

REGISTERED REPORT

Effects of Action Video Game Play on Cognitive Skills: A Meta-Analysis





Benoit Bediou^{1, 2}, Melissa A. Rodgers^{3, 4}, Elizabeth Tipton⁵, Richard E. Mayer⁶,
C. Shawn Green⁷, and Daphne Bavelier^{1, 2}¹ Faculty of Psychology and Education Science, University of Geneva² Fondation Campus Biotech, Geneva, Switzerland³ Department of Educational Psychology, The University of Texas at Austin⁴ American Institutes for Research, Arlington, Virginia, United States⁵ Department of Statistics and Data Science, Northwestern University⁶ Department of Psychological and Brain Sciences, University of California, Santa Barbara⁷ Department of Psychology, University of Wisconsin-Madison

The possibility that playing action video games is associated with enhancements in certain aspects of cognitive function has attracted significant interest from researchers in education, psychology, and neuroscience. Previous meta-analyses indicated an overall positive relationship between action video game play and cognitive skills. However, follow-up to this previous work is warranted, not only because the amount of data has grown significantly since previous meta-analyses were conducted, but also because previous work left several issues unresolved (e.g., certain meta-analytic procedures). We conducted a literature search using predefined keywords and inclusion criteria to identify studies that examined the relationship between action video game play and cognitive skills. Data from (a) 105 cross-sectional studies (221 effect sizes) and (b) 28 intervention studies with an active control group (91 effect sizes) were analyzed separately via meta-analytic models for dependent effect sizes with robust variance estimates for correlated and hierarchical effects (CHE) and small-sample corrections. Consistent with our hypotheses, action video game players outperformed nonvideo game players in the cross-sectional meta-analysis (large effect, $g = 0.64$, 95% CI [0.53, 0.74]). Action video game play was causally related to improvements in cognitive skills in the intervention meta-analysis (small effect, $g = 0.30$, 95% CI [0.11, 0.50]). Publication bias was detected in the cross-sectional data set, with sensitivity analysis showing high heterogeneous estimates of the average unbiased effect. Publication bias was not detected in the intervention data set, but sensitivity analyses also point to heterogeneity. No significant moderators were found for either data set; however, this may be limited by small sample sizes.

Keywords: action video games, cognitive skills, perception, attention, meta-analysis**Supplemental materials:** <https://doi.org/10.1037/tmb0000102.supp>

There is a great deal of current interest in the extent to which playing *action video games* (AVG) is associated with and/or directly causes enhancements in cognitive function. Within this single statement there are three main topics that must be unpacked in order to effectively set the stage for the present work.

First, why is there particular interest in *action video games*? Here, although it is unfortunately commonplace in the popular media (and even parts of the scientific literature) to treat video games as a unitary construct, all video games are not equal (Powers & Brooks, 2014). As such, lumping all video games together when considering their

Action Editor: Nicholas David Bowman was the action editor for this article.**ORCID iDs:** Benoit Bediou  <https://orcid.org/0000-0002-3477-7948>; Elizabeth Tipton  <https://orcid.org/0000-0001-5608-1282>; C. Shawn Green  <https://orcid.org/0000-0002-9290-0262>; Daphne Bavelier  <https://orcid.org/0000-0002-5904-1240>.**Acknowledgments:** The authors are thankful to Claire Holman, Elisa Gallerie, and Hanna Gohlke for their help with the literature search, review, coding, and data extraction processes.**Funding:** This research was supported by funding from the Office of Naval Research Global Grant N00014-18-1-2633 and Office of Naval Research Grant N00014-20-1-2074 and from the European Research

Council Grant ERC-2018-SyG and the Swiss National Fund (schweizerischer nationalfonds zur förderung der wissenschaftlichen forschung) Grant 100014_178814 awarded to Daphne Bavelier. Funding was partially provided by Grants N00014-17-1-2049 and N00014-22-1-2283 from the Office of Naval Research awarded to C. Shawn Green and Grants N000141612046 and N000142112047 from the Office of Naval Research awarded to Richard E. Mayer.

Disclosures: Daphne Bavelier is a founder and scientific advisor to Akili Interactive Inc. Boston.**Author Contributions:** Benoit Bediou coordinated all aspects of the project including design, data collection, analyses, and write-up. Melissa A. Rodgers helped with the analyses and the write-up of the article.*continued*

effects on the brain and behavior would be roughly equivalent to lumping together all types of food when considering relationships with body composition. Action video games are of primary interest in the field because the elements inherent in such games (e.g., enemies that can appear anywhere within a large field of view, cluttered visual scenes, many quickly, and independently moving targets) are expected to place heavy load on certain perceptual, attentional, and cognitive skills in a way that is not true of most other video games genres. Specifically, for the purposes of this article, we utilize the term action video games to encompass first- and third-person shooter video games, as this has been the dominant definition in the literature to date. As has been noted in a great deal of recent work (Dale et al., 2020; Dale & Green, 2017), many game genres beyond just first- and third-person shooters have, over the past 20 years, incorporated at least some action-like characteristics into their gameplay (e.g., action role playing video games or action adventure video games that combine elements of classic role-playing or adventure video games with certain first- or third-person shooter mechanics; see Table 1). Yet, the extent to which this is the case varies significantly from game to game within these other genres, and there is not yet consensus in the field with respect to how to treat these newer genres with respect to the longer term action video game literature (Choi et al., 2020). Thus, to ensure that we are truly comparing apples to apples and that we are using definitions that most closely align with the modal definition utilized in the field to-date, we chose to utilize the more stringent classification reflected in the top line of Table 1. This point is discussed further in our inclusion/exclusion criteria below.

Second, why is there interest in whether being a regular player of such action video games is *associated with* differences in cognitive skills? This question, which is typically addressed via cross-sectional designs contrasting individuals who naturally choose to play action video games with individuals who naturally engage in little video gaming, has both theoretical implications (e.g., about motivated behavior and why players engage with some, but not other games) as well as practical implications (e.g., with respect to identifying individuals who may excel in certain Esports).

Third, why is there interest in whether playing action video games *directly causes* enhancements in cognitive skills? This question, which is typically examined via intervention designs contrasting an experimental group trained on an action video game and a control group trained on a control video game, addresses one of the core issues in cognitive training—that of generalization/transfer of learning. Indeed, while there are many examples of behavioral training paradigms that serve to improve performance on the exact experiences the training calls for, it is rare to observe benefits that extend beyond the experienced context, such as performance on untrained cognitive and/or academic tasks. We have proposed that the game




mechanics inherent in action video games may be ideally suited to train core cognitive functions, such as attentional control, and in turn facilitate transfer to nongame contexts (learning to learn) at least in the cognitive domain (Bavelier et al., 2012).

The existing literature to-date points toward generally positive associations between action video game play and higher levels of cognitive performance (i.e., in cross-sectional/correlational work) as well as toward a positive causal relation existing between playing action video games and improvements in cognitive performance. For instance, previous meta-analyses that have considered all games together (Powers et al., 2013; Sala et al., 2018) have largely reached similar overall conclusions including that (a) video game play is associated with differences in cognition, (b) stronger effects are found in cross-sectional work compared to interventions,¹ and (c) the effects are complex and vary across both types of games and cognitive skills. This last issue is particularly key, as meta-analyses specifically examining the relationship between action video game play and cognitive skills remain rare (Bediou et al., 2018; Hilgard et al., 2019; Mayer, 2014; Powers & Brooks, 2014; Powers et al., 2013; Wang et al., 2016). Meta-analyses that have focused on the action video game genre (Bediou et al., 2018; Wang et al., 2016) indicate that frequent players of action video games consistently outperform individuals who seldom play video games in a number of cognitive skills ($g = 0.55$; Bediou et al., 2018). Meta-analyses of action video game interventions show that training with an action video game results in numerically smaller (in comparison to cross-sectional differences), but still significant improvements in cognitive skills as indicated by a Hedges' $g = 0.33$ in the meta-analysis by Bediou et al. (2018) and a Cohen's $d = 0.58$ in the meta-analysis by Wang et al. (2016). These values are close to the estimates obtained from subgroup analyses focusing at least partially on what we refer to as action video games in other less-selective meta-analyses. For example, Powers et al. (2013) found that the cognitive effects associated with action/violent video games (which involved mostly, but not exclusively action video games) were large in quasi-experiments ($g = 0.62$) but small in true experiments ($g = 0.22$). A similar effect ($g = 0.23$) was found for the category of first-person shooter games by Powers and Brooks (2014). Sala et al. (2018) found a Hedges' $g = 0.40$ for action video games in cross-sectional studies but only a marginal effect in intervention studies contrasting an experimental group trained with an action video game and a control group trained with a control video game training ($g = 0.10$, $p = .068$). We note here that the definition of action video game in this latter work was considerably more expansive

¹ Cohen's $d = 0.61$ for quasiexperiments and $d = 0.48$ for true experiments in Powers et al. (2013); Hedges' $g = 0.33$ for cross-sectional studies and $g = 0.07$ for interventions in Sala et al. (2018).

Elizabeth Tipton advised on all statistically related aspects of data analysis. Richard E. Mayer contributed to theoretical and conceptual aspects. C. Shawn Green and Daphne Bavelier contributed to design, analyses interpretation and write-up, review data extracted.

Open Science Disclosures:

-  The data are available at <https://osf.io/3xdh8/> (Bediou et al., 2020).
-  The experimental materials are available at <https://osf.io/3xdh8/>.
-  The preregistered design (transparent changes notation) is available at <https://osf.io/6qpye>.

Open Access License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC-BY-NC-ND). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Contact Information: Correspondence concerning this article should be addressed to Benoit Bediou, Faculty of Psychology and Education Science, University of Geneva, Boulevard Pont d'Arve 40, 1205 Geneva, Switzerland. Email: benoit.bediou@unige.ch

Table 1*Common Video Game Genres With Examples*

Video game genre	Example video game titles
First- or third-person shooters (FPS) (AVG for this work)	<i>Call Of Duty, Halo, Battlefield, Half-Life, Overwatch, Counterstrike</i>
Action role-playing games (RPG) or action-adventure games	<i>The Witcher, Mass Effect, Fallout 4, Skyrim, Grand Theft Auto, Assassin's Creed, Tomb Raider, The Last of Us</i>
Sports, driving	<i>Fifa, National Hockey League, Mario Kart, Need for Speed, Forza</i>
Real-time strategy (RTS) or multiple online battle arena (MOBA) games	<i>Starcraft, Warcraft I, II & III, DotA, Command & Conquer, League of Legends, Age of Empires</i>
Turn-based nonaction role-playing, or fantasy games	<i>World of Warcraft, Final Fantasy, Fable, Pokemon, Dragon Age</i>
Fighting games	<i>Mortal Kombat, Injustice, Marvel versus Capcom</i>
Turn-based strategy (TBS), life simulation, or puzzle games	<i>Civilization, Hearthstone, The Sims, Restaurant Empire, Puzzle Quest, Bejeweled, Solitaire, Candy Crush</i>
Music games	<i>Guitar Hero, DDR, Rock Band</i>
Other games	<i>Phone games, Browser games</i>

Note. We use the label “action” video games as only first- and third-person shooter video games and exclude other genres that frequently contain some degree of action characteristics (e.g., action-RPG, action-adventure, certain sports/driving games, real-time strategy games, and MOBAs). AVG = action video games; RPG = role-playing games; MOBA = multiple online battle arena; DDR = Danss Danse Revolution.

than what is utilized in the current work, including not only genres that might be considered “action-like” but arguably games that would not fall under even this broader label.

With respect to which domains of cognition are more or less impacted by action video game, Powers and Brooks (2014) found that training with a first-person shooter video game improved perceptual processing ($d = 0.45$) and spatial imagery ($d = 0.17$), but not motor skills ($d = 0.07, p = .7$) or executive functions ($d = -0.17, p = .3$). Top-down attention ($g = 0.31$) and spatial cognition ($g = 0.45$) were also seen to be improved following training with action video games in comparison to control video games in the intervention meta-analysis by Bediou et al. (2018). A somewhat similar pattern was obtained in the subgroup meta-analyses by Sala et al. (2018) who found that action video game training improved visual attention/processing ($g = 0.22$) when compared to control video game training, and spatial ability ($g = .12$) when compared to nonvideo game training control group who did not receive any video game training. No significant benefit was found for cognitive control or memory.

In sum, previous meta-analyses consistently point to action video game players (AVGP) demonstrating enhancements in a number of cognitive skills as compared to individuals who seldom play video games (Bediou et al., 2018; Hilgard et al., 2019; Mayer, 2014; Powers & Brooks, 2014; Powers et al., 2013; Wang et al., 2016), with the strongest advantages being observed for perception, top-down attention, and spatial cognition. In the case of intervention studies, individuals trained on action video games tend to show larger improvements in cognitive skills as compared to those trained on control video games. The intervention effect though is weaker than the cross-sectional effect.

Unfortunately, the intervention literature is also considerably less rich than the cross-sectional literature. Indeed, there are far fewer intervention studies than cross-sectional studies and intervention studies often include smaller sample sizes than cross-sectional work. Furthermore, a number of other issues also make our understanding of intervention studies results more tenuous. For example, methodological variations between intervention studies are vast, but due to the correlated nature of these design elements, it is difficult to isolate the impact of any particular type of methodological variation. For example, certain research groups tend to use rather long-training durations and strict recruitment criteria, while other groups tend to use shorter training durations and less strict

recruitment criteria. As a result, it is difficult to tease apart whether any differences are due to training duration, recruitment criteria, or something else inherent to the various research groups.

Finally, a host of complexities also make it difficult to assess publication bias—a major issue that threatens the validity of inferences from meta-analyses. If nonsignificant results are disproportionately not included in the meta-analysis because they are less likely to be published, the resulting inference about the totality of the field could be incorrect. Many meta-analytic procedures used to detect publication bias focus on funnel plot asymmetry (e.g., Egger’s regression, trim and fill). Yet, some of the factors discussed above could also produce significant funnel plot asymmetry. For instance, primary studies with smaller sample sizes could produce relatively larger effect sizes as a result of design-related factors (e.g., stricter inclusion criteria, longer training durations). Despite the known limitations of techniques for detecting and correcting publication bias, especially in the presence of few and dependent effect sizes, some authors have questioned the validity of previous meta-analytic findings due to concerns regarding possible publication bias (Hilgard et al., 2019), calling for an updated meta-analysis incorporating more recent work from the past 5 years. To this aim, we preregistered the present meta-analytic work in order to make all our methods and hypotheses transparent, and we share the data and code in order to facilitate future meta-analytic work in the field.

Here we explored whether action video game play is associated with greater levels of cognitive task performance (via a meta-analysis focused on cross-sectional studies) and whether action video game play causes enhancements in cognitive task performance (via a meta-analysis focused on intervention studies) using the most up-to-date data available (adding approximately 5 years of research as compared to our previous meta-analysis). Furthermore, we used new state of the art meta-analytic techniques that take into account the structure of the data that exists in the field (e.g., multiple effect sizes from the same study as well as multiple studies from the same research groups).

In brief, we expected to find (a) a significant positive average effect size for action video game play in both the cross-sectional and intervention meta-analyses, but that (b) the effect sizes will vary and will be particularly strong for perception, top-down attention, and spatial skills. We also expected to find that these effect sizes cannot be explained by publication bias alone.

Method

A detailed version of the preregistered protocol is available at <https://osf.io/6qpye>. All procedures and analyses proceeded exactly as per the preregistration except where explicitly noted as a deviation in the text below.

Search Strategy

Given the fact that the cognitive domain is quite broad, and that the areas of behavioral science where researchers are interested in the impact of video game play is quite varied, our literature search covered an extensive list of databases: PsycINFO, PsycARTICLES, APA books, Psyn dex, ERIC, MEDLINE, PubMed, Web of Science, ScienceDirect, and Google Scholar. Additional resources were reviewed to identify unpublished data (gray literature), including dissertations and theses, base-search, as well as searching through the abstracts from three annual conferences (Cognitive Neuroscience Society, Society for Neuroscience, and Vision Science Society). Authors were also contacted in an effort to identify possible unreported data. Finally, references from the retrieved studies as well as from reviews and meta-analyses on the topic were checked for additional references that were not gathered via the above methods.

Search keywords related to video game or cognitive skills were used (Table 2) and combined into the following Boolean expression: (“video game” OR “computer game”) AND (“attention” OR “attentional” OR “attend” OR “cognitive” OR “cognition” OR “perception” OR “perceptual”). When Boolean search was not permitted, we repeated the search with all possible combinations of keywords. Open ended terms were used whenever possible in order to cover all possible variants of each keyword (e.g., cognit*, attent*, attend*, percept*, perceiv*). We note that this search strategy purposely achieved a low precision (i.e., 0.01% in our previous meta-analysis, according to Gusenbauer & Haddaway, 2020). Our goal was to favor sensitivity over specificity and we therefore combined the use of (a) broad search keywords (to be as inclusive as possible in our literature search) with (b) thorough screening and filtering of retrieved studies using specific selection criteria (in order to identify more studies with particular predefined characteristics). Although more demanding, this strategy has the main advantage of reducing the risk of missing relevant work.

Table 2
Literature Search Terms

Video game	Perception attention cognition
Video game*	Perception, perceptual, percept*, perceiv*
Computer game*	Attention*, attend*
Gam*	Cognition, cognitive, cognit*

Note. To be indexed, studies need to mention at least one term from each column (e.g., video game AND cognition). Before preregistering this study, we conducted a separate search using keywords related to specific cognitive skills (executive function, inhibition, task switching, multitasking, verbal, spatial, problem-solving, motor control, working memory). Although this strategy increased the number of retrieved studies, it did not result in any new inclusions that had not already been captured with the term “cognition.”

* indicates open completion of the term (e.g., cognit * refers to both cognitive and cognition).

Inclusion and Exclusion Criteria

Our search covered the period from January 2000 to June 2020. We selected the January 2000 start date as roughly aligning with the rise in the popularity of home gaming consoles as first- or third-person shooters platforms, as well as to match with previous meta-analytic work. Again, as noted in the introduction we emphasize that for the current work, we use the term *action video games* to encompass only first- and third-person shooter video games. Accordingly, only studies reporting a measure of action video game experience in terms of hours per week (for cross-sectional studies) or training hours (for intervention studies) were considered eligible. The number of excluded studies can be found in Figure 1.

Participants

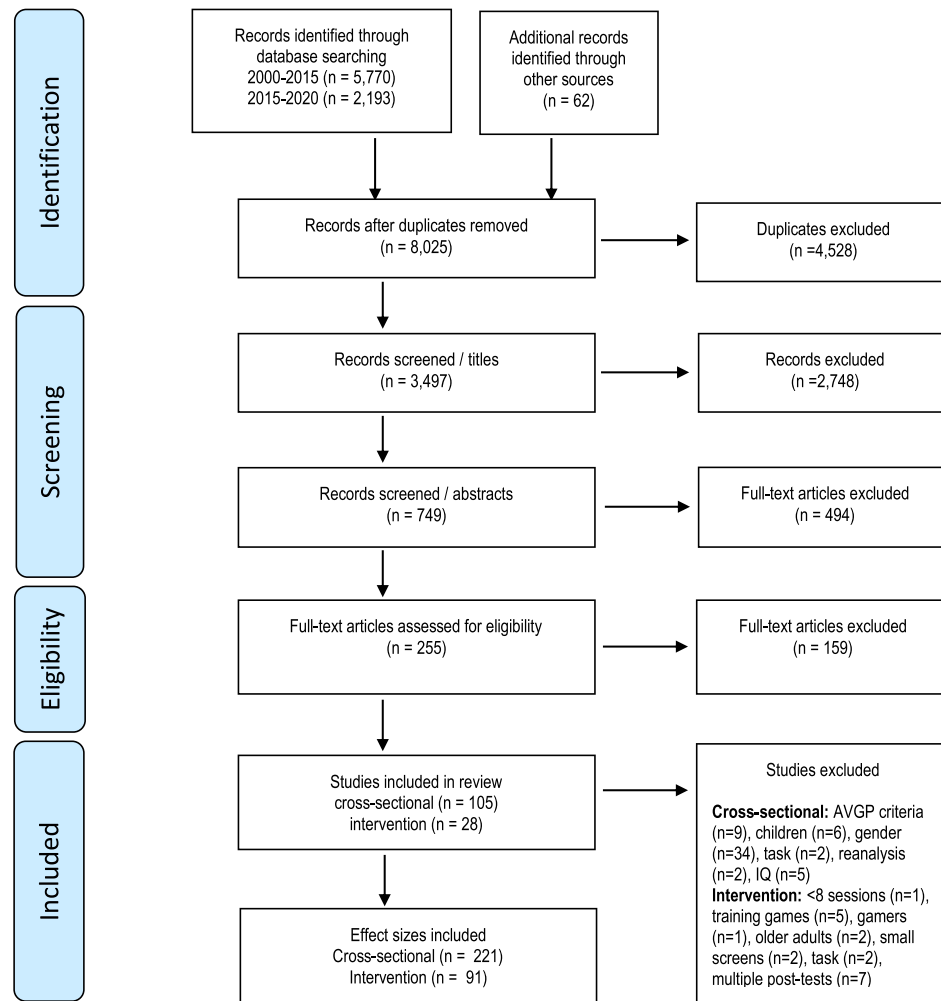
Only studies involving healthy young adults (i.e., without any diagnosed neurological, psychological or mental disorder) and including an objective measure of cognitive ability were included. Based on our previous meta-analysis, we decided to focus on healthy young adults because the literature on action video games is still largely dominated by studies (cross-sectional and intervention) involving this age group. Two additional considerations further supported this choice. First, among the very few studies conducted with children (below 18), all were cross-sectional in nature (probably because most action games are not suitable for children and thus inappropriate for experimental interventions) and relied on parental reports of their child’s video game experience, which may diverge from that of the child. Similarly, only very few studies have involved older adults, and there is strong reason to suspect that action video games designed for healthy young avid players are simply too difficult for older adults (let alone older adults with no gaming experience). For these reasons, the present meta-analysis focused on healthy young adults (i.e., without any diagnosed neurological, psychological, or mental disorder) at the peak of their cognitive abilities. Finally, to avoid possible confounds related to past video game experience, we excluded intervention studies that involved participants who could qualify as experienced video game players of any genre. We defined experienced video game players as individuals who play more than 3 hr per week of any specific video game genre and did so in the past 6 months.

Type of Studies: Cross-Sectional and Intervention

The two most common types of study designs were included in separate meta-analyses: (a) cross-sectional studies comparing cognitive abilities of individuals with high versus low prior action video game play experience and (b) intervention studies in which cognitive abilities are measured before and after training with an action video game versus a nonaction (control) video game.

Cross-Sectional Studies. Only cross-sectional studies involving healthy young adults (18–45 years old) were considered eligible. Studies contrasting individuals playing a minimum of 3 hr per week of action video games (i.e., AVGP) with individuals playing at most 1 hr per week of action video games and less than 3 hr of video gaming per week in total across all video game genres (i.e., nonvideo game players, NVGP) were considered for inclusion. Only outcomes related to one of the predefined measures of a cognitive ability (see below) were included. To reduce possible bias due to gender differences in cognitive abilities, cross-sectional studies in which

Figure 1
Flow Diagram of Selection Procedure



Note. For cross-sectional exclusions: AVGP criteria = the criteria utilized in the study for defining AVGP or NVGP did not match our inclusion criteria; children = children < 18 years old; gender = unmatched gender ratios; task = outside cognitive domain (e.g., affective stimuli, eye-tracking). Additional intervention exclusions: < 8 sessions = fewer than 8 hr or eight sessions; training games = outside inclusion criteria; older adults = older adult participants; small screens = training on a mobile device. In case of multiple posttests only the first posttest measure was included. Note that the discrepancy between the numbers of full texts (manuscripts) and studies is due to manuscripts that included multiple studies, and the fact that some full texts also included both cross-sectional and intervention studies. AVGP = action video game players; AVGDP = action video game players; NVGP = nonvideo game players; IQ = intelligence quotient.

the gender ratio difference exceeded 20% were excluded ($n = 34$ effect sizes, 12 of which were included in Bediou et al., 2018).

Intervention Studies. To limit possible confounds related to placebo effects or differences in expectations/participant reactivity, only active controlled studies with pre- and posttest measures of cognitive performance of interest were included. Eligible interventions had to contrast an experimental group playing a commercially available action video game with an active control group who played commercially available video games that were not of the action-like genres nor of the brain training genres. We excluded action-like games that contain some, but not all, of the mechanics of action video games, such as action role playing games, action real time strategy games, multiplayer online

battle arenas games, and action sport or driving games (Dale et al., 2020; Dale & Green, 2017). Because these games share some of the core mechanics of action games, they also were not considered suitable to be included as control group games. In addition, we excluded brain-training games from the control group because, unlike most action and nonaction video games which have been designed for leisure or entertainment, brain training games are designed to enhance cognitive skills. Therefore, studies contrasting action with brain training games do not address the issue of whether action video games impact cognition or not, but rather whether action games enhance cognition more than brain training games, an issue outside of the present scope of work. In sum, possible control games included turn-based nonaction role-playing or

fantasy games, turn-based strategy games, life simulation games, puzzle games, music games, mobile games, browser games, and fighting games. Studies where action video game training took place on mobile devices were excluded (e.g., Hutchinson et al., 2015; Oei & Patterson 2013, 2015), because the small size of screens does not allow the expected load on divided attention (Cardoso-Leite et al., 2019). In line with our focus on long-term training effects (as opposed to acute physiological arousal effects), the training had to be equivalent in both experimental and control groups in terms of duration (minimum 8 hr) and number of sessions (minimum 8 days) and posttest performance measures had to be performed at least 24 hr after training. Finally, in the case of studies that took the same measures of cognitive ability at multiple time points (i.e., after 10 hr of training and then again after 20 hr of training), only the first posttest measure was included (provided it met our minimum training criterion of 8 hr of training) as this measure is more immune to practice or test–retest effects.

Type of Outcome Measure

Our primary outcome of interest was cognitive ability as measured with objective task performance. The following nine subtypes of cognitive abilities were identified based on our previous meta-analysis:

- *Perception*. Tasks measuring the precision or speed of perceptual information processing (e.g., perceptual discrimination, contrast sensitivity, visual acuity).
- *Bottom-up attention*. Tasks measuring the sensitivity to exogenous sources of attention (e.g., attentional capture, Posner cueing with exogenous cues, oculomotor capture by distractors).
- *Top-down attention*. Tasks requiring an endogenous or goal-driven control of attention (e.g., serial visual search tasks, multiple object tracking).
- *Inhibition*. Paradigms requiring the suppression of prepotent (frequent or automatic) responses (e.g., Stroop, Go–Nogo).
- *Task-switching/multitasking*. Tasks requiring maintaining several goals and switching between task sets or rules (e.g., dual task and task switching paradigms).
- *Verbal cognition*.² Tasks that do not fall in the five first domains and load primarily on a verbal encoding of the presented and manipulated information (e.g., verbal n-back, digit span, o-span).
- *Visual spatial cognition* (see footnote 2). Tasks that do not fall in the five first domains and load primarily on a spatial encoding of the presented and manipulated information (e.g., Shepard mental rotation, spatial n-back).
- *Problem-solving*. Tasks requiring complex thinking and planning to achieve a goal (e.g., Tower of London, Raven's progressive matrices).
- *Motor control*. Tasks requiring sensorimotor coordination (e.g., visuo-motor coordination tasks, sensorimotor learning).

Tasks that did not fall into any of these cognitive domains, such as measures of crystallized intelligence (e.g., Wechsler Adult Intelligence

Scale, intelligence quotient, e.g., Latham et al., 2013; Schubert et al., 2015; Strobach et al., 2012), semantic knowledge (e.g., Trivia questions in Donohue et al., 2012), or educational outcomes (e.g., mathematics performance, Libertus et al., 2017; Novak & Tassell, 2015) were excluded. We note here that some tasks that were previously categorized as, for instance, “verbal cognition” in Bediou et al. (2018) were now excluded from this meta-analysis based on this criterion (e.g., Trivia questions which measure general knowledge). Similar reasoning led us to also exclude tasks designed to elicit eye movements, physiological responses, or brain responses (e.g., eye-tracking, electroencephalography, functional Magnetic Resonance Imaging or electrodermal activity, heart rate, etc.) because the primary goal of these tasks was not to address differences in cognitive skills per se, but instead to examine various behavioral markers. These markers also frequently had the additional problem that it was not always clear which direction was “better.”

Selection of Studies

After removal of duplicates, all titles and abstracts retrieved were processed by two independent reviewers who were trained to exclude studies that did not fit inclusion criteria and to identify potential studies that should be evaluated further as a full text. Any disagreement regarding the inclusion or exclusion of a particular study was resolved through meetings between the two independent reviewers and the three main authors of the meta-analysis. Figure 1 summarizes the study selection process.

Although the inclusion and exclusion criteria differed slightly from Bediou et al. (2018), our search strategy (keywords, databases, etc.) remained unchanged. For Step 1, *Identification*, we started from the 5,770 abstracts that had passed initial screening in Bediou et al. (2018) and which covered the period 2000–2015. An additional search covering the period 2015–2020 identified 2,193 articles, as well as 19 new conference abstracts, and 43 articles that were found through other sources, including reference lists from systematic reviews and alerts on ResearchGate. In total, 2,255 records were identified for the period 2015–2020 and added to the 5,770 records for the period 2000–2015, resulted in a total of 8,025 records.

In the *Screening* step, the references were first screened by two independent reviewers who read the titles of the articles and excluded studies that fell outside of the scope of the meta-analysis. This screening excluded 4,528 records, which left a total of 3,497 titles. The two independent reviewers then read the abstracts and excluded 2,748 articles, with a total of 749 abstracts passed our eligibility criteria and entered the full text screening stage.

For the third *Eligibility* step, the reviewers read the full documents and excluded 506 records that did not meet the selection criteria. A total of 255 articles passed our inclusion criteria and were further thoroughly processed for inclusion in the cross-sectional or intervention meta-analysis. The final data set comprised 74 manuscripts for the cross-sectional meta-analysis and 22 for the intervention meta-analysis. Fourteen manuscripts comprised several experiments which were included both in the cross-sectional meta-analysis and the intervention meta-analysis. The files used during the literature search and screening process are available on Open Science Framework (OSF).

² For verbal cognition and spatial cognition, we also coded whether the task primarily loads on working memory and use this new coding in exploratory analyses (see Appendix and Supplemental Results).

Data Extraction and Management

Coding

Two independent reviewers extracted key characteristics of the population(s) (participants) and intervention(s) using standardized coding sheets. Different sheets were used for cross-sectional studies and intervention studies. Both contained similar information regarding (a) study description (e.g., title, author(s), year of publication, publication status, experiment within article); (b) sample characteristics (e.g., the number of the participants, mean age, and percentage of males in the experimental and control groups); and (c) outcome characteristics (e.g., task and condition, dependent measure, type of effect, category of cognitive skill). Additional information was extracted specifically for cross-sectional studies (e.g., whether overt or covert recruitment was used) or interventions (e.g., total duration of the training in hours), which also required supplementary fields related to the pretest and posttest data. The fully coded cross-sectional and intervention data sets, including the statistical information to calculate the effect size for each comparison, are available on OSF. The coding was thoroughly reviewed by three authors, such that all ambiguities were discussed until unanimous agreement was achieved. All notes from these discussions are available on our OSF repository for this project.

Dealing With Duplicates

When multiple sources of the same data were retrieved (e.g., published articles and PhD dissertation), we reviewed both studies to determine any differences or if additional information was provided. A number of cross-sectional studies involved groups with unmatched gender ratios (i.e., difference greater than 20%). In these cases, we restricted the analysis to only male participants whenever possible (e.g., Kowal et al., 2018; Schmidt et al., 2018, 2019; Unsworth et al., 2015; Wong & Chang, 2018). If the information reported in the article was either missing, incomplete, or ambiguous, authors were contacted to obtain additional information necessary for effect size computation or moderator coding. For data that were reported in different manuscripts, we reviewed and extracted all study information from both sources. However, if the exact same data (participants and task) were reported in distinct manuscripts, then only the effect size from the most recent source was used. Effect sizes from studies existing as both published and unpublished reports were coded as “published” as long as any of the data included in the meta-analysis was part of a peer-reviewed publication.

Unit of Analysis Issues

Multiple measures were often extracted from the same participants (e.g., different tasks, measures or experimental conditions). All effect sizes involving the same subjects, including cases of partial overlap reported in different articles, were assigned the same study identification number. To reduce possible biases due to the inclusion of multiple effect sizes from nonindependent samples, we used multilevel meta-analytic models to estimate the average effect size and explore heterogeneity, then estimated standard errors and hypothesis tests using robust variance estimation (RVE). This combination of a working model for the dependence structure

(the multilevel model) with RVE allows for greater precision in the estimation while also guarding against misspecification.

Although our preregistered model assumed dependent effect sizes to be either correlated or hierarchically related, a recent extension of this model offered a better match to our data structure, allowing for both types of dependencies. Specifically, we used the correlated and hierarchical model (CHE; Pustejovsky & Tipton, 2021), a more recently developed method than the preregistered correlated effects model. This deviation thus represents an improvement over the preregistered method. We note though (a) that none of the main inferences differed as a function of the use of this improved analysis technique and (b) although the preregistered analysis is less appropriate for the data structure than the newer technique we utilized, to maximize transparency, the results of the preregistered analytic model are available in the Supplemental Materials.

Missing Data

Forty-eight authors were contacted in order to reduce the amount of missing data, and while only four did not respond, authors who replied were not always able to answer our questions or provide the requested data. Effect sizes that could not be computed despite our attempts to obtain missing data from the authors were excluded from all analyses. The numbers of missing effects are reported in the results sections of the cross-sectional and intervention meta-analyses.

Outliers

To limit the influence of extreme effect sizes, outliers were identified and replaced with their winzorized values. This allowed us to run the analyses without excluding any effect size data, while reducing their influence on the overall results. An analysis including nonwinzorized effects is presented as a sensitivity analysis in the Supplemental Material (Appendix A, Table S2 and Appendix B, Table S7).

Meta-Analytic Procedure

Effect Sizes

We used Hedges' g as our effect size of the standardized mean difference between two groups. Cohen's d was initially computed and converted into Hedges's g , which applies an additional correction for small samples. Whenever possible, effect sizes were computed from the means and standard deviations. When group means and standard deviations (SD s) were unavailable, we relied on other statistics (e.g., T tests of group differences, F tests of group differences only if they had 1 degree of freedom or chi-square test). All equations used are available on OSF.

For cross-sectional studies, effect sizes quantify the performance difference between AVGPs and NVGPs. We used the formulas for independent groups from Borenstein (2009; equations 12.11–12.18, p. 226) corresponding to g in Lakens (2013, equation 4, p. 3).

For intervention studies, effect sizes quantify the between-group differences in performance changes from pretraining to posttraining, and thus reflect the causal effect of action video game training in comparison to an active control group trained with a nonaction video game. Although pretest differences are not expected due to random group assignment, differences present at pretest can be controlled for by using the formula for independent groups with pre-post scores,

from Borenstein (2009; equations 12.19–12.22, p. 227). Multiple computational considerations were applied and are detailed in Appendix B in terms of preference and rationale from recent methodological studies.

Statistical Analyses

All analyses were conducted in R using the *metafor* (Viechtbauer & Viechtbauer, 2015) and *clubSandwich* (Pustejovsky, 2017) packages, as well as *robumeta* (Fisher et al., 2016). Both data sets as well as analysis code are available at (<https://osf.io/3xdh8/>; Bediou et al., 2020). Analyses and hypotheses were preregistered on OSF (<https://osf.io/6qppe/>), and any deviations from the analysis plan are summarized in the deviation documentation. In particular, our primary analysis relied on the combination of a multilevel meta-analytic model with RVE (Hedges et al., 2010; Tipton, 2015). In this approach, first the dependence structure of the effect size data was approximated using a working model and this model allowed for the development of inverse-variance weights. Here we used the CHE model (Pustejovsky & Tipton, 2021). In this model, we assumed that effect sizes in the same study were correlated because they were measured on the same participants but that these true effect sizes might differ from one another (e.g., since the measures include different constructs or time points). Notably, the estimation of this model required knowledge of the correlation between effect sizes that is typically unreported; for our analyses, we assumed that effect sizes on the same participants were correlated $r = 0.80$. Second, in order to guard against misspecification—of this correlation or the working model more generally—we estimated standard errors and conducted hypothesis tests using RVE. The combination of the working model plus RVE resulted in hypothesis tests that were valid while also increasing the precision of estimates. As noted above, the use of CHE is a deviation from our preregistration, where we proposed to use a RVE model with a correlated *or* multilevel model. The combination of both into CHE was not available until 2021 (i.e., after our preregistration was accepted). This deviation was deemed critical as our data included effect sizes that were BOTH correlated AND hierarchical in nature and only the CHE model is capable of appropriately handling data with both types of structure simultaneously.

The average effect size was estimated using an intercept-only meta-analytic model. For moderator analysis, we added all moderators to the model, including cognitive domain, type of outcome measure, type of statistical effect, and recruitment method—for cross-sectional meta-analysis—or training duration—for intervention meta-analysis. Moderator effects were evaluated using the Wald *F* test, using the Approximate Hotelling's *T*-squared (AHT) approach (Tipton & Pustejovsky, 2015). Approximate Hotelling's *T*-squared test (AHT-F) tests were used to assess differences between levels of a given moderator, and *t* tests were used to test the significance of each moderator level. This was achieved by running a model in which the moderator of interest was considered (e.g., cognitive domain) while the other moderators (e.g., dependent measure type [DV], effect, and recruitment) were considered secondary or control moderators, using a loop across each moderator. In each model, only the primary moderator was assessed, and its effect was estimated by correcting for the relative frequency of the reference levels of the included control moderators in the data set (see code on OSF).

Our primary method for publication bias analysis used an alternative form of Egger's regression funnel plot asymmetry test using a

modified covariate (Pustejovsky & Rodgers, 2019) and accounting for dependent effect sizes with the CHE model (Pustejovsky & Tipton, 2021). We evaluated this alternate Egger's regression using the CHE model both with and without the moderators included in our full meta-analytic model. Our preregistered methods also included applying the three-parameter selection mode (Vevea & Hedges, 1995) to both detect and adjust for publication bias. This test is a more powerful test to detect publication bias and has great flexibility in assessing the sensitivity of estimating effect sizes adjusted for publication bias. However, the selection model methods (e.g., 3 Parameter Selection Model) is unable to handle dependent effect sizes. When applying the 3 Parameter Selection Model method, we first randomly selected one effect size per study (i.e., article) and cluster bootstrapped (1,000 repetitions), and then calculated the mean effect and mean variance across the distribution of the repetitions. Additional analyses to detect publication bias were conducted for comparability with previous meta-analyses and to continue tracking publication bias as this literature grows (e.g., Trim & Fill; Duval & Tweedie, 2000; Precision Estimate Test Precision Estimate Effect with Standard Errors; Stanley & Doucouliagos, 2014). These results are reported in the Supplemental Material.

Results

All data sets and analysis code are available at <https://osf.io/3xdh8/>.

Cross-Sectional Meta-Analysis

Data Set Description

The cross-sectional data set included 221 effect sizes (from which 170 were included in Bediou et al., 2018) extracted from 104 studies involving 91 independent samples of participants found in 73 manuscripts. Given our new inclusion criteria designed to better isolate effects specifically associated with AVGs, 30 studies that were part of Bediou et al. (2018) were not included here. This includes 11 studies (12 effect sizes) that were not included in the current meta-analysis based upon gender imbalances, nine studies (nine effect sizes) based upon criteria for AVGP, five studies (five effect sizes) involving children participants, as well as three studies (three effect sizes) focused on intelligence quotient outcome measures and one study (one effect size) in which the task used affective stimuli (Chisholm & Kingstone, 2015). Finally, one unpublished study (Föcker et al., 2014) that was included in Bediou et al. (2018) was replaced by a reanalysis included in the published version (Föcker et al., 2018).

Outcome domains with less than four effect sizes were excluded from all analyses (average effect, moderator analysis, and publication bias) because including these studies systematically affected the degrees of freedom of the other moderators and thus the ability to assess their effects (an analysis including these effects is reported in Table S2). As a result, three effect sizes from motor control, as well as six missing effects that could not be computed, were excluded from all analyses. The total number of effect sizes excluded is shown in Figure 1.

Average Effect

A total of six effect sizes that could not be computed were excluded from the analysis. The number of effects extracted from each study varied between 1 and 18 (*Mdn* = 2). The average effect of

action video games on cognition, using a multilevel meta-analytic model for dependent effect sizes with correlated and hierarchical weights and small sample correction, was estimated as $g = 0.64$, 95% CI [0.53, 0.74], $k = 212$, $m = 70$, $p < .001$, $\tau^2 = .15$, $\omega^2 = .34$. We conclude that, on average, young health adults who heavily play action video games tend to score higher on cognitive skills targeted by the games but performed outside the game context than those who do not heavily play action video games. This analysis also indicated that there was significant between study heterogeneity, $Q(df = 211) = 934.89$, $p < .0001$. In combination, this means that 95% of the effect sizes could be predicted to be between -0.73 and 2.01 (i.e., $g \pm 1.96\sqrt{\tau^2 + \omega^2}$). Thus, another goal of this meta-analysis was to try to explain these differences between and within the cross-sectional studies.

Moderator Analysis

Table 3 presents the results for the primary model including all moderators. None of the moderators showed significant moderating influence. There was not sufficient evidence to reject the null hypothesis given our preregistered α level with regard to (a) differences across cognitive domains ($p = .144$), (b) DV ($p = .389$) or effect category ($p = .884$), or (c) recruitment method ($p = .057$). Across cognitive domains, large effects were found for perception and multitasking, spatial cognition, and top-down attention followed by inhibition, verbal cognition, and problem-solving which showed moderately strong effects. Significant effects were found for both speed and accuracy and for both main effects and statistical interactions (e.g., difference scores) and for overt and covert means of recruitment. In combination, these moderators explained about 12% of the variation in effect sizes (i.e., residual $\tau^2 = .01$, $\omega^2 = .11$). However, the remaining heterogeneity was still significant, $QE(df = 201) = 844.65$, $p < .0001$, suggesting additional moderating influences are involved. We conclude that AVGP outperformed nonaction

video game players on cognitive tasks targeted by the games but performed outside the game context, regardless of the cognitive domain, DV, effect category, or recruitment method.

Publication Bias Analyses

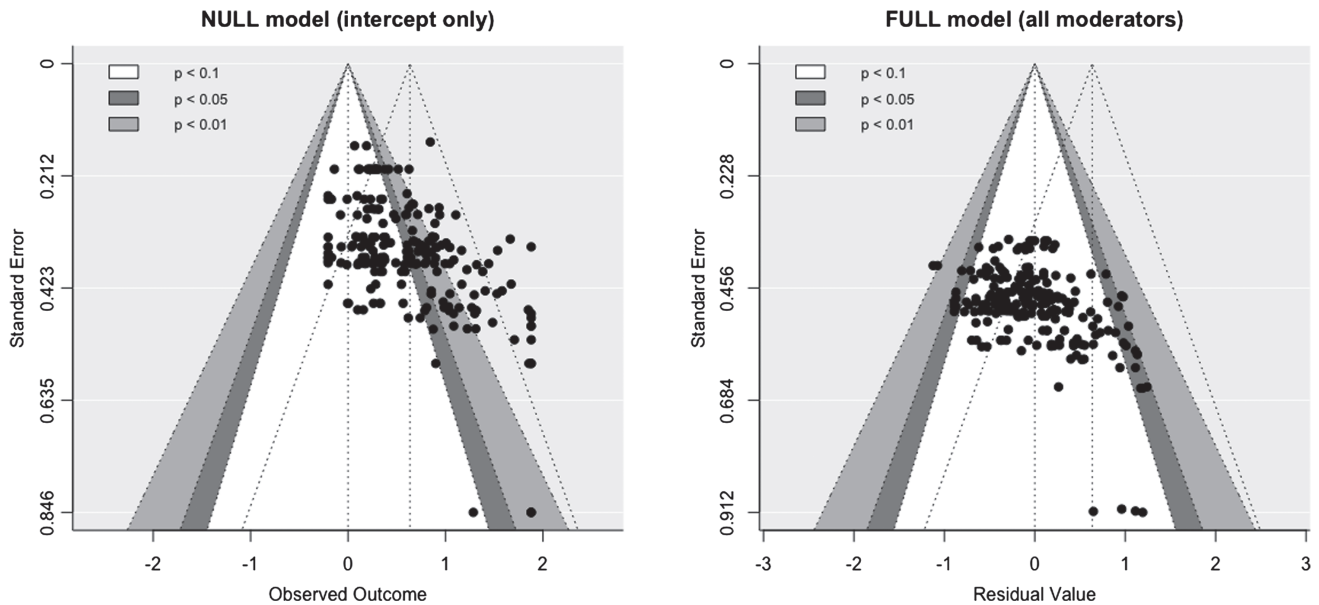
Detection of Publication Bias. Contour funnel plots (shown in Figure 2) were generated (Peters et al., 2008) to visually assess the asymmetry in the distribution of effect sizes and their standard error (Light & Pillemer, 1984; Sterne & Egger, 2001). Table 4 shows the results of our two primary methods for detecting and correcting publication bias. The modified Egger's tests showed significant small-study effects, indicating signs of publication bias, both in the model without moderators used to assess the average impact of action video game experience on cognition, NULL model, $\beta = 2.10$, $SE = 0.59$, $t(17.0) = 3.57$, $p = .002$, and also in the model that included all moderators, FULL model, $\beta = 1.99$, $SE = 0.68$, $t(21.3) = 2.91$, $p = .008$. The 3-parameter selection model (3-PSM) also detected significant publication bias both in the NULL model (mean $\chi^2 = 7.49$, $SD = 3.15$, mean p value = .021, $SD = 0.041$) and in the FULL model including all moderators (mean $\chi^2 = 7.57$, $SD = 3.16$, mean p value = .021, $SD = 0.041$). Additional preregistered analyses using more traditional as well as some advanced methods (e.g., trim and fill, Precision Estimate Test Precision Estimate Effect with Standard Errors, p-uniform) all likewise indicated the presence of small-study effects (see Appendix A and Table S1).

Correction of Publication Bias. Following convention, we also attempted to correct the average effect size for publication bias using several approaches (as summarized in Table S1 and Supplemental Figures S1–S5). In line with our preregistered method, we performed a sensitivity analysis focusing on the distribution of adjusted estimates, rather than emphasizing a single selected method. This allowed us to combine the desirable features of the various correction methods and to estimate the variability of the adjusted

Table 3
Results of Moderator Analysis—Cross-Sectional Meta-Analysis

Moderator	Level	F_{stat}	k	m	g	CI	df	p value
Cognitive domain		2.23					7.84	0.144
	Perception		38	23	0.71	[0.526, 0.893]	23.26	0.000
	Bottom-up attention		7	4	0.24	[-0.171, 0.652]	3.23	0.166
	Top-down attention		74	44	0.63	[0.489, 0.771]	39.58	0.000
	Inhibition		10	9	0.53	[0.211, 0.842]	8.04	0.005
	Spatial cognition		25	14	0.67	[0.452, 0.896]	12.90	0.000
	Multitasking		19	11	0.86	[0.33, 1.394]	9.89	0.005
	Verbal cognition		31	14	0.47	[0.279, 0.66]	12.79	0.000
Dependent measure type	Problem-solving		8	6	0.31	[0.005, 0.623]	4.46	0.048
		0.77					25.24	0.389
	Accuracy		115	49	0.65	[0.522, 0.771]	43.31	0.000
	Speed		97	44	0.58	[0.456, 0.708]	35.24	0.000
Effect		0.02					22.31	0.884
	Main		155	60	0.61	[0.506, 0.72]	46.23	0.000
Recruitment	Interaction		57	26	0.63	[0.427, 0.83]	24.56	0.000
		4.98					7.85	0.057
	Covert		32	9	0.44	[0.238, 0.643]	6.00	0.002
	Overt		180	64	0.65	[0.538, 0.759]	54.45	0.000

Note. Moderator effects were estimated using models without an intercept. df is the degrees of freedom of the denominator; the df of the numerator is equal to the number of levels. Each moderator estimate marginalizes over the other variables. Italicized rows indicate low degrees of freedom below 4. CI = confidence interval; AHT-F = Approximate Hotelling's T-squared test. Bold values correspond to moderator effect tested using AHT-F test comparing the different levels.

Figure 2*Funnel Plots for the Cross-Sectional Meta-Analysis*

Note. Left: NULL model with intercept only showing strong asymmetry; Right: FULL model with all moderators showing less asymmetry. Note the different x axis.

average estimates, considering the 3-PSM as currently being the best possible estimate of the unbiased average effect.

Finally, a significance funnel plot (shown in Figure S1) was also generated in order to better visualize the effect sizes separately for affirmative (significant) and nonaffirmative (non-significant) studies (Mathur & VanderWeele, 2020). Additional results of our publication bias sensitivity analysis are presented in Appendix A. Overall, these analyses indicated detection of publication bias, yet correction and estimation of average adjusted effects remains problematic.

Exploratory Subgroup Analyses and Sensitivity Analyses

In line with our preregistered method, additional exploratory analyses assessed the sensitivity of our results to (a) the choice of meta-analytic model (comparing with results from correlated or

hierarchical models), (b) the presence of publication bias (and the choice of method to assess it), and (c) the presence of outliers. In addition, we also examined the impact on the cross-sectional results of (d) recoding outcomes from spatial or verbal cognition into a working memory category, and of (e) controlling for joint publication group (JPG; clustering based on coauthorship frequency) by adding an additional random factor account for this. These additional analyses are available in Appendix A and produced identical inferential results as those presented above.

Intervention Meta-Analysis

Data Set Description

The intervention data set included 91 effect sizes (from which 78 were included in Bediou et al., 2018) extracted from 23 manuscripts

Table 4
Publication Bias Detection and Correction—Cross-Sectional Meta-Analysis

Method	Model	<i>b</i>	<i>SE</i>	95% CI	Statistics	<i>dfs</i>	<i>p</i>
Detection							
Egger CHE (primary model)	NULL	2.10	0.56	[0.92, 3.28]	3.76	16.99	0.00
	FULL	1.99	0.68	[0.57, 3.4]	2.91	21.32	0.008
3-PSM	NULL				7.49		0.021
	FULL				7.57		0.021
Correction							
Egger CHE (primary model)	NULL	−0.07	0.19	[−0.48, 0.33]	−0.39	12.34	0.700
	FULL	0.00	0.22	[−0.45, 0.45]	0.00	18.82	0.997
3-PSM	NULL	0.48	0.05	[0.39, 0.57]			0.000
	FULL	0.53	0.16	[0.23, 0.86]			0.049

Note. *SE* = standard error; *CI* = confidence interval; CHE = correlated and hierarchical effects; 3-PSM = 3-parameter selection model.

reporting data from 29 studies involving 18 independent samples of participants (total $N = 2,739$). Differences in inclusion criteria resulted in nine studies (73 effect sizes) from Bediou et al. (2018) being excluded: one study (14 effects) because of the control game (e.g., the control group trained with the action-like video game *Rise of Nations* in Boot et al., 2008, which included 12 effects), two studies (11 effects) because participants were older adults, three studies (43 effects) that involved video game interventions on mobile devices with small screens (Oei & Patterson, 2013, included four control groups and seven tasks resulting in 28 effect sizes, and Hutchinson et al., 2015 had two control groups trained on nintendo-DS using either a brain training game or a sight-training program which are designed to train cognitive or perceptual skills) and two studies (four effect sizes) that did not assess cognitive skills (affective task in Oei & Patterson, 2013, math performance in Novak & Tassell, 2015). Moreover, we note that for Boot et al. (2008), we included the comparison between the action game *Medal of Honor* and the control game *Tetris* at the first posttest, whereas Bediou et al. (2018), included the second posttest, which was thus confounded with test–retest effects. This data set includes 24 new effect sizes that were not included in Bediou et al. (2018). Three effect sizes from problem-solving and one from inhibition, as well as four missing effect sizes that could not be computed were also excluded. The analyzed data set thus contained a total of 83 effect sizes. In order to control for the overlap of participants across studies (i.e., the same participants reported in distinct studies or manuscripts; Bediou et al., 2018; Hilgard et al., 2019), the meta-analysis of intervention studies was conducted using participant sample as the clustering variable, instead of the article. The total numbers of excluded effects is included in Figure 1.

Average Effect

The average effect of action video games on cognition, using a multilevel meta-analytic model for dependent effect sizes with correlated and hierarchical weights and small sample correction, was $g = 0.30$, 95% CI [0.11, 0.50], $k = 83$, $m = 18$, $p = .004$, $\tau^2 = .00$, $\omega^2 = .41$, corresponding to a small effect size. We conclude that being assigned to play an action video game for 8–50 hr had, on average, a positive effect on cognitive skills target by the game but performed outside the game context as compared to being assigned to play a nonaction video game. As in the previous meta-analysis, this residual heterogeneity was significant, $QE(82) = 458.23$, $p \leq .001$. However, this suggests a wide range of possible effect sizes across samples, with 95% of effect sizes between -0.97 and 1.57 . Thus, we sought to explore possible moderating influences.

Moderator Analysis

We first ran a multilevel model with all moderators. As seen in Table 5, none of the moderators showed significant moderating influence according to the AHT-F test on the multilevel model. Thus, there was not sufficient evidence to reject the given null hypotheses (i.e., that the effect of action video games training did not differ as a function of cognitive domain, DV or type of effect). We conclude that the positive effects of being assigned to play action video games for 8–50 hr did not depend on the cognitive domain, DV, or effect category. However, the low degrees of freedom for cognitive domain suggests that this test has low power. Looking across cognitive

domains, moderate to large effects were found for top-down attention. The estimated effect of action video games play on bottom-up attention and perception were unreliable as indicated by the low degrees of freedom. Significant effects were found for accuracy only and for both main effects and statistical interactions (e.g., group difference found in a specific experimental condition or when subtracting one *baseline* or *control* condition from a condition of interest). In combination, these moderators explained about 8.5% of the variation in effect sizes (i.e., residual $\tau^2 = .00$, $\omega^2 = .16$). However, the remaining heterogeneity was still significant, $QE(df = 74) = 374.58$, $p < .001$, suggesting additional, unobserved moderating influences may be involved.

When included in the full model with all moderators, the effect of training duration was not significant ($\beta = 0.01$, 95% CI $[-0.26, 0.51]$, $t = 2.2$, $df = 5.62$, $p = .192$, $\tau^2 = .00$, $\omega^2 = .16$). Given the theoretical and practical importance of this moderator, additional analyses were conducted to better understand whether the lack of effect was due to the inclusion of other moderators. As in Bediou et al. (2018), we thus ran a model including only training duration as the single moderator. When considered in isolation, the effect of training duration was highly significant ($\beta = 0.02$, 95% CI $[0.01, 0.02]$, $t = 20.2$, $df = 5.28$, $p < .006$, $\tau^2 = .00$, $\omega^2 = .17$). This suggested that differences in training duration can be explained by other features of the studies. The effect of training duration is also illustrated in Figure 3 using the moving constant technique (Johnson & Huedo-Medina, 2011). This technique aims to estimate the effect size with a variable intercept to provide better estimates of the confidence interval across training durations. As can be seen in Figure 3, the large confidence intervals at high training duration suggest that training duration could not be properly isolated because it may be confounded with other moderators. Indeed, studies with the longest training durations almost exclusively measured the impact of AVG training on perception ($k = 9$, $m = 5$), except for one study focusing on multitasking abilities. The meta-regression analysis comparing the single-moderator model with the full model demonstrated that inclusion of moderators explained some of the heterogeneity, as the effect size stabilized across training duration for the full model including all moderators.

Publication Bias

Funnel plots were generated to visually assess the asymmetry in the distribution of effect sizes and their standard errors (Figure 4). As seen in Table 6, the robust Egger's regression test with the CHE model did not show statistically significant evidence of small-study effects/publication bias, either in the null (i.e., no moderators) model, $\beta = 2.12$, $SE = 1.11$, $t(2.58) = 1.09$, $p = .17$, or in the full model that included all moderators, $\beta = 2.19$, $SE = 1.39$, $t(2.59) = 1.57$, $p = .23$. The 3-PSM also did not detect significant publication bias both in the NULL model (mean $\chi^2 = 1.10$, $SD = 1.29$, mean p value = .46, $SD = .29$) and in the FULL model (mean $\chi^2 = 1.15$, $SD = 1.31$, mean p value = .44, $SD = .29$). Additional analyses using several other methods likewise did not indicate a significant degree of publication bias (see Table S6 and Figures S2–S5).

Exploratory Subgroup Analyses and Sensitivity Analyses

Similar to the cross-sectional meta-analysis, we performed a number of additional analyses (all preregistered), which are presented in Appendix B, including (a) an analysis using the preregistered RVE models with a correlated effects working model, (b) sensitivity to

Table 5
Results of Moderator Analysis—Intervention Meta-Analysis

Moderator	Level	F_{stat}	k	m	g	CI	df	p value
Cognitive domain		0.97					3.51	0.534
	Perception		19	6	0.23	[−0.06, 0.525]	4.71	0.095
	Bottom-up attention		7	3	0.19	[−0.957, 1.33]	3.01	0.640
	Top-down attention		24	12	0.52	[0.248, 0.793]	11.04	0.001
	Spatial cognition		15	6	0.26	[−0.269, 0.791]	5.22	0.265
	Multitasking		10	5	0.41	[−0.289, 1.107]	4.70	0.189
Dependent measure type	Verbal cognition		8	7	0.01	[−0.452, 0.469]	6.30	0.965
		0.59					6.54	0.471
	Accuracy		59	16	0.34	[0.154, 0.533]	12.03	0.002
Effect	Speed		24	8	0.25	[−0.036, 0.538]	7.38	0.078
		1.14					6.95	0.321
Interaction			61	16	0.25	[0.023, 0.475]	11.99	0.033
	Main		22	8	0.50	[0.042, 0.966]	6.63	0.037
Training duration		2.2					5.62	0.192
	Continuous		83	18	0.01	[−0.26, 0.51]	10.1	0.491

Note. All models included training duration as a moderator (see text and Figure 3 below, for the effect of training duration). Effects were estimated using a model without an intercept. Each moderator estimate marginalizes over the other variables. CI = confidence interval; AHT-F = Approximate Hotelling's T-squared test. Bold corresponds to AHT-F test comparing moderator levels.

outliers, (c) publication bias sensitivity analyses, (d) recoding spatial and verbal outcomes into working memory component, and (e) adding a random factor to account for differences between JPGs. Finally, we also tested (f) the sensitivity of our results to the impact of effect sizes standardization method by repeating the analysis with effect sizes standardized based on change scores (as in Bediou et al., 2018) or using the variance of the pretest rather than the variance of the posttest. Overall, these analyses confirmed the pattern of same

results both in terms of average effect (Table S7) and the lack of significant moderating influences (Tables S8 and S22).

Discussion

Consistent with our preregistered hypotheses, AVGP were observed to have superior cognitive skills, on average, compared with individuals who engage less in video game play as per the

Figure 3
Effect of Training Duration in Intervention Studies Using the Moving Constant Technique

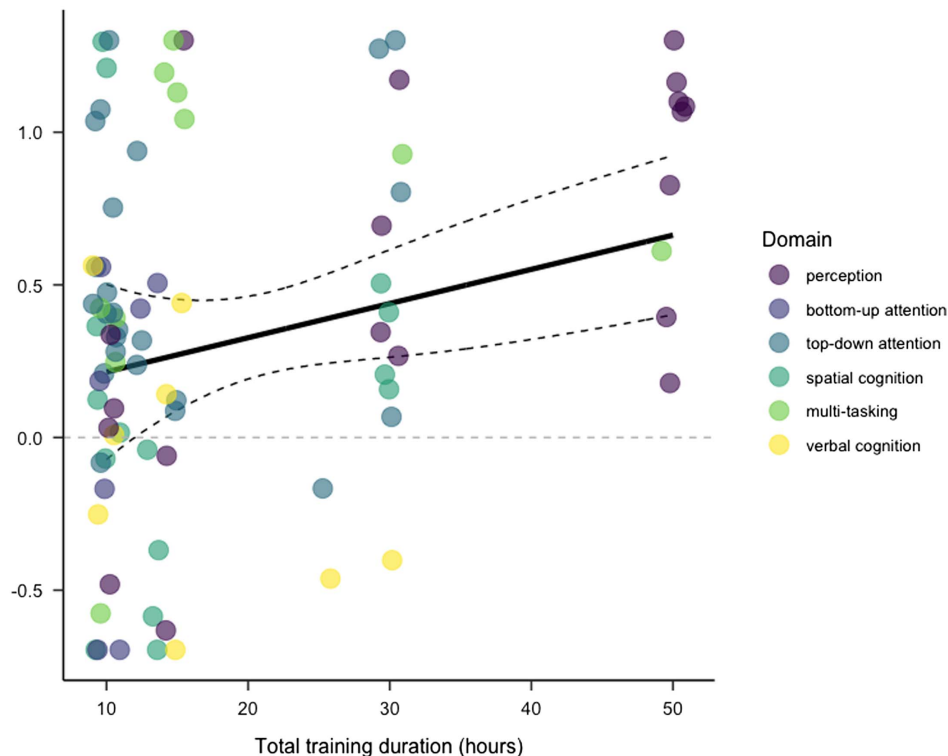
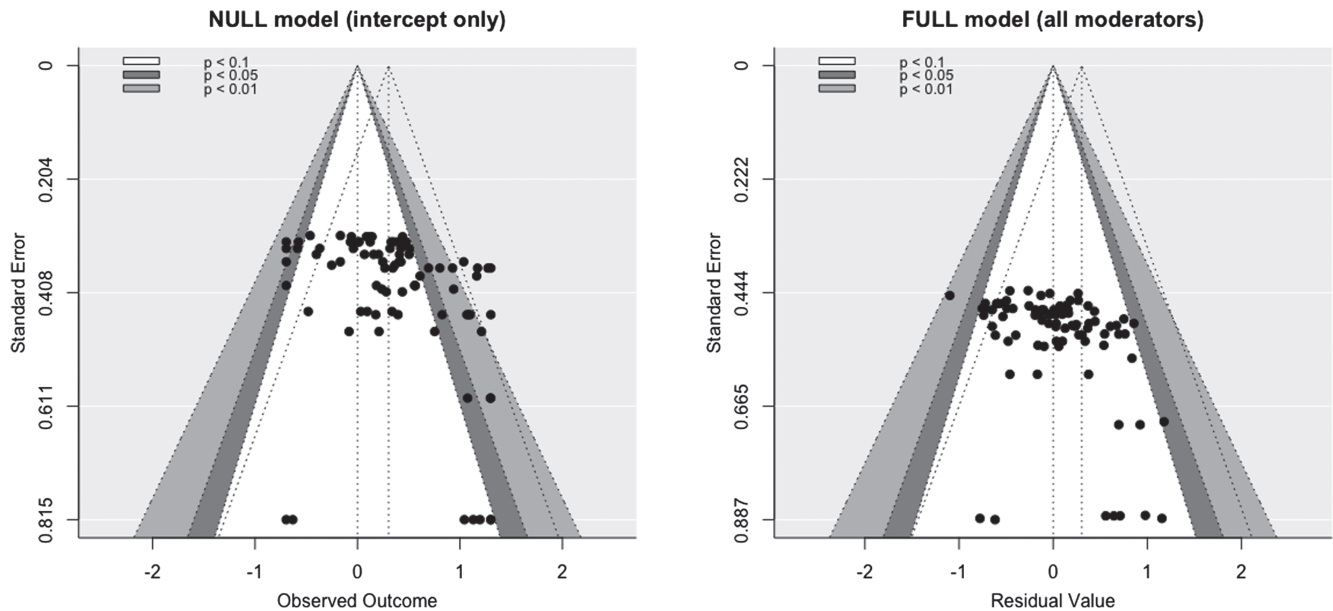


Figure 4*Funnel Plots for the Intervention Metanalysis*

Note. Left: NULL model with intercept only showing strong asymmetry; Right: FULL model with all moderators showing less asymmetry. Note the different x axis.

cross-sectional meta-analysis ($g = 0.64$, 95% CI [0.53, 0.74], $k = 212$). Results from the intervention meta-analysis also align with the preregistered hypotheses. On average, action video game play caused improvements in cognitive skills in the intervention meta-analysis ($g = 0.30$, 95% CI [0.11, 0.50], $k = 83$). In short, this meta-analysis found a positive relation, on average, between playing action video games and performance on cognitive skills that are targeted by the games but performed outside the game context. The intervention analysis supported the idea that cognitive skills trained by playing an action video game can, on average, transfer to improvements in performing the skills outside the game context. Both analyses also revealed substantial heterogeneity among the effect sizes. We discuss below the impact of the moderators we considered on the average effect of action video game play on cognition before turning to how the heterogeneity highlighted in this work may serve to advance work in the field.

Impact of Action Video Games Play Across Cognitive Domains

For cross-sectional studies, the moderator analyses indicated medium to large effect for each of the eight cognitive domains considered. The numerically strongest differences were seen in the perceptual ($g = 0.71$), top-down attentional ($g = 0.63$), and spatial ($g = 0.67$) domains. These results align well with those reported in Bediou et al. (2018). In addition, a large effect was seen for multitasking ($g = 0.81$). Medium effects were found for inhibition, verbal cognition, and problem-solving. Although a cognitive analysis of action video game play remains to be fully executed, these results most likely reflect the fact that action video games, in addition to putting a high load on top-down attention also require perceptual, spatial, and multitasking skills; in contrast, this game genre may load less on verbal cognition, inhibition, and problem-solving than many

Table 6
Publication Bias Detection and Correction—Intervention Meta-Analysis

Method	Model	b	SE	95% CI	stat	dfs	p
Detection							
Egger CHE (primary model)	NULL	2.12	1.11	[−1.77, 6.01]	1.91	2.58	0.17
	FULL	2.19	1.39	[−2.66, 7.04]	1.57	2.59	0.228
3-PSM	NULL				1.10		0.455
	FULL				1.15		0.444
Correction							
Egger CHE (primary)	NULL	−0.48	0.43	[−1.81, 0.85]	−1.12	3.19	0.341
	FULL	−0.61	0.63	[−2.27, 1.05]	−0.97	4.62	0.379
3-PSM	NULL	0.33	0.12	[0.1, 0.58]			
	FULL	0.56	0.52	[−0.33, 1.52]			

Note. SE = standard error; CI = confidence interval; CHE = correlated and hierarchical effects; 3-PSM = 3-parameter selection model.

other daily activities of a young adult (which may very well include playing nonaction video games that involve, for instance, puzzles). Only bottom-up attention was associated with a nonsignificant effect. While this may be surprising at first sight as action video game play encompasses many abrupt onset/offsets at the core of the bottom-up attention experimental designs used to assess this domain, it is consistent with the hypothesis that bottom-attention processes are phylogenetically quite ancient, early to develop and in great part mediated by subcortical structures and thus likely to be less plastic throughout life (Petersen & Posner, 2012).

For intervention studies, among the six cognitive domains considered, top-down attention ($g = 0.52$) and multitasking ($g = 0.41$) showed medium effect sizes, while perception, bottom-up attention, spatial cognition showed a small effect and verbal cognition a null effect. Confidence intervals indicated both negative and positive bounds for all domains but top-down attention, which is the only domain where a significant effect of action game play was noted. The low degrees of freedom for the effect of cognitive domain ($df = 3.51$, $p = .53$) suggested top-down attention should not be seen as differing from other domains.

Of note, all intervention studies contrasted action video games against another commercially available game from a different genre. From this point of view, the effect on top-down attention is in line with our hypothesis that action video games play enhances attentional control more than other nonaction like game genres such as puzzle games or social simulation games (Choi et al., 2020; Dale et al., 2020; Dale & Green, 2017). Training duration was a significant moderator of effect size magnitude, with longer interventions leading to stronger effects. Adding other moderators in the model reduced this effect suggesting that the effect of training duration may vary across domains which is consistent with the idea that some brain functions are more plastic than others (Bavelier & Neville, 2002; Böckler & Singer, 2022). Yet, the finding that training duration also differs across moderators (e.g., see Figure 3B showing that most 50 hr training studies focused on perception) makes it difficult to disentangle their respective effects.

In both analyses, the cognitive domain moderator failed to highlight differences across cognitive domains. Yet, greater numerical effects were observed for domains that action video games are likely to challenge to a greater extent than other everyday activities in the cross-sectional data set. Similarly, greater numerical effects were observed for domains that action video games are likely to challenge to a greater extent than other control video games in the intervention data set. As video games tend to be more similar to each other than other daily activities, it is no surprise to see intervention studies showing numerically smaller effect sizes and less broad cognitive improvement across domains than cross-sectional studies.

Speed Versus Accuracy Measures

Effect sizes were comparable for accuracy and speed measures in both cross-sectional and intervention studies with numerically stronger effect sizes for accuracy than speed. This speaks against the oft-raised criticism that action video game play merely facilitates motor execution, rather than providing cognitive benefits.

The finding that the impact of action video game play on cognition is equally detectable through both accuracy and speed measures aligns well with several studies showing enhanced sensitivity or

information accumulation in the service of decision-making after action video game play (Bejjanki et al., 2014; Green et al., 2010). Although the finding of enhanced sensitivity was challenged by van Ravenzwaaij et al. (2014), this work had to be excluded due to massed practice (i.e., either 10 hr training in 5 days or 20 hr in 5 days) and repeated testing (cognitive performance was repeatedly measured after each video game play session).

As the field advances this is an important factor to keep track of, as reaction times alone, the preferred measure of speed in Psychology, does not allow these different mechanisms of improvement to be distinguished (i.e., from reaction time alone it is difficult to disentangle increases in sensory integration rate from speed–accuracy trade-offs).

Expectation Effects

For the cross-sectional meta-analysis, group differences were observed not only when participants were recruited overtly but also when recruited covertly, with medium-to-large effects for both covert and overt recruitment. Similar group differences were observed for main effects measured via overall performance (e.g., AVGP should respond faster), compared to interaction effects measured via difference scores (e.g., AVGP should respond disproportionately faster on switch trials than on nonswitch trials, resulting in smaller switch costs) in both the cross-sectional and intervention analyses. These latter results speak against an effect of expectations; indeed, while it may be possible for participants to intuit what is expected of them in terms of main effects such as faster responding, doing so for interaction effects is nearly impossible (i.e., when the researcher's expectation is that action game players will be disproportionately fast to respond on only some subset of trials in a full experiment). The lack of difference between main and interaction is in line with our previous meta-analysis (Bediou et al., 2018).

Numerically stronger effects were found when overt recruitment was used compared to covert recruitment in cross-sectional studies. Although these effects did not reach the threshold for statistical significance and thus expectation effects due to recruitment could not be detected, if they were to be detected, this would be a result of high practical significance. Indeed, as demonstrated by the field of cognitive training, inducing durable cognitive enhancement, especially in young adults, is a tall order. If the cognitive benefits of AVGP could be augmented through the manipulation of expectations, it would add a complementary pathway to enhance cognition (Denkinger et al., 2021). Unfortunately, the few studies that have tried to manipulate expectations in this way, whether cross-sectional or interventions, suggest such effects require significant and purposeful effort to induce (Parong et al., 2022; Vodyanyk et al., 2021).

Publication Bias

For cross-sectional studies, the Egger's CHE test indicated significant funnel plot asymmetry and the 3-PSM indicated an average positive effect after correction ($g = 0.48$ for the null model or $g = 0.53$ for the full model). In this context, the 3-PSM technique, which is less sensitive to heterogeneity may thus provide a more reliable estimate of the average bias-corrected association between action video game experience and cognitive performance in cross-sectional studies and indicated a positive relationship after adjustment for publication bias.

The intervention meta-analysis revealed a different picture. Our primary methods (Egger's test and 3-PSM) failed to detect significant publication bias. This lack of significant publication bias may be less surprising considering our focus on randomized controlled designs as well as the smaller number of studies included. Indeed, randomized controlled designs with minimal attrition are less susceptible to introducing bias in causal estimates due to minimal baseline differences between control and treatment groups. While this does not directly reduce the potential of publication bias, the rigor of these designs improves the precision of causal effect size estimates. Nonrandomized studies in the cross-sectional design synthesis can be more susceptible to variable, imprecise estimates due to smaller sample sizes or variation in groups at baseline. Again, high heterogeneity may limit the reliability of regression techniques, especially when it comes to estimating the bias-corrected average adjusted effect.

While publication bias analyses are critical to ensure the validity of meta-analytic results, there are still both conceptual and methodological issues associated with its detection, and even more so when it comes to estimating a bias-corrected average effect size. Currently available methods to detect and especially adjust for publication bias lack consistent power and are known to perform poorly in the presence of heterogeneity (Fernández-Castilla et al., 2019; McShane et al., 2016; Rodgers & Pustejovsky, 2019; Stanley, 2017; Terrin et al., 2003; van Aert et al., 2016; van Assen et al., 2015). Finally, the fact that small-study effects can arise from multiple causes including genuine methodological differences (e.g., true heterogeneity) warrants further caution when interpreting the results of publication bias analyses (Rothstein et al., 2005; Sterne & Egger, 2005). Therefore, we followed the increasingly common approach of using sensitivity analyses and relying on the range of estimates and their confidence intervals, rather than a single adjusted pooled effect size estimate.

Exploratory Results

Preregistered exploratory analyses produced identical inferential results as those just discussed. Controlling for JPG resulted in numerically smaller effects along with an increase in between-study heterogeneity; as discussed in Bediou et al. (2018), some research groups have applied stricter criteria, whether in cross-sectional (requiring 5+ hr per week and not three as selected here) or in intervention studies (running interventions for 20+ hr), which may be at the source of these effects. Combining spatial and verbal domains under a working memory domain indicated an effect of action video game play on working memory, in both the cross-sectional and intervention meta-analyses. In addition, including outliers resulted in smaller average effects in the cross-sectional meta-analysis, whereas the inclusion of outliers increased the average estimate of action video games impact in intervention studies, although the CI was also larger (see results of analyses comparing winsorized and nonwinsorized data sets in the Appendix A and B). Finally, the standardization method to calculate Hedge's g based on posttest SD s gave numerically smaller effects compared to the SD of pre-post differences (change scores) as was used in Bediou et al. (2018), together with smaller between-study heterogeneity but larger within-study heterogeneity. Additional details about these exploratory findings can be found in the Supplemental Material.

Toward an Understanding of the Sources of Heterogeneity

Beyond estimation of average effect, a distinctive strength of meta-analytic work is to evaluate heterogeneity, both in terms of between-studies and within-study heterogeneity. The present work indicates high heterogeneity especially for the cross-sectional meta-analysis. We discuss below important challenges to address in future research that could significantly contribute in our understanding of the heterogeneity in results. Progress in these dimensions would help refine both cross-sectional and intervention work.

Participants Inclusion Criteria in an Ever-Changing Video Game Landscape

The high heterogeneity of cross-sectional studies may be, at least in part, attributable to the fast-changing ecosystems around video games. Over the past 40 years, video game play has evolved from a relatively niche activity, mostly limited to young males, to a form of entertainment enjoyed by 90%+ of young adults. One of the biggest challenges faced by contemporary cross-sectional studies then is the almost impossible task of finding participants with little to no video game play experience. While this was quite possible in the early 2000s (i.e., before mobile devices), in just 20 years, the population of nongamers have substantially dwindled. Thus, the nongamers in studies conducted over the past 4 years likely have had significantly more gaming experience than nongamers in studies conducted in the first 4 years covered by the meta-analysis. Interestingly, the evolving video game ecosystem has likewise made it more difficult to find true AVGP—that is, individuals who really only play action video games, rather than who play action video games as one of many different types of games that they play. This is partially due to the increasing prevalence of hybrid genres, which mix action game mechanics with characteristics of other game genres (e.g., combining action-game-based combat with role-playing characteristics, or adventure characteristics). Finally, unlike action games of the early 2000s—which were largely linear in nature (sometimes referred to as corridor shooters, in that players were shunted through a virtual hallway, always in the same order), games today are much more frequently sandbox or online in nature. This has the knock-on effect of reducing the similarity of the experience of two individuals who nominally played the same game (i.e., in an early 2000s first-person shooter game, all players would have experienced the same levels in the same order; in a 2020s first-person shooter game, all the gameplay might be online and thus no two players would have the same experience as each other). Critically, each and every one of these cohort effects would tend to have the effect of reducing the differences between AVGPs and NVGPs today as compared to the early 2000s. In future meta-analytic work, it would be interesting to consider year of publication as a moderator.

The Task Impurity Problem

A major challenge undertaken by the present meta-analyses was to carve cognition into separate subdomains. Indeed, in doing so, we had to face head-on what is known as the task impurity problem, or the fact that any task requiring an informed decision is likely to load on a combination of cognitive domains. For example, an n -back task not only requires working memory, but also inhibition of past, now

irrelevant items (and perception of a temporally evolving stream of items that could mask one another, and potentially switching between different n-levels, and so forth). While we went through extraordinary lengths in classifying each and every task reported in this meta-analysis along the nine cognitive domains discussed above, it's critical to note our preferred classifications and interpretations did not necessarily always align with that of the authors. As preregistered, in such cases, the classification/interpretation the authors originally used in their work was kept. However, as we made clear in our 2018 meta-analytic work, this state of affairs can be especially thorny. First, the very same name—n-back task or visual short-term memory task—may refer to quite different implementations and thus load on rather different cognitive constructs. Second, and even more problematically, for some tasks such as the Eriksen flanker task, a larger flanker effect can be interpreted either as a positive (more attentional resources) or a negative (less inhibitory control) behavior depending on the study context (see Bediou et al., 2018, for a full discussion).

In short, the wide variety of implementations that exists for cognitive tasks is likely to contribute to the heterogeneity highlighted in the present work. While this diversity of implementation means that no two articles tested for a given cognitive constructs using exactly the same instrument, this variety also makes the richness of the cognitive, and more generally, task-based literature. While some may argue benchmarked tasks should be used, utilizing only such tasks runs the risk of characterizing more the very task parameters chosen for that benchmark, than the cognitive constructs underlying its performance. Rather, it may be helpful in future work to think through the use of latent models to first characterize from task performance the relevant constructs, and to carry out the meta-analytic work on the latent variables rather than raw data from each and every tasks. Yet, doing so requires at least two and preferably three or more tasks per construct, a feature which is not met by any of the articles included here.

The path forward here appears therefore difficult—but it seems advisable to keep domains of cognition relatively large to avoid an overcharacterization that may not reflect the true cognitive component under study.

Understanding the Effect of Video Game Play, Game Genres, and Mechanics

Because all intervention studies used another nonaction video game as an active control, our results above provide further evidence indicating that not all games are created equal with respect to their impact on cognitive skills. What drives the positive effects of action video game play is still a topic of debate with a number of possibilities being put forward related to specific mechanics or dynamics (e.g., the need to make decisions under extreme time pressure, the need to utilize both extremely focused attentional states and extremely diffuse attentional states and to rapidly switch between them, etc. Dale et al., 2020). Yet, the nature of commercial video games makes these hypotheses difficult to test—both in intervention studies (e.g., because it would be necessary to find games that somehow allow for perfect contrasts of these mechanisms) and even more so in cross-sectional studies, where individuals are typically playing a huge mixture of games, often across genres, but also platforms (e.g., console, computer and mobile). As a result,

correlational and cross-sectional studies do not report the hardware players use, whereas only few studies have exclusively used mobile devices (e.g., Oei & Paterson 2013, 2015) or a mixture of devices (e.g., Hutchinson et al., 2015). To be more complete, the only article we are aware of that questioned participants about genre/platform/hardware/mode is a master thesis by Brostek (2019) from Rochester Institute of Technology (see Table 4, p. 41); however, this work does not measure cognition. Given these issues, research groups have utilized a variety of definitions for AVGPs as well as different contrast sets/games in intervention studies depending on the exact mechanisms they are interested in. This heterogeneity in the definitions of the action video game genres, and the heterogeneity of criteria for defining AVGPs and NVGPs groups, creates definite difficulties for meta-analyses. For example, here we excluded studies in which the games that authors listed as belonging to the action genre did not match our emphasis on mechanics. Similarly, we would likely consider games to be inappropriate as prospective training games if they do not allow a proper control on the type of gameplay (competitive vs. cooperative) as the difficulty level of the game, which is a key determinant of their cognitive effects. In short, if a game is online only, it may always be too difficult for nonaction gaming individuals and thus would not a priori be expected to produce any benefits. Finally, the choice of control game is also critical and generally determined by the research question. For example, studies examining the role of specific mechanics might contrast two action games that differ in one specific dimension (e.g., pace/speed) and hence did not qualify for this meta-analysis (e.g., Hoseini et al., 2022 used *Uncharted* in 2D vs. 3D).

Ongoing efforts to understand which ingredients in action games are responsible for their cognitive enhancing properties, and what are the mechanisms of these effects, should therefore be encouraged (Arnab et al., 2015; Ben-Sadoun & Alvarez, 2022; Proulx et al., 2017). Future work should focus on examining the particular features of action games that may foster learning to learn and thereby favor the transfer of cognitive benefits of video game play to real-world outcomes (Pasqualotto, Parong, et al., 2022), such as educational ones (Pasqualotto, Altarelli, et al., 2022).

Conclusion

Many strong claims are made for the power of video game playing to improve cognitive skills, but these claims are not often supported by research evidence (Mayer, 2019). In contrast, this meta-analysis adds substantial support to the proposal that there is a connection between playing action video games and human cognition. In cross-sectional studies, AVGP tended to score higher than nonplayers on tests of cognitive skill involved in the games, with an average effect size of $g = 0.64$. In intervention studies, nongame players who were assigned to play action video games for an extended time period showed greater improvements on tests of cognitive skills involved in the games than those assigned to engage in a control activity, with an average effect size of $g = 0.30$. This meta-analysis is consistent with a previous one (Bediou, 2018; Bediou et al., 2018), thereby updating evidence for the role of action video game playing in human cognition based on more recent studies. Overall, this meta-analysis encourages further work on the design of computer games for cognitive training.

References

- Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., De, S., Louchart, S., Suttie, N., Berta, R., & Gloria, A. D. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2), 391–411. <https://doi.org/10.1111/bjet.12113>
- Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, 35(1), 391–416. <https://doi.org/10.1146/annurev-neuro-060909-152832>
- Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nature Reviews Neuroscience*, 3(6), 443–452. <https://doi.org/10.1038/nrn848>
- Bediou, B. (2018). “Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills”: Correction to Bediou et al. (2018). *Psychological Bulletin*, 144(9), 978–979. <https://doi.org/10.1037/bul0000168>
- 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. *Psychological Bulletin*, 144(1), 77–110. <https://doi.org/10.1037/bul0000130>
- Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2020). Effects of action-shooter video game play on cognitive skills: A meta-analysis. <https://osf.io/3xdh8/>
- Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z. L., & Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. *Proceedings of the National Academy of Sciences of the United States of America*, 111(47), 16961–16966. <https://doi.org/10.1073/pnas.1417056111>
- Ben-Sadoun, G., & Alvarez, J. (2022). Gameplay Bricks Model, a theoretical framework to match game mechanics and cognitive functions. *Games and Culture*. Advance online publication. <https://doi.org/10.1177/15554120221080925>
- Böckler, A., & Singer, T. (2022). Longitudinal evidence for differential plasticity of cognitive functions: Mindfulness-based mental training enhances working memory, but not perceptual discrimination, response inhibition, and metacognition. *Journal of Experimental Psychology: General*, 151(7), 1573–1590. <https://doi.org/10.1037/xge0001143>
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, 129(3), 387–398. <https://doi.org/10.1016/j.actpsy.2008.09.005>
- Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). Sage Publications.
- Brostek, C. (2019). *The effect of video game experience in different genres on task-switching ability* [Unpublished doctoral dissertation]. Rochester Institute of Technology. <https://scholarworks.rit.edu/theses/10104>
- Cardoso-Leite, P., Joessel, A., & Bavelier, D. D. (2019). Games for enhancing cognitive abilities. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 437–468). MIT Press. <https://mitpress.mit.edu/books/handbook-game-based-learning>
- Chisholm, J. D., & Kingstone, A. (2015). Action video game players’ visual search advantage extends to biologically relevant stimuli. *Acta Psychologica*, 159, 93–99. <https://doi.org/10.1016/j.actpsy.2015.06.001>
- Choi, E., Shin, S.-H., Ryu, J.-K., Jung, K.-I., Kim, S.-Y., & Park, M.-H. (2020). Commercial video games and cognitive functions: Video game genres and modulating factors of cognitive enhancement. *Behavioral and Brain Functions*, 16(1), Article 2. <https://doi.org/10.1186/s12993-020-0165-z>
- Dale, G., & Green, C. S. (2017). The changing face of video games and video gamers: Future directions in the scientific study of video game play and cognitive performance. *Journal of Cognitive Enhancement: Towards the Integration of Theory and Practice*, 1(3), 280–294. <https://doi.org/10.1007/s41465-017-0015-6>
- Dale, G., Joessel, A., Bavelier, D., & Green, C. S. (2020). A new look at the cognitive neuroscience of video game play. *Annals of the New York Academy of Sciences*, 1464(1), 192–203. <https://doi.org/10.1111/nyas.14295>
- Denkinger, S., Spano, L., Bingel, U., Witt, C. M., Bavelier, D., & Green, C. S. (2021). Assessing the impact of expectations in cognitive training and beyond. *Journal of Cognitive Enhancement*, 5(4), 502–518. <https://doi.org/10.1007/s41465-021-00206-7>
- Donohue, S. E., James, B., Eslick, A. N., & Mitroff, S. R. (2012). Cognitive pitfall! Video game players are not immune to dual-task costs. *Attention, Perception, & Psychophysics*, 74(5), 803–809. <https://doi.org/10.3758/s13414-012-0323-y>
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & den Noortgate, W. V. (2019). Detecting selection bias in meta-analyses with multiple outcomes: A Simulation study. *Journal of Experimental Education*, 89(1), 125–144. <https://doi.org/10.1080/00220973.2019.1582470>
- Fisher, Z., Tipton, E., & Hou, Z. (2016). *Robumeta: Robust variance meta-regression*. R package version 1.8. <https://cran.r-project.org/web/package/s/robumeta/robumeta.pdf>
- Föcker, J., Cole, D., & Bavelier, D. (2014). *Attentional network recruitment in action video game players*. Human Brain Mapping.
- Föcker, J., Cole, D., Beer, A. L., & Bavelier, D. (2018). Neural bases of enhanced attentional control: Lessons from action video game players. *Brain and Behavior*, 8(7), e01019. <https://doi.org/10.1002/brb3.1019>
- Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. *Current Biology*, 20(17), 1573–1579. <https://doi.org/10.1016/j.cub.2010.07.040>
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, Pubmed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181–217. <https://doi.org/10.1002/jrsm.1378>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Hilgard, J., Sala, G., Boot, W. R., & Simons, D. J. (2019). Overestimation of action-game training effects: Publication bias and salami slicing. *Collabra: Psychology*, 5(1), Article 30. <https://doi.org/10.1525/collabra.231>
- Hirst, R. J., Cragg, L., & Allen, H. A. (2018). Vision dominates audition in adults but not children: A meta-analysis of the Colavita effect. *Neuroscience and Biobehavioral Reviews*, 94, 286–301. <https://doi.org/10.1016/j.neubiorev.2018.07.012>
- Hoseini, F. S., Khodadadi, M., & Khorambakht, A. (2022). The effect of 2D and 3D action video game interventions on executive functions in male students. *Simulation & Gaming*, 53(5), 405–422. <https://doi.org/10.1177/10468781221110577>
- Hutchinson, C. V., Barrett, D. J., Nitka, A., & Raynes, K. (2015). Action video game training reduces the Simon Effect. *Psychonomic Bulletin & Review*, 23, 587–592. <https://doi.org/10.3758/s13423-015-0912-6>
- Johnson, B. T., & Huedo-Medina, T. B. (2011). Depicting estimates using the intercept in meta-regression models: The moving constant technique. *Research Synthesis Methods*, 2(3), 204–220. <https://doi.org/10.1002/jrsm.49>
- Kowal, M., Toth, A. J., Exton, C., & Campbell, M. J. (2018). Different cognitive abilities displayed by action video gamers and non-gamers. *Computers in Human Behavior*, 88, 255–262. <https://doi.org/10.1016/j.chb.2018.07.010>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs.

- Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Latham, A. J., Patston, L. L., & Tippett, L. J. (2013). The virtual brain: 30 years of video-game play and cognitive abilities. *Frontiers in Psychology*, 4, 629. <https://doi.org/10.3389/fpsyg.2013.00629>
- Libertus, M. E., Liu, A., Pikul, O., Jacques, T., Cardoso-Leite, P., Halberda, J., & Bavelier, D. (2017). The impact of action video game training on mathematical abilities in adults. *AERA Open*, 3(4). <https://doi.org/10.1177/2332858417740857>
- Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Harvard University Press. https://scholars.unh.edu/psych_facpub/194
- Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Applied Statistics*, 69(5), 1091–1119. <https://doi.org/10.1111/rssc.12440>
- Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press. <https://doi.org/10.7551/mitpress/9427.001.0001>
- Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, 70(1), 531–549. <https://doi.org/10.1146/annurev-psych-010418-102744>
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, 11(5), 730–749. <https://doi.org/10.1177/1745691616662243>
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105–125. <https://doi.org/10.1037/1082-989X.7.1.105>
- Novak, E., & Tassell, J. (2015). Using video game play to improve education-majors' mathematical performance: An experimental study. *Computers in Human Behavior*, 53, 124–130. <https://doi.org/10.1016/j.chb.2015.07.001>
- Oei, A. C., & Patterson, M. D. (2013). Enhancing cognition with video games: A multiple game training study. *PLOS ONE*, 8(3), e58546. <https://doi.org/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. <https://doi.org/10.3389/fpsyg.2015.00113>
- Parong, J., Seitz, A. R., Jaeggi, S. M., & Green, C. S. (2022). Expectation effects in working memory training. *Proceedings of the National Academy of Sciences*, 119(37), e2209308119. <https://doi.org/10.1073/pnas.2209308119>
- Pasqualotto, A., Altarelli, I., De Angeli, A., Menestrina, Z., Bavelier, D., & Venuti, P. (2022). Enhancing reading skills through a video game mixing action mechanics and cognitive training. *Nature Human Behaviour*, 6, 545–554. <https://doi.org/10.1038/s41562-021-01254-x>
- Pasqualotto, A., Parong, J., Green, C. S., & Bavelier, D. (2022). Video game design for learning to learn. *International Journal of Human-Computer Interaction*. Advance online publication. <https://doi.org/10.1080/10447318.2022.2110684>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Petersen, S., & Posner, M. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, 35, 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>
- Powers, K. L., & Brooks, P. J. (2014). Evaluating the specificity of effects of video game training. In F. Blumberg (Ed.), *Learning by playing: Frontiers of video gaming in education* (pp. 302–329). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199896646.003.0021>
- 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. <https://doi.org/10.3758/s13423-013-0418-z>
- Proulx, J.-N., Romero, M., & Arnab, S. (2017). Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning. *Simulation & Gaming*, 48(1), 81–97. <https://doi.org/10.1177/1046878116674399>
- Pustejovsky, J. (2017). clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections. R package Version 0.2.3. *R Foundation for Statistical Computing*, Vienna.
- Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, 10(1), 57–71. <https://doi.org/10.1002/jrsm.1332>
- Pustejovsky, J. E., & Tipton, E. (2021). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*. Advance online publication. <https://doi.org/10.1007/s11121-021-01246-3>
- Rodgers, M., & Pustejovsky, J. (2019). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*. Advance online publication. <https://doi.org/10.31222/osf.io/vqp8u>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. <https://doi.org/10.1002/0470870168>
- Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, 144(2), 111–139. <https://doi.org/10.1037/bul0000139>
- Schmidt, A., Geringswald, F., & Pollmann, S. (2019). Spatial contextual cueing, assessed in a computerized task, is not a limiting factor for expert performance in the domain of team sports or action video game playing. *Journal of Cognitive Enhancement*, 3(3), 281–292. <https://doi.org/10.1007/s41465-018-0096-x>
- Schmidt, A., Geringswald, F., Sharifian, F., & Pollmann, S. (2018). *Not scene learning, but attentional processing is superior in team sport athletes and action video game players*. bioRxiv. <https://doi.org/10.1101/353953>
- Schubert, T., Finke, K., Redel, P., Kluckow, S., Muller, 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. <https://doi.org/10.1016/j.actpsy.2015.03.005>
- Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological & Personality Science*, 8(5), 581–591. <https://doi.org/10.1177/1948550617693062>
- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60–78. <https://doi.org/10.1002/jrsm.1095>
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046–1055. [https://doi.org/10.1016/S0895-4356\(01\)00377-8](https://doi.org/10.1016/S0895-4356(01)00377-8)
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In A. J. Sutton & M. B. H. R. Rothstein (Eds.), *Publication bias in meta-analysis—Prevention, assessment and adjustments* (pp. 99–110). Wiley. <https://doi.org/10.1002/0470870168.ch6>
- Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychologica*, 140(1), 13–24. <https://doi.org/10.1016/j.actpsy.2012.02.001>
- Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, 22(13), 2113–2126. <https://doi.org/10.1002/sim.1461>
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393. <https://doi.org/10.1037/met0000011>

- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, 40(6), 604–634. <https://doi.org/10.3102/1076998615606099>
- Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is playing video games related to cognitive abilities? *Psychological Science*, 26(6), 759–774. <https://doi.org/10.1177/0956797615570367>
- van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p -uniform and p -curve. *Perspectives on Psychological Science*, 11(5), 713–729. <https://doi.org/10.1177/1745691616650874>
- van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, 20(3), 293–309. <https://doi.org/10.1037/met0000025>
- van Ravenzwaaij, D., Boekel, W., Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2014). Action video games do not improve the speed of information processing in simple perceptual tasks. *Journal of Experimental Psychology: General*, 143(5), 1794–1805. <https://doi.org/10.1037/a0036923>
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435. <https://doi.org/10.1007/BF02294384>
- Viechtbauer, W., & Viechtbauer, M. W. (2015). *Package 'metafor'*. *The comprehensive R archive network*. <https://cran.r-project.org/web/package/metafor/metafor.pdf>
- Vodyanyk, M., Cochrane, A., Corriveau, A., Demko, Z., & Green, C. S. (2021). No evidence for expectation effects in cognitive training tasks. *Journal of Cognitive Enhancement*, 5(3), 296–310. <https://doi.org/10.1007/s41465-021-00207-6>
- Wang, P., Liu, H. H., Zhu, X. T., Meng, T., Li, H. J., & Zuo, X. N. (2016). Action video game training for healthy adults: A meta-analytic study. *Frontiers in Psychology*, 7, Article 907. <https://doi.org/10.3389/fpsyg.2016.00907>
- What Works Clearinghouse Procedures Handbook, Version 4.1. (2020). 76. <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf>
- 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), Article 5528. <https://doi.org/10.1038/s41598-018-23819-z>

Appendix A

Supplementary Analyses of Cross-Sectional Studies

Publication Bias Sensitivity Analyses

Given the importance of detecting and correcting for publication bias when interpreting meta-analytic results, we performed sensitivity analyses for both detection and correction of publication bias using the two main types of methods. First, we used regression-based techniques because they can handle dependent effect sizes. Publication bias was estimated by adding the standard error or variance to meta-analytic models or a modified covariate. This includes the Egger's tests and Precision Estimate Test Precision Estimate Effect with Standard Errors approach, to which we applied the two types of RVE approaches (correlated or hierarchical), as well as the more recent CHE (correlated and hierarchical) modeling approach. This was done on our two main meta-analytic models, the null model with intercept-only and the full model with all moderators. The second group of methods (trim and fill, p -uniform and three-parameter models) require independent effect sizes and thus necessitate that dependent effects are either aggregated into a single, average estimate or that only one effect be randomly selected prior to the analysis. Here, we use random selection with a bootstrapping procedure in order to quantify the distribution of publication bias estimates. The 3-PSM was applied to both the null model and the full model, whereas the trim and fill and p -uniform do not allow for the inclusion of any moderator and were used only with the null model. The results are summarized in Table S1 and Figures S1A–S5A. Except for the PEESE with CHE modeling approach, all analyses detected significant publication bias. As can be seen on Figure S5A, “bias-corrected” average effect size estimates varied between $g = -0.43$ (PET with CHE estimate and no moderator) and $g = 0.73$ (p -uniform). Regression-based correction methods also consistently suggested that there was no significant difference in cognition between AVGP and NVGP after adjustment, despite one exception (e.g., PET with

CHE modeling approach produced an average negative estimate when applied to the null model). In contrast, the trim and fill, p -uniform, and 3-PSM consistently found significant effects after adjustment. Finally, to understand the source of the small study effect, a cumulative forest plot was generated showing that while early studies were small and reported large effects, each additional study reduced the average effect size until a stabilization was observed around an average effect of $g = 0.6$ (Figure S6).

Choice of Meta-Analytic Models

We preregistered an analysis based on RVE models with correlated weights. However, because the CHE model was better suited to the structure of our data set, we decided to report this model as our primary analysis. The average estimates from the preregistered RVE model with correlated weights was very similar, with an average effect $g = 0.69$, 95% CI [0.58, 0.79], $p < .001$ (see Table S2 for comparison with CHE and other models). The model used in Bediou et al. (2018; RVE with hierarchical weights), produced an average effect $g = 0.54$, 95% CI [0.42, 0.66], $p < .001$.

Sensitivity to Outliers

The main analysis included 22 winsorized effects (10%) with a mean effect size of $g = 1.18$ before winsorizing, and $g = 0.84$ after correction (compared to a mean $g = 0.60$ for the 190 nonwinsorized effects). Using nonwinsorized effect sizes the average effect of AVG in cross-sectional studies was $g = 0.64$, 95% CI [0.53, 0.75], $p < .001$, $\tau^2 = .14$, $\omega^2 = .43$ (Table S2). This means that 95% of the effect sizes could be predicted to be between -0.85 and 2.13 . Including outliers also affected the moderator results.

Recoding Spatial and Verbal Tasks With a Working Memory Component

The results were identical when tasks were coded based on the type of material (e.g., verbal, spatial) or the presence of working memory demands. The average effect was unchanged (Table S2). The moderator analysis indicated a strong effect of AVG experience on working memory measures, $g = 0.64$, 95% CI [0.54, 0.75], $k = 200$, $m = 70$, $\tau^2 = .16$, $\omega^2 = .34$, $p < .001$, 95% Precision Interval, PI [-0.75, 2.08]. The effect of recruitment also became significant, AHT $F(8,11) = 5.90$, $p = .041$, indicating stronger effects with overt recruitment ($g = 0.67$, 95% CI [0.55, 0.78], $p < .001$) compared to covert recruitment, $g = 0.44$, 95% CI [0.24, 0.64], $p < .001$. Importantly, significant effects for both types of recruitment. This result provides partial support of the expectation hypothesis (Table S3).

Controlling for Joint Publication Group

We added an additional random factor to control for differences between JPGs and repeated the analysis.^{A1} Accounting for differences between groups slightly reduced the average estimate ($g = 0.55$, 95% CI [0.42, 0.67], $p < .001$, $k = 212$, $m_paper = 70$, $m_jpg = 32$) without affecting significance ($p < .001$) or residual heterogeneity, $Q(212) = 1,391$, $p < .001$ (Table S2). Including the additional JPG clustering level (32 unique levels) did not account for any additional heterogeneity between effect sizes (Table S5). Controlling for the JPG also changed some moderator results: the effect for inhibition and problem-solving became nonsignificant (Table S4). In line with

our preregistered methods, to further examined whether differences in methods used by specific research groups, based on coauthorship, can account for differences between the effect sizes (Bediou et al., 2018; Hilgard et al., 2019), we also ran a full model including an additional moderator for “JPG,” which aimed at capturing differences in methodological approaches (e.g., recruitment criteria, task duration, etc.). We first assessed the impact of JPG as in Bediou et al. (2018) considering Bavelier lab versus other labs. This analysis showed that studies conducted by the Bavelier lab ($g = 0.85$, CI [0.63, 1.07], $p < .001$, $k = 69$, $m = 23$) report significantly stronger effects compared to Other labs ($g = 0.53$, CI [0.41, 0.64], $p < .001$, $k = 143$, $m = 47$), AHT-F test $F(34,74) = 6.65$, $p = .014$. To go beyond this dichotomous approach, we performed additional exploratory analysis that grouped together all studies that had one author in common. Because the analysis is limited to 12 categories, studies from authors that contributed only one article were included as “Other.” With these 10 clusters (grouping all unique studies as one cluster), no significant effect of Lab was observed, however, both the AHT-F test as well as seven of the 10 clusters had low $df < 4$. To further reduce the number of clusters (and consequently increase the degrees of freedom), we then iteratively grouped together authors that were associated with low df in a full moderator model. This led us to a model with five clusters in all labs reported significant effect despite some differences in magnitude as indicated by a significant AHT-F(18,23) = 4.22, $p = .014$. Adding this moderator did not affect the pattern of effect of the other moderators.

^{A1} This corresponds to a hierarchical structure with three nested levels in which effect sizes are embedded in papers, which are clustered into JPGs.

Appendix B

Supplementary Analyses of Intervention Studies

Publication Bias Sensitivity Analyses

We followed the same approach that we used to assess sensitivity to publication bias in our cross-sectional meta-analysis. As can be seen in Table S6, only two of the 16 analyses detected significant publication bias. The most notable exception to this pattern was the Egger’s sandwich test, which uses robust variance estimation for correlated effects (Table S6). We suggest a cautious interpretation of the significance of this test for two reasons. First, we believe the CHE model more closely aligns with the structure of this meta-analytic sample, with dependence resulting from multiple effect sizes per study (correlated) and studies conducted by the same lab or research group (hierarchical). Additionally, the significance of the modified measure of precision has low degrees of freedom (less than 4) and this has been shown to increase false positive results. Tipton and Pustejovsky (2015) suggest using an α level of .01 or lower instead of .05 when interpreting tests with degrees of freedom less than 4. A cumulative forest plot was also generated to visualize the changes in average effect size following each new included study, chronologically (Figure S7)

Choice of Meta-Analytic Models

We preregistered an analysis based on RVE models with correlated weights. However, because the CHE model was better suited to

the structure of our data set, we decided to report this model as our primary analysis which resulted in an average effect size $g = 0.30$, 95% CI [0.11, 0.50]. The average effect size estimate from the preregistered RVE model with correlated weights was very similar with an $g = 0.33$, 95% CI [0.13, 0.52], $p = .003$ (see Table S7, for comparison with the CHE and other models).

Sensitivity to Outliers

Our reported primary analyses winsorized 10 effect sizes with a mean of $g = -0.66$ before winsorizing, and $g = 0.303$ after winsorizing (compared to a mean $g = 0.379$ for the 73 nonwinsorized effects). When nonwinsorized effect sizes were used, the average effect of action game on cognition in intervention studies increased to $g = 0.42$, 95% CI [0.09, 0.75], $k = 83$, $m = 18$, $p = .017$, $\tau^2 = .23$, $\omega^2 = 1.03$, 95% PI [-1.78, 2.62], but both the confidence interval and the prediction interval were also larger (Table S6). It is noteworthy that one of our effect sizes was both extremely large and negative ($g = -10.58$, $N = 20$). This was obtained from Strobach et al. (2012) in the single task condition of their verbal task (odd/even or consonant/vowel discrimination). The Tetris group showed higher error rates at pretest leading to larger reduction in error rates in this group compared to the Medal of Honor (and consequently larger variance at pretest too). The high estimates are

due to standardizing based on the posttest and not the pretest. Using the variance of the pretest results in a much smaller estimate of $g = 2.37$. Importantly though, an analysis performed on effect sizes standardized using the SD of the pretest showed similar results (Table S7).

Recoding Spatial and Verbal Tasks With a Working Memory Component

Considering the working memory component of spatial and verbal tasks, resulted in an average effect of AVG play that was very close to the main model, $g = 0.38$, 95% CI [0.20, 0.59], $k = 77$, $m = 18$, $p < .001$ (Table S7). The pattern of moderator effects was unchanged with a significant effect of AVG interventions on improving working memory, $g = 0.40$, 95% CI [0.01, 0.79], $p = .045$, $k = 17$, $m = 7$, $\tau^2 = .00$, $\omega^2 = 1.41$, 95% PI [−0.87, 1.63] (Table S6). Moderator results were also unchanged (Table S8).

Controlling for JPG

Adding a random factor to account for differences between JPGs resulted in a numerically smaller estimate, $g = 0.25$, 95% CI [0.03, 0.47], $k = 83$, $m = 18$, $JPG = 10$, $p = .03$, $\tau^2 = .00$, $\omega^2 = 0.27$, 95% PI [−0.76, 1.26] (Table S7). The pattern of moderator effects was also unchanged (Table S9).

Effect Size Standardization in Intervention Studies

The effect sizes reported in the main analyses were obtained by dividing the difference between posttest and pretest means (i.e., difference in change scores) by the pooled standard deviation at posttest (Morris & DeShon, 2002). Some intervention studies reported the standard deviation (SD) of change scores rather than the pretest and posttest standard. Although it is possible to compute the standard deviation of a difference from the standard deviations of pretest or posttest, this requires knowing the correlation between pretest and posttest, which is often not reported in primary studies. Alternatively, methodologists have suggested using the neutral value of $\rho = .5$ for the correlation between pre- and postmeasures. This strategy is thought to provide the least biased estimate because it substitutes the variance of pretest or posttest with the variance of the pre–post difference score. Several authors (Hirst et al., 2018) have argued that this method should be privileged over most of the alternative approaches that have been proposed. We used the same strategy as in Hirst et al. (2018) and adapted the formula for independent groups, which is equivalent to either ignoring the pre–post correlation or assuming a value of $\rho = .5$ for both the effect size estimate and variance of the effect size. This is a more conservative approach than other approximations and tends to produce larger confidence intervals around each estimate.

Several methods have been proposed to compute effect size in pre–post between group designs when the pre–post correlation is missing. The first approach proposed by Morris and DeShon (2002,

equation 6, p. 108) is obtained by computing the effect sizes of pre–post changes in each group and then subtracting the effect size of the control from the effect size of the experimental group. This approach takes into account the within-subject design but does not correct for between group differences in variance. Another approach relies on sensitivity analyses using different values for the pre–post correlation to more systematically examine their impact on the results. However, this makes several assumptions that are likely to be incorrect or at least unverifiable, one of which is that the correlation coefficient is constant across studies and measures. Moreover, in its latest recommendations, the What Works Clearinghouse has argued against this strategy (*What Works Clearinghouse Procedures Handbook, Version 4.1*, 2020) which can introduce bias by ignoring that some coefficient values are more representative than others, especially in the context of a within-subject pre–post cognitive intervention design. For examples, values of $\rho < .5$ mean that the variability (SD) of the difference scores is larger than the average of pretest and posttest, resulting in an underestimated effect. Conversely, a value of $\rho > .5$, which is reasonable to expect in the context of pre–post intervention designs, will result in lower variability for difference scores than for pretest or posttest, which in turn leads to in overestimated (i.e., less conservative) effect sizes. As such, using a neutral value of $r = .5$ is most likely to provide a least biased and relatively more conservative estimate.

Meta-analytic results are sensitive to the methods used to calculate effect size variance and thus the related weights in effect size estimation. The selection of the standard deviation used to calculate the variance of an effect size has implications for standardization. In this study we used the SD of the posttest means,^{B1} instead of the pooled SD of the pre and posttest difference scores used in Bediou et al. (2018).

To test the impact of the standardization method, we conducted sensitivity analysis on the subset of effect sizes that were included in both the present meta-analysis and our previous one (Bediou et al., 2018). Specifically, we examined the differences in the effect size estimates and related standard errors when the distinct standardization methods are used. As can be seen in Table S7, the analysis resulted in smaller average estimates when using the SD of the posttest ($g = 0.38$, $p = .001$, $k = 82$, $m = 21$, $\tau^2 = .00$, $\omega^2 = 0.42$, 95% PI [−0.96, 1.57]) compared to the SD of pre–post differences ($g = 0.52$, $p < .001$, $k = 71$, $m = 19$, $\tau^2 = .00$, $\omega^2 = 0.40$, 95% PI [−0.86, 1.62]). Note that for this analysis, the clustering was performed at the level of the article instead of the participants' sample.

^{B1} Standardization using the SD of pretest instead of posttest did not change the results (Table S2).

Received June 21, 2021

Revision received December 8, 2022

Accepted December 12, 2022 ■