cue word is presented to participants, it activates a semantic network related to the target and creates an attentional template to direct

participant's attention to other semantically congruent inputs from other modalities while inhibiting unrelated cross-modal information. If inputs come from other modalities that are incongruent with the target search and is easy to inhibit, it cannot enter in the processing to create a distractor effect. The second possibility is that in our study, some auditory stimuli may not be informative enough to direct participant's attention and affect visual search because some of them are physically similar (e.g., the sound of the coin and keys) or semantically similar (e.g., sound of musical instruments). Thus, we cannot be sure that they were powerful enough to play a proper role of a distractor or neutral sound. To clarify this point, further studies should be run on the effect of high informative (hard to inhibit) versus low informative (easy to inhibit) cross-modal cues.

The low prevalence of female AVGPs has typically caused a great gender imbalance in cross-sectional studies comparing AVGPs with NPVGs, which has been almost exclusively based on a male population (Cohen et al., 2007; C Shawn Green & Bavelier, 2003; C. S. Green et al., 2010; Adam C. Oei & Patterson, 2013). Here, we used a gender-balanced sample of participants, with equal number of males and females in both AVPG and NVPG groups. Therefore, despite the sample size was not calculated to detect possible gender effects, we believe that exploratory analyses with this variable may provide a first impression unavailable in previous studies. In terms of the putative general advantage of AVPGs in the search task overall (first hypothesis proposes), our results suggested that the main AVGP effect applies across the board. Playing AVGs enhances speed of processing or attentional control capacity while accuracy of responses (*d'* scores) will remain unchanged irrespective of gender, as far as the sensitivity of our design allows us to tell. Our results also hinted at a basic difference in strategy of responding between experienced male and female AVG players in dealing using cross-modal cues. Please note upfront, that these analyses are probably underpowered, and they will need general confirmation. However, what we observed in the analyses by gender is that female AVGPs were faster in the distractor-consistent condition than male AVGPs, but male AVGPs provided more accurate responses, than female AVGPs. If this finding were confirmed, it may be interpreted as slightly different strategies. The increased speed of processing noted in female AVGPs in distractor-consistent condition can be viewed as impulsive behavior, in which female AVGPs respond faster but make more errors, compared to males who sacrifice time for precision (see figure 4). With the due cautiousness, given the exploratory nature of this analysis, this finding could be in line with some previous studies suggesting that there is a gender difference in selective attention and spatial cognition (Evans & Hampson, 2015; Halpern, 2013; Merritt et al., 2007; Posner & Marin, 2016; Stoet, 2017). More studies, with larger samples and with the adequate statistical should be performed to follow up this result. For now, one conclusion that can be drawn from this study is that there could be potential for studying gender differences in the effects of AVGP, an area that has been neglected in most studies so far.

Perhaps it is worth mentioning that this study was performed online with the experiment presented in a browser at the participant's personal device (personal computer or laptop). This adds to the generalization of our findings to more ecological situations. Despite our efforts, probably the auditory stimuli and video clips quality, size, and luminance have been varied between participants. It is our assumption that this variation has been similar between the groups (AVGPs vs. NVGPs and men vs, women). There is reason to assume that such group differences were negligible because our data replicated the finding of the cross-modal semantic effect, very similar to the laboratory-based study of Kvasova et al. (2019). However, more works in other natural contexts (e.g., virtual reality or other situations that do not involve sitting in front of a display) are needed to evaluate the extent to which the observed advantages can be generalized.

Conclusion

The results of the present investigation demonstrate that AVGP advantage in attention tasks can be extended to naturalistic multisensory situations. We also failed to demonstrate that AVGP experience creates any specific advantage to benefit more from cross-modal cues, or more resistance to distractors, compared than NVGPs. Our findings from exploratory analyses shed some light on semantic aspects of multisensory integration by generalizing (and confirm) previous laboratory findings on semantic congruency/incongruency effect on cross-modal interactions. Other findings from exploratory analyses with gender indicate that despite the AVPG advantage applies across the board, there could be strategic differences regarding how this advantage is achieved. To gain a full understanding about AVG advantages in males and females it will be important to include female participants in this field of study and directly compare results between genders more systematically.

Funding

Salvador Soto-Faraco has been funded by Ministerio de Ciencia e Innovación (Ref: PID2019-108531GB-I00 AEI/FEDER), and AGAUR Generalitat de Catalunya (2017 SGR 1545). This project has been co-funded with 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€.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

References

Bavelier, D., & Green, C. S. (2019). Enhancing Attentional Control: Lessons from Action Video Games. *Neuron, 104*(1), 147-163. doi:10.1016/j.neuron.2019.09.031

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychol Bull, 144*(1), 77-110. doi:10.1037/bul0000130

Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica, 119*(2), 217-230.

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*(3), 389-404.

Chen, Y.-C., Yeh, S.-L., & Spence, C. (2011). Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry. *Frontiers in psychology, 2*, 212.

Chiappe, D., Conger, M., Liao, J., Caldwell, J. L., & Vu, K. P. (2013). Improving multi-tasking ability through action videogames. *Appl Ergon, 44*(2), 278-284. doi:10.1016/j.apergo.2012.08.002

Chisholm, J. D., Hickey, C., Theeuwes, J., & Kingstone, A. (2010). Reduced attentional capture in action video game players. *Attention, Perception, & Psychophysics, 72*(3), 667-671.

Chisholm, J. D., & Kingstone, A. (2012). Improved top-down control reduces oculomotor capture: The case of action video game players. *Attention, Perception, & Psychophysics, 74*(2), 257-262. doi:10.3758/s13414-011-0253-0

Chisholm, J. D., & Kingstone, A. (2015a). Action video game players' visual search advantage extends to biologically relevant stimuli. *Acta Psychologica, 159*, 93-99. doi:https://doi.org/10.1016/j.actpsy.2015.06.001

Chisholm, J. D., & Kingstone, A. (2015b). Action video games and improved attentional control: Disentangling selection- and response-based processes. *Psychonomic Bulletin & Review, 22*(5), 1430-1436. doi:10.3758/s13423-015-0818-3

Clark, K., Fleck, M. S., & Mitroff, S. R. (2011). Enhanced change detection performance reveals improved strategy use in avid action video game players. *Acta Psychologica, 136*(1), 67-72.

Cohen, J. E., Green, C. S., & Bavelier, D. (2007). Training visual attention with video games: not all games are created equal. In H. F. O'Neil, H. O'Neil, & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp. 205–227). Amsterdam, the Netherlands: Elsevier Group Inc.

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

Cox, D., & Hong, S. W. (2015). Semantic-based crossmodal processing during visual suppression. *6*(722). doi:10.3389/fpsyg.2015.00722

Donohue, S. E., Woldorff, M. G., & Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities. *Atten Percept Psychophys, 72*(4), 1120-1129. doi:10.3758/app.72.4.1120

Ducrocq, E., Wilson, M., Vine, S., & Derakshan, N. (2016). Training Attentional Control Improves Cognitive and Motor Task Performance. *Journal of Sport and Exercise Psychology, 38*, 1-36. doi:10.1123/jsep.2016-0052

Durlach, P. J., Kring, J. P., & Bowens, L. D. (2009). Effects of action video game experience on change detection. *Military Psychology, 21*(1), 24-39. doi:10.1080/08995600802565694

Dye, M. W., Green, C. S., & Bavelier, D. (2009). The development of attention skills in action video game players. *Neuropsychologia, 47*(8-9), 1780-1789.

Dye, M. W., Green, C. S., & Bavelier, D. (2009). Increasing Speed of Processing With Action Video Games. *Current directions in psychological science, 18*(6), 321-326. doi:10.1111/j.1467-8721.2009.01660.x

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). The development of attention skills in action video game players. *Neuropsychologia, 47*(8), 1780-1789. doi:https://doi.org/10.1016/j.neuropsychologia.2009.02.002

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science, 11*(1), 19-23.

Evans, K. L., & Hampson, E. (2015). Sex-dependent effects on tasks assessing reinforcement learning and interference inhibition. *Frontiers in psychology, 6*, 1044. doi:10.3389/fpsyg.2015.01044

Föcker, J., Mortazavi, M., Khoe, W., Hillyard, S. A., & Bavelier, D. (2019). Neural correlates of enhanced visual attentional control in action video game players: An event-related potential study. *J. Cognitive Neuroscience, 31*(3), 377–389. doi:10.1162/jocn_a_01230

Gao, Y.-L., Zhang, S.-H., Zhang, Y., Wang, M.-X., Li, Y.-X., Wang, Z.-N., . . . 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.

Green, C., Gorman, T., & Bavelier, D. (2016). Action Video-Game Training and Its Effects on Perception and Attentional Control. In (pp. 107-116).

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534-537.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465-1478. doi:10.1037/0096-1523.32.6.1465

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology: Cell Press, 22*(6), R197-R206. doi:10.1016/j.cub.2012.02.012

Green, C. S., & Bavelier, D. (2015). Action video game training for cognitive enhancement. *Current Opinion in Behavioral Sciences, 4*, 103-108. doi:https://doi.org/10.1016/j.cobeha.2015.04.012

Green, C. S., Kattner, F., Eichenbaum, A., Bediou, B., Adams, D. M., Mayer, R. E., & 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*(5), 679-682. doi:10.1177/0956797616644837

Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. *Curr Biol, 20*(17), 1573-1579. doi:10.1016/j.cub.2010.07.040

Greenfield, P. M., DeWinstanley, P., Kilpatrick, H., & Kaye, D. (1994). Action video games and informal education: Effects on strategies for dividing visual attention. *Journal of Applied Developmental Psychology, 15*(1), 105-123. doi:https://doi.org/10.1016/0193-3973(94)90008-6

Halpern, D. F. (2013). *Sex differences in cognitive abilities*: Psychology press.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience, 27*(30), 7881-7887.

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

Hisam, A., Mashhadi, S. F., Faheem, M., Sohail, M., Ikhlaq, B., & Iqbal, I. (2018). Does playing video games effect cognitive abilities in Pakistani children? *Pak J Med Sci, 34*(6), 1507-1511. doi:10.12669/pjms.346.15532

Hsiao, J.-Y., Chen, Y.-C., Spence, C., & Yeh, S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Consciousness and Cognition, 21*(2), 775-787.

Hubert-Wallander, B., Green, C. S., Sugarman, M., & Bavelier, D. (2011). Changes in search rate but not in the dynamics of exogenous attention in action videogame players. *Attention, Perception, & Psychophysics, 73*(8), 2399-2412.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review, 15*(3), 548-554.

Irons, J. L., Remington, R. W., & McLean, J. P. (2011). Not so fast: Rethinking the effects of action video games on attentional capacity. *Australian Journal of Psychology, 63*(4), 224-231.

Knoeferle, K. M., Knoeferle, P., Velasco, C., & Spence, C. (2016). Multisensory brand search: How the meaning of sounds guides consumers' visual attention. *Journal of Experimental Psychology: Applied, 22*(2), 196-210. doi:10.1037/xap0000084

Kvasova, D., Garcia-Vernet, L., & Soto-Faraco, S. (2019). Characteristic Sounds Facilitate Object Search in Real-Life Scenes. *Frontiers in psychology, 10*(2511). doi:10.3389/fpsyg.2019.02511

Kvasova, D., & 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

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human brain mapping, 19*(4), 213-223.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nat Neurosci, 12*(5), 549-551. doi:10.1038/nn.2296

List, A., Iordanescu, L., Grabowecky, M., & Suzuki, S. (2014). Haptic guidance of overt visual attention. *Atten Percept Psychophys, 76*(8), 2221-2228. doi:10.3758/s13414-014-0696-1

Mastroberardino, S., Santangelo, V., & Macaluso, E. (2015). Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks. *Frontiers in Integrative Neuroscience, 9*(45). doi:10.3389/fnint.2015.00045

Merritt, P., Hirshman, E., Wharton, W., Stangl, B., Devlin, J., & Lenz, A. (2007). Evidence for gender differences in visual selective attention. *Personality and Individual Differences, 43*(3), 597-609. doi:https://doi.org/10.1016/j.paid.2007.01.016

Nuyens, F. M., Kuss, D. J., Lopez-Fernandez, O., & Griffiths, M. D. (2019). The Empirical Analysis of Non-problematic Video Gaming and Cognitive Skills: A Systematic Review. *International Journal of Mental Health and Addiction, 17*(2), 389-414. doi:10.1007/s11469-018-9946-0

Oei, A. C., & Patterson, M. D. (2013). Enhancing Cognition with Video Games: A Multiple Game Training Study. *PLOS ONE, 8*(3), e58546. doi:10.1371/journal.pone.0058546

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

Pesquita, A., Brennan, A. A., Enns, J. T., & Soto-Faraco, S. (2013). Isolating shape from semantics in haptic-visual priming. *Exp Brain Res, 227*(3), 311-322. doi:10.1007/s00221-013-3489-1

Posner, M. I., & Marin, O. S. (2016). *Attention and performance XI*: Routledge.

Powers, K. L., & Brooks, P. J. (2014). Evaluating the Specificity of Effects of Video Game Training. In F. C. Blumberg (Ed.), *Learning by Playing: Video Gaming in Education* (pp. 302–330). Oxford, UK: Oxford University Press.

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review, 20*(6), 1055-1079. doi:10.3758/s13423-013-0418-z

Roberts, K. L., & 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. *Journal of cognitive neuroscience, 20*(6), 1063-1078.

Rupp, M. A., McConnell, D. S., & Smither, J. A. (2016). Examining the Relationship Between Action Video Game Experience and Performance in a Distracted Driving Task. *Current Psychology, 35*(4), 527-539. doi:10.1007/s12144-015-9318-x

Sauter, M., Draschkow, D., & Mack, W. J. B. S. (2020). Building, Hosting and Recruiting: A Brief Introduction to Running Behavioral Experiments Online. *10*(4), 251.

Schubert, T., Finke, K., Redel, P., Kluckow, S., Müller, H., & 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 Psychologica, 157*, 200-214. doi:https://doi.org/10.1016/j.actpsy.2015.03.005

Soto-Faraco, S., Kvasova, D., Biau, E., Ikumi, N., Ruzzoli, M., Morís-Fernández, L., & Torralba, M. (2019). *Multisensory Interactions in the Real World*. In Elements in Perception. Retrieved from https://www.cambridge.org/core/elements/multisensory-interactions-in-the-real-world/549C5120543E2298E1295696FBC39F3E doi:DOI: 10.1017/9781108578738

Spagna, A., Mackie, M. A., & Fan, J. (2015). Supramodal executive control of attention. *Frontiers in psychology, 6*, 65. doi:10.3389/fpsyg.2015.00065

Stewart, H., Martinez, J., Perdew, A., Green, S., & Moore, D. (2019). *Auditory cognition and perception of action video game players*.

Stoet, G. (2017). Sex differences in the Simon task help to interpret sex differences in selective attention. *Psychological research, 81*(3), 571-581. doi:10.1007/s00426-016-0763-4

Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychol (Amst), 140*(1), 13-24. doi:10.1016/j.actpsy.2012.02.001

Tomczak, M., Tomczak, E., Kleka, P., & Lew, R. (2014). Using power analysis to estimate appropriate sample size. *Trends in Sport Sciences, 21*(4), 195-206.

Torner, H. P., Carbonell, X., & Castejón, M. (2019). A comparative analysis of the processing speed between video game players and non-players. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport, 37*(1).

Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In *Multisensory object perception in the primate brain.* (pp. 95-121). New York, NY, US: Springer Science + Business Media.

Wong, N. H. L., & Chang, D. H. F. (2018). Attentional advantages in video-game experts are not related to perceptual tendencies. *Scientific Reports, 8*(1), 5528. doi:10.1038/s41598-018-23819-z

Xuemin, Z., & Bin, Y. (2010). Effects of action video game on spatial attention distribution in low and high perceptual load task. *Journal of Software, 5*(12), 1434-1441.

Zhang, Y.-X., Tang, D.-L., Moore, D. R., & Amitay, S. (2017). Supramodal Enhancement of Auditory Perceptual and Cognitive Learning by Video Game Playing. *8*(1086). doi:10.3389/fpsyg.2017.01086