Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills

BEDIOU, Benoît, et al.

Abstract
The ubiquity of video games in today's society has led to significant interest in their impact on the brain and behavior and in the possibility of harnessing games for good. The present meta-analyses focus on one specific game genre that has been of particular interest to the scientific community—action video games, and cover the period 2000–2015. To assess the long-lasting impact of action video game play on various domains of cognition, we first consider cross-sectional studies that inform us about the cognitive profile of habitual action video game players, and document a positive average effect of about half a standard deviation (g = 0.55). We then turn to long-term intervention studies that inform us about the possibility of causally inducing changes in cognition via playing action video games, and show a smaller average effect of a third of a standard deviation (g = 0.34). Because only intervention studies using other commercially available video game genres as controls were included, this latter result highlights the fact that not all games equally impact cognition. Moderator analyses indicated that action [...]
The ubiquity of video games in today’s society has led to significant interest in their impact on the brain and behavior and in the possibility of harnessing games for good. The present meta-analyses focus on one specific game genre that has been of particular interest to the scientific community—action video games, and cover the period 2000–2015. To assess the long-lasting impact of action video game play on various domains of cognition, we first consider cross-sectional studies that inform us about the cognitive profile of habitual action video game players, and document a positive average effect of about half a standard deviation ($g = 0.55$). We then turn to long-term intervention studies that inform us about the possibility of causally inducing changes in cognition via playing action video games, and show a smaller average effect of a third of a standard deviation ($g = 0.34$). Because only intervention studies using other commercially available video game genres as controls were included, this latter result highlights the fact that not all games equally impact cognition. Moderator analyses indicated that action video game play robustly enhances the domains of top-down attention and spatial cognition, with encouraging signs for perception. Publication bias remains, however, a threat with average effects in the published literature estimated to be 30% larger than in the full literature. As a result, we encourage the field to conduct larger cohort studies and more intervention studies, especially those with more than 30 hours of training.

Public Significance Statement
Understanding the effects of action video game play is essential given that (a) a large number of individuals regularly spend many hours on these types of games, and (b) proponents are offering suites of video games that are claimed to change behavior or enhance cognition. The 2 meta-analyses in this paper present the current status of this field, concluding that playing action video games has some positive effects on improving cognitive skills. This review also identifies some limitations of current research.

Keywords: action video games, attention, cognition, meta-analysis, perception

Supplemental materials: http://dx.doi.org/10.1037/bul0000130.supp
Over the past 20 years, there has been significant scientific interest in examining the potential behavioral consequences of playing video games. This research is still in its relative infancy, but one repeatedly observed finding, from social psychology to clinical psychology to educational psychology, to the focus of this meta-analysis, cognitive psychology, is that not all video games have the same impact. Indeed, given the enormous range of completely different experiences that fall under the label of video games, attempting to identify how playing video games affects behavior is analogous to attempting to identify how eating food impacts physiology. The current study therefore focuses on one specific game genre, termed action video games, for which there are now sufficient data to examine impact via meta-analytic techniques, covering the period from year 2000 until 2015. Here we thus ask, “Do people learn anything useful from playing action video games?”

The rationale for this study is both practical and theoretical, placing it in the category of what Stokes (1997) calls use-inspired basic research. Specifically, whether people learn anything useful from playing action video games is an important practical question in light of the large number of individuals who play action video games for extended periods of time. Indeed, according to recent reports, more than 1.2 billion individuals worldwide (including more than 150 million Americans) are video gamers, with action games consistently ranking at or near the top for most popular game type (Spil Games, 2013; The Entertainment Software Association, 2015). Perhaps not surprisingly then, as video gaming has surged in popularity, so too has popular interest in the potential practical ramifications of such gaming on our everyday lives. Perhaps not surprisingly then, as video gaming has surged in popularity, so too has popular interest in the potential practical ramifications of such gaming on our everyday lives (Bavelier & Green, 2016). Examining the cognitive profile of habitual action video game players, who have spent hundreds of hours playing this particular type of games, is an important first step for understanding the long-lasting cognitive changes associated with action video game play. This is not the only practical concern related to the impact of action video games. For instance, many research groups have started to use off-the-shelf commercial action video games in translational applications (Mayer, 2014, 2016). Thus, confidence in the overall accuracy of that foundational science is critical, as has been recently highlighted in the context of brain training games (Simons et al., 2016).

It is also an important theoretical question in light of the proposal that the general skills learned from playing in a game context can transfer to nongame contexts that require the same underlying skills (Green & Bavelier, 2012; Mayer, 2014; Sims & Mayer, 2002). Such growing interest into the gamification of various interventions is exemplified by the recent surge of new genres of games, such as therapeutic or educational serious games, games for impact, crowdsourcing games, or ‘so-called’ brain-training games. Here, we used a meta-analytic approach to paint a more coherent and global picture of the effects of playing action video games.

**Action Video Games**

As noted above, there are many types of video games that can differ dramatically from one another. Video games within the action genre all share a set of qualitative features, such as: (a) a fast pace (in terms of the speed of moving objects, the presence of many highly transient events, and the need to make motor responses under severe time constraints); (b) a high degree of perceptual and motor load, but also working memory, planning and goal setting (e.g., many items to keep track of simultaneously, many possible goal states that need to be constantly reevaluated, many motor plans that need to be executed rapidly); (c) an emphasis on constantly switching between a highly focused state of attention (e.g., toward aimed targets) and a more distributed state of attention (e.g., to monitor the whole field of view); and (d) a high degree of clutter and distraction (i.e., items of interest are distributed among many nontarget items). The main subtypes of games generally considered to be action video games include first-person shooter games, wherein the player views the world through the eyes of his or her avatar (e.g., the Halo, Call of Duty, and Medal of Honor series of games) and third-person shooter games, wherein the player sees the back of his or her avatar (e.g., the Gears of War and Grand Theft Auto series of games). Thus, games such as Rise of Nations, Pac-Man, or Space Fortress, although occasionally also given the label of action games in various parts of the literature, do not qualify as action games for the purposes of this meta-analysis because they lack the game mechanics described above.

It should be noted that although the action game features listed above were reasonably unique to action video games circa the year 2000, today some, albeit not all, of these critical characteristics can now be found in other genres such as multiplayer online battle arena games, real-time strategy games, or role-playing games (Dale & Green, 2017). In the present work, however, we use a narrow definition of action video game, whereby genres such as real-time strategy, role-playing, and fighting video games do not qualify, as this is the definition that has been employed in the literature to date.

**Previous Syntheses of Gaming and/or Action Gaming Research**

The use of the meta-analytic approach, combined with a focus on action video games and their impact on cognition, departs from other literature reviews that provide qualitative descriptions of research studies rather than measures of effect size (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Hays, 2005; Honey & Hilton, 2011; Randel, Morris, Wetzel, & Whitehill, 1992; Tobias, Fletcher, & Bediou, 2015; Tobias, Fletcher, Bediou, Wind, & Chen, 2014; Tobias, Fletcher, Dai, & Wind, 2011; Young et al., 2012) or that focus on academic learning outcomes rather than on skills (D. B. Clark, Tanner-Smith, & Killingsworth, 2014; Sitzmann, 2011; Vogel et al., 2006). A few reviews focusing more squarely on cognition and carried out at the level of all video games have certainly provided interesting pointers (Adams & Mayer, 2014; Connolly et al., 2012; Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013; Simons et al., 2016; Tobias et al., 2011; Young et al., 2012). Yet, it is important to consider that the label of video games applies equally well to experiences that may differ dramatically from one another. Experiences that fit the superordinate category label of video games may involve incredibly simple graphics or alternatively may involve highly realistic and lifelike environments, solitary single-player activities or rich social structures, minutes-long play or years-on-end play, mostly reactive activities or substantial long-term planning, distinctly pro-social behavior or rather antisocial behavior, elaborate game mechanics...
or nothing more than a set of rules, pure entertainment or primarily translational, and so forth. In short, the extreme flexibility of the media allows for all kinds of experiences. Given that the behavioral outcomes of an experience depend strongly on the nature of the experience (D. B. Clark et al., 2014), a superordinate category label such as video games is likely to have limited predictive power. Indeed, it seems highly unlikely that a slow-paced game with essentially no perceptual, attentional, or cognitive dimensions (e.g., Farmville) would produce the same outcomes as a fast-paced game that places heavy demands on the perceptual, attentional, and cognitive systems (e.g., Call of Duty) – despite both clearly being video games.

To date, only three published meta-analyses have examined the effect of playing specific video game genres on cognition. The first two come from the group of Powers and colleagues (Powers & Brooks, 2014; Powers et al., 2013). In their first meta-analysis, Powers et al. (2013) included all video games and coded for game genre as a moderator (using 5 categories: action/violent, mimetic, nonaction, puzzle, nonspecific). This moderator was found to have a significant effect, indicating differences between game genres, although the exact source of that effect was difficult to pinpoint (see their Table 5 on page 1067). Interestingly, the effect of action video games was significant in both quasi-experiments (i.e., cross-sectional studies comparing selected groups of habitual players of action video games with nonvideo game players, matched for as many factors as possible other than their video gaming habits) and in true or intervention studies examining whether video game training produced changes in performance.

In their subsequent report (Powers & Brooks, 2014), the authors focused squarely on intervention studies alone. This reanalysis confirmed an effect of First Person Shooter games (more in line with our definition of action video games), and demonstrated that training with these games produced a significant overall improvement in cognition ($d = 0.23$, 95% CI [0.07, 0.39], $p = .005$). The authors also further examined a number of subdomains of cognition, with significant effects being found for perceptual skills ($k = 35$, $d = 0.45$, 95% CI [0.17, 0.72], $p = .001$), and spatial imagery ($k = 11$, $d = 0.17$, 95% CI [0.01, 0.34], $p = .04$), but not for executive functions ($k = 10$, $d = −0.17$, 95% CI [−0.47, 0.14], $p = .28$) or for motor skills ($k = 1$, $d = 0.07$, 95% CI [−0.31, 0.45], $p = .72$).

A third and more recent meta-analysis by Wang et al. (2017) arrived at the same conclusion about the positive effects of action video game training, despite using a quite different definition of action video games. In particular, Wang et al. (2017) included games such as Pac-Man that would clearly not be considered action games in the present meta-analysis or in the meta-analyses by Powers and colleagues. As different games may have different impact on cognition, the present meta-analyses stick to a more coherent definition of action video games, as described above.

The present study thus builds on these previous meta-analyses, while adding a number of novel aspects. The first notable difference lies in the strict definition of action video games used in the present analyses, which imposed different constraints on our literature search and selection criteria. For example, our study covers the period from year 2000–2015, whereas Powers and colleagues (Powers & Brooks, 2014; Powers et al., 2013) covered from 1980 to 2012.

The inclusion of three extra years at the end of the range (i.e., 2013–2015) is of clear importance, as the rapid growth of publications on video games over the past few years means that 48% of the studies we included in the current analyses (2013–2015) were not included in Powers et al. (2013, 2014). However, the difference in start date is perhaps even more important. In particular, by considering only work starting with the year 2000, we ensure that we are categorizing under the action video game genre relatively homogenous game experiences. Indeed, games that are recognizable in terms of modern action mechanics only came into being in the late-1990s (this includes acclaimed titles such as Doom (1993), Goldeneye (1997), Half-Life (1998), and Counterstrike (1999)). Although games today have obvious graphical and computational advantages as compared with the games of the late 1990s and early 2000s, the core components are nonetheless strongly shared across this range (i.e., most of the base mechanics as related to movement and aiming are essentially identical). This is not true though of games from the early 1980s because most action-like games from that era may include games such as Centipede or Super Mario Bros., which bear little resemblance in terms of action mechanics and content to modern action games. Thus, lumping games together from as early as the 1980s will also tend to confound efforts to examine the impact of what are now considered key action game characteristics.

We have also taken special care to not conflate action and violent video games. A relatively common misbelief about the world of gaming is that these terms are interchangeable; yet they are not. Simply put, not all action games contain violence and not all games that contain violence are action games. It is undeniably the case that most first-person and third-person shooter games are violent. Yet, violence is not essential to the action mechanics we describe above. As such, it is perfectly possible for a game to utilize action mechanics and dynamics in the absence of any violent content (e.g., as the case in cooperative paintball games such as Splatoo, or child friendly shooter games such as Rayman’s Raving Rabbids). It is also equally possible to have violent content in games without any action characteristics (e.g., many turn-based role playing games, such as Final Fantasy VII, have a great deal of violence in the absence of any action components). This latter point is particularly critical with respect to the creation of a combined action/violent category, as the presence of violent, but not action games in an analysis can potentially confound attempts to isolate the effect of action games.

Our meta-analyses further depart from previous meta-analyses on the impact of action video games by restricting intervention studies to only those that used other commercially available game genres as controls. This point is critical if we are to understand the specific features of action games that impact cognition. It also sets a high bar for observing an effect, as intervention studies with well-matched active controls are known to result in smaller effect sizes (Uttal et al., 2013). For example, our meta-analytic approach includes studies comparing the impact of action video games to that of Tetris on spatial cognition, when Tetris has been documented to improve spatial cognition (Uttal et al., 2013).

Furthermore, another unique aspect of our meta-analytic approach is to focus on the long-lasting impact of action video game play on cognition. In particular, intervention studies were restricted only to those studies that tested impact at least 24 hours after training completion, thus ruling out a number of potentially fleeting confounds.
(e.g., improvements attributable only to arousal). As long-lasting plastic changes are notoriously hard to induce, our meta-analysis is quite representative of the field by also requiring at least 8 hours of training distributed over 8 days.

Finally, the present meta-analysis makes use of recent meta-analysis methods (robust variance estimation, hierarchical model, multiple moderator analysis) to examine a host of previously unaddressed methodological issues both with respect to the action gaming literature itself, such as the impact of moderators related to motivation or expectation biases, as well as with respect to issues unique to meta-analyses, such as publication bias and small study effects (see below for additional description).

Key Issues to Be Considered

Which Cognitive Domains Are/ Are Not Impacted by Action Gaming?

The extent to which a given experience alters behavior should depend strongly on the match between the experience and the neural process underlying the measured behavior as well as the extent to which the processes are themselves plastic. The mechan- ics and dynamics inherent in action games do not place equivalent load on all cognitive domains (e.g., strong load is placed on processes related to top-down attention, perception, and multitasking, but there is little to no verbal cognition at play; Spence & Feng, 2010). Therefore, it would be surprising if all cognitive domains were equally altered—or altered at all—by action gaming. We use the broad term cognitive domain to refer to all aspects of cognition, including perceptual skills, attentional skills, and cognitive skills.

Understanding the cognitive domains that are and are not modified by action gaming is not only important for our theoretical understanding of how action game play promotes behavioral changes, but is also particularly crucial for those researchers attempting to utilize action games for practical ends. For example, some reports of action video game play improving mental rotation (Feng, Spence, & Pratt, 2007), have led to the proposal that these games may be useful for education in the sciences, technology, engineering, and mathematics disciplines (Uttal et al., 2013). Similarly, reports of faster and more accurate visuomotor control after action game play has led to studies probing their usefulness to train laparoscopic surgeons to perform surgeries faster without making more errors (Schlickum, Hedman, Enochsson, Kjellin, & Fellander-Tsai, 2009). Reports of enhanced perceptual (primarily visual) processing after action gaming has led to studies assessing their utility in the rehabilitation of visual disorders such as amblyopia (J. Li et al., 2015; R. W. Li, Ngo, Nguyen, & Levi, 2011; Vedamurthy, Nahum, Bavelier, & Levi, 2015). Finally, reports of enhanced visual attention after action game play have led researchers to use these games to train Italian dyslexic children, for whom attention appears to be one of the major bottlenecks that constrains the fluency of reading (Franceschini et al., 2013).

Given this strong movement toward translational work, it would thus be reassuring to validate that these games have an impact on cognition, and identify which skills may be more reliably affected. To this end, we categorized measures of cognition into the following eight cognitive domains, guided by available data: (a) perception (e.g., contrast sensitivity, lateral masking), (b) bottom-up attention (e.g., pop-out search, exogenous cueing), (c) top-down attention (e.g., complex search, flanker tasks, multiple object tracking), (d) spatial cognition (e.g., mental rotation, spatial working memory tasks), (e) task-switching/multitasking (e.g., dual-task or task-switch paradigms), (f) inhibition (e.g., go-nogo, stopsignal tasks, proactive interference), (g) problem solving (e.g., Tower of Hanoi, Tower of London, Raven’s matrices), and (h) verbal cognition (e.g., verbal working memory, reading).

Type of Study Design: Cross-Sectional and Intervention Studies

Studies examining the long-lasting effects of action gaming on cognition have typically taken one of two forms. The first type are cross-sectional designs. Here, the performance of self-selected individuals who naturally play a large amount of action video games (often labeled to as action video game players or AVGPs) is contrasted with the performance of individuals who specifically do not play those kinds of fast-paced action-packed video games and rarely, if at all, play other nonaction types of games (often referred to as nonvideo game players or NVGPs). Although these groups differ in their video game play, they are matched along many potentially confounding dimensions as possible, including factors such as age-range, gender, and years of education. The critical measure in these studies is thus related to a difference in performance between these two extreme self-selected groups (i.e., whether AVGPs show better performance than NVGPs).

Although self-selection bias is always a concern for such cross-sectional studies, they nonetheless serve a dual role in the characterization of the long-lasting effects of action video game play on cognition. First, they document the cognitive profile of a growing segment of the population (AVGPs), an important societal question. Second, they provide a useful pointer as to where it may be worth investing in a training study. Indeed, documenting different cognitive skill between AVGPs and NVGPs clearly calls for an intervention study. In contrast, if hundreds of hours of action game play over months to years do not result in a group difference in the cognitive skill tested, the expectation that durable changes of that cognitive skill will be observed after only tens of hours of video game play is lessened. Accordingly, 65% of the intervention studies identified in the literature first established a cross-sectional effect of action video game play.

The second broad type of study involves intervention studies (also called true experiments). Here, individuals who do not, as part of their normal life, tend to play much video games are specifically trained on either an action video game or a control video game with performance on the skills of interest being measured both before and after training. The critical measure in these studies is thus related to a difference of differences—specifically assessing whether the action trained group showed greater improvements from pretest to posttest than the control trained groups.

Among intervention studies, we focus exclusively on those which contrasted training on an action video game with training on a control video game. As is standard in the field, the experimental action and the control, nonaction games were required to be commercially available games, ensuring that both arms of the study were engaged in a quality, entertaining and challenging video game experience. Studies using, either no control, passive controls or repeated practice on the task of interest (e.g., group
several key dimensions of interest. First, to the extent that the stimula-
tion is of appropriate difficulty and specifically targeted toward the
to-be-enhanced skills, it seems computer- or game-based training induces small sized benefits in
older adults, especially in verbal memory, speed of processing, and
verbal/spatial working memory with the impact on attention and
executive functions being less reliable (Ball et al., 2002; Karr,
Areshenkoff, Rast, & García-Barrera, 2014; Lampit, Hallock, &
Valenzuela, 2014; Toril, Reales, & Ballesteros, 2014; Wang et al.,
2017). In contrast to the experiences above, which were carefully
titrated to the abilities of older participants, action video games are
designed to be challenging for young individuals, and perhaps
more to the point, young individuals who are already well versed
in the demands of action games. Not surprisingly, such games are
typically too challenging for older adults. As a result, very few
studies have used action video games as defined above as an
intervention tool in older adults (i.e., 2 records; 12 effect sizes).
Given the importance of providing users with a training regimen
within their proximal zone of development to induce learning, one
can expect rather different outcomes of training with an action
video game in young and older adults. The present analysis is
poised to shed some light on this issue.

Type of Dependent Measure

The most common dependent measures in cognitive psychology
are accuracy and reaction time (RT). Many tasks involve collecting
only one or the other. Even in those tasks that nominally measure
both accuracy and RT, it is typically the case that one is the
primary measure of interest, with the other being used mainly to
rule out potential confounds such as speed–accuracy trade-offs. Of
interest here is whether RT or accuracy is differentially affected by
action video game play. Specifically, by examining whether the
type of dependent variable used in studies moderates the size of the
action video game play effect, we can test the popular belief that
action video game play involves a certain degree of “trigger
happy” behavior, whereby speed is valued over accuracy. If so, we
would expect to see larger effect sizes in speed as compared with
accuracy measures. If, however, the effects of action games are
truly at the level of the core processes themselves, then we would
expect to see effects on both accuracy and RT.

Main Versus Difference Effects

There have been a number of recently published criticisms
suggesting that the effects attributed to action video game experi-
ence in the literature could instead be ascribed to participant
expectation effects (Boot, Blakely, & Simons, 2011; Boot, Simons,
Stothart, & Stutts, 2013; Kristjansson, 2013, but see Bisoglio et al.,
2014; Green et al., 2014). A first foray in assessing this expectation
bias hypothesis has led us to separately code studies where the
effect of interest is a main effect (e.g., overall change in RT or
accuracy) versus studies where the effect of interest is a difference
between conditions (e.g., disproportionately faster RTs or higher
accuracy in some conditions vs. others). Main effect hypotheses
such as “action game players should respond faster” are more
likely for a naïve participant to intuit, as well as potentially easier
to match behavior to, than a difference effect hypotheses such as
“action game players should respond disproportionately faster on
switch trials than on non-switch trials.” Thus, if action video game

Participant Age

Most available studies examining the impact of action video
games on cognitive function have focused on college-aged indi-
viduals. A few studies, though, have examined the effect of action
video game play in normal children (under age 18) or in older
adults. Although the current literature is not ideal with respect to
exploring effects from a life span perspective, in that there are no
cross-sectional studies including older adults and no intervention
studies including children, we can nonetheless look at possible age
effects within each of these types of studies.

In general, plastic changes are typically greatest in children and
then decrease in magnitude with aging. Recent work though high-
lights the potential for brain plasticity throughout the life span,
even into old age (Mahncke, Bronstone, & Merzenich, 2006).

receiving Useful Field of View [UFOV] training in Belchior et al.,
2013), are not included.

Focusing exclusively on intervention studies with an entertain-
ment quality video game control group makes the present meta-
analysis quite unique in comparison to those of others (Powers et
al., 2013; Powers & Brooks, 2014; Wang et al., 2017). As noted
previously, it also sets a high bar for the magnitude of effects that
must be produced by action video game training. Indeed, not only
is it the case that recent work indicates that some game genres that
have, in some studies, been used as a control may also enhance
cognition (Blumen, Gopher, Steinerman, & Stern, 2010; Glass,
Maddox, & Love, 2013; Powers & Brooks, 2014; Powers et al.,
2013; Wang et al., 2017), it is also the case that, more generally,
effect sizes of interventions with well-matched active control are
typically smaller than those obtained when including all interven-
tion studies (Uttal et al., 2013). Thus, the present meta-analysis of
intervention studies is a departure from most, if not all intervention
meta-analytic work published thus far as the control groups con-
sidered here go well beyond a simple placebo group (e.g., being
given an inert pill) to control for possible expectation effects.
Instead, many of the video games used as controls provide rich
immersive experiences that are likely to have an impact of their
own, whether on cognition or on other aspects of behavior.

Importantly, active control intervention studies fall under the
label of what Mayer (2011, 2014) has called cognitive conse-
quences research, which is essential in defining the impact of the
video game genre studied on various cognitive outcomes, as our
aim is here (see also Boot & Simons, 2012; Green, Strobach, &
Schubert, 2014; Jacoby & Ahissar, 2013; Schellenberg & Weiss,
2013 for best practices in behavioral intervention studies). Unlike
in cross-sectional studies, all participants in an intervention study
are recruited using the same method, and participants are then
randomly assigned to their treatment group. Therefore, recruitment
method is not a relevant moderator for these types of studies.
Conversely, training duration only applies to intervention studies.
These constraints determine the moderators considered below dur-
ing the analysis.

A number of factors are likely to moderate the effects of action
games, some of which relate to participant characteristics (e.g.,
age), whereas others relate to methodological aspects (e.g., recruit-
ment, type of measure, or duration of training). One of our aims is
to take advantage of the meta-analytic approach to examine how
different factors may alter the impact of action video game play.
effects result from participants trying to match their behavior to what is expected of their assigned group, we would predict larger group differences in studies where the effect of interest was a main effect and smaller group differences in cases where the effect of interest was a difference term. The strength of action video game play on difference effects will thus provide an indirect assessment of the validity of the expectation bias hypothesis.

**Recruitment Method in Cross-Sectional Studies**

Although the analysis of main versus difference effects is partially relevant to this point, a more direct assessment of the as-yet-untested expectation hypothesis comes from the analysis of recruitment methods. Specifically, many studies comparing action video game players with nonaction players in the literature have employed overt recruitment methods (e.g., posters asking about gaming habits). Because the participants in these studies know that they have been selected based upon their gaming habits, expectation effects are possible. Other studies though have employed purely covert recruitment methods (i.e., participants do not know they are being selected based on their gaming habits). Expectation effects are unlikely, because the participants in these studies do not know that gaming habits are of interest. If knowledge of one’s video game status drives the entirety of the reported effects on action video game play (which we label here as the most extreme form of the expectation hypothesis), then no effect of action gaming status should be observed in studies that have employed covert methods.

**Training Duration in Intervention Studies**

Intervention studies vary in terms of training duration. Because time-on-task is a significant predictor of learning, studies using short training durations may report smaller effects than studies with longer training duration. Here we used metaregression methods to test whether training duration is linearly related to the strength of action video game effects.

**Laboratory**

Most of the earliest work around the topic of action games was performed by the Bavelier laboratory. Although work from this laboratory now makes up a considerably smaller total percentage of the literature, it is nevertheless the case that whenever a single study is labeled as the Bavelier laboratory, it includes additional domains such as top-down attention and verbal cognition. Fourth, we examine the impact of a number of moderators, in particular that of possible expectation biases which have been the focus of a number of recent critiques of this work, and were not addressed in previous meta-analyses. Finally, we take full advantage of the hierarchical structure of our data set and invested specific effort to explore methodological issues related to small study effects and moderator effects.

**The Current Meta-Analyses**

In sum, the present meta-analyses focus on the impact of action video games on behavior considering a range of cognitive domains and does so separately for cross-sectional and intervention designs. We also address timely issues in the field such as the relative impact of game play on measures of RTs versus accuracy, potential confounding factors such as those related to the expectation bias hypothesis, as well as the impact of training duration in intervention studies.

Our meta-analyses thus not only extend prior meta-analyses (e.g., Powers & Brooks, 2014; Powers et al., 2013; Toril et al., 2014; Wang et al., 2017), but they also depart from previous work on methodological and theoretical grounds. First, our meta-analyses focus on the action video game genre defined based on specific mechanics likely to have differential impact across cognitive domains. This is important, as both the two meta-analytic studies of Powers and colleagues (2013, 2014) and our own studies (Cohen, Green, & Bavelier, 2007) have indicated that not all video games have the same impact on different aspects of cognition. Second, we perform separate meta-analyses of cross-sectional and intervention studies. By including only intervention studies with active controls that made use of commercially available video games, the current analysis of action video game impact is unique in addressing a number of confounding variables related to possible differences in novelty, engagement, motivation, and fun, as both experimental and control groups are faced with a commercial grade experience. Third, we provide a more fine-grained analysis of impact across different domains of cognition, which diverges from the classification chosen by Powers and colleagues, and includes additional domains such as top-down attention and verbal cognition. Fourth, we examine the impact of a number of moderators, in particular that of possible expectation biases which have been the focus of a number of recent critiques of this work, and were not addressed in previous meta-analyses. Finally, we take full advantage of the hierarchical structure of our data set and invested specific effort to explore methodological issues related to small study effects and moderator effects.

**Method**

**Study Selection**

The literature search covered the period between January 2000 and November 2015. We queried the databases PsycINFO, PsYINFO, ERIC, FRANCIS, MEDLINE, SCOPUS, Web of Science and ScienceDirect, using terms combined in the following Boolean expression (“video game” OR “computer game”) AND (“attention” OR “attentional” OR “attend” OR “cognitive” OR “perception” OR “perceptual”), or else repeating the search with different combinations when the Boolean search was not permitted. Databases were either interrogated using their web interface, or using multidatabase search interfaces such as Ovid, EBSCO and PROQUEST. Results from the database PsycINFO can be viewed via the link provided in the supplementary material.

The validity of a systematic review or meta-analysis is highly dependent on the underlying data, and more specifically the ability to reduce the potential sources of publication bias by including data from unpublished sources (Rothstein, Sutton, & Borenstein, 2005). One significant concern in this endeavor is the search for what is generally labeled as gray literature (Borenstein, Hedges, Higgins, & Rothstein, 2009a; Mahood, 2006; Rothstein & Hopewell, 2009), which comprises work that is not published in research journals covered by these databases or not published at all. Although the Internet provides increasing access to information about published and unpublished studies, the search for, and inclusion of, gray literature remains a challenge and is still unanimously recognized to be particularly difficult (Hopewell, Clarke, & Mallett, 2006).

Thus, in an effort to be as exhaustive as possible, we took a number of steps to identify potential sources of unpublished stud-
ies. First, we identified possible sources of gray literature related to the impact of (action) video games on perception, attention, or cognition, by following the recommendations found in various books, articles and websites dedicated to publication bias and gray literature, such as the Cochrane Handbook and the Library of the University of Western Australia. Given these recommendations we: (a) queried several databases that specialize in gray literature (e.g., PsycEXTRA, ScholarOne,.opengrey, base-search); (b) searched the abstracts of the annual conferences of the Society for Neuroscience, the Vision Science Society, the Cognitive Neuroscience Society, and the annual convention of the American Psychological Association (APA); (c) conducted additional searches in Google Scholar, and in the database Dissertations Abstracts International; (d) posted to a number of relevant list-servers including those for the Vision Sciences Society (VSS) and the Color & Vision Network (CVNet); and (e) directly contacted 67 authors who are known to have worked on the topic and asked whether they had or were aware of unpublished data examining the effects of video games on attention, perception or cognition.

We further extended the literature search to include other languages: Chinese, French, German, Italian, Portuguese, Romanian, Russian, and Spanish. The literature search in non-English languages was performed using the same databases (but with the keywords translated into the corresponding language), as well as additional tools specific to each language, such as tesionline (Italian), wanfang (Chinese), plural (Romanian), and the National Library of Romania to cite a few. Whenever possible, we confined the search to examine only the title, abstract, subject, and keyword fields. Only if a paper’s potential for inclusion was unclear was the full text consulted. We also limited our search to relevant disciplines (e.g., psychology, neuroscience, computer science, education, sociology), and only included written material (e.g., books, chapters, articles, reviews), and thus discarded other types of documents (e.g., video, audio, biography), as well as irrelevant subjects or topics (e.g., academic guidance counseling, music, mental health, robotics, nutrition).

Throughout the search process, we paid particular attention to literature reviews or comments (Achtman, Green, & Bavelier, 2008; Bavelier et al., 2011; Bavelier, Green, Pouget, & Schrater, 2012; Bisoglio et al., 2014; Boot et al., 2011; Boot & Simons, 2012; Boot, Simons, et al., 2013; Granic, Lobel, & Engels, 2014; Green & Bavelier, 2012; Kristjansson, 2013; Latham, Patston, & Tippett, 2013; Oei & Patterson, 2014; Spence & Feng, 2010), even when the focus was not directly relevant to the present meta-analysis (Connolly et al., 2012; Sitzmann, 2011; Tobias et al., 2011; Vogel et al., 2006; Wouters, van Nijmegen, van Oostendorp, & van der Spek, 2013; Young et al., 2012), to ensure all potentially relevant references were examined for inclusion. We further cross-checked with recently published meta-analyses (Powers & Brooks, 2014; Powers et al., 2013; Wang et al., 2017), one of which examined a much larger range of video game related research. Finally, the reference lists of all the documents that passed the inclusion criteria were also consulted.

Our search, combining the results of all databases and languages as well as all keywords combinations described above, yielded a total of 958,147 hits, including 48 records in Chinese, 2,360 in French, 59,360 in German, 11,575 in Italian, 120,012 in Portuguese, 26 in Romanian, and 29,720 in Spanish. As expected, there was substantial redundancy across databases. After removal of duplicates, the remaining 676,102 references were screened by 4 raters (two authors of this article and two graduate assistants), who only read the titles of the articles (as well as the tables of contents in case of books and theses) and were trained to exclude studies that fell outside the scope of the present meta-analysis. Studies were excluded if they were only theoretical, did not involve a group of video game players or video game training, if they included only patients, or if they did not involve a measure of perception, attention or cognition. Most of the studies excluded at Step 1 dealt with topics such as the relationship(s) between video game habits (e.g., frequency/types of games played) and sociodemographic information, measures of psychopathology, or of social/personality factors (in particular measures related to aggression). In addition, we noted a substantial body of literature focused on the development and potential impact of serious video games for educational purposes, as well as a growing literature on the use of video games as therapeutic tools designed for targeted clinical populations. Lastly, a large body of work examining which game characteristics can increase the motivation to play or induce a flow experience was also excluded.

The 5,770 documents that remained after Step 1 were then reduced to 630 documents in Step 2 after a reading of the abstracts. Finally, after careful reading of the methods sections of these sources in Step 3, we further excluded 549 references that did not meet the inclusion criteria detailed in the next section below (which also describes representative examples of studies excluded at this stage and the reason(s) they were excluded). The final dataset thus contained 82 studies, including 65 published studies (all in English) and 17 unpublished studies. The 17 unpublished studies consisted of 6 doctoral theses (five of which were in English and one in Spanish), 3 personal communications, and 8 posters from various conferences. Several of these records contained both cross-sectional and intervention studies. Of the 73 records with cross-sectional designs, 15 were unpublished (20%), whereas of the 23 records for intervention studies only 2 were unpublished (9%). It is worth noting that a substantial proportion of the unpublished work identified in our initial search contained data that were later published (and is utilized in the published form here), a fact that likely lowered the prevalence of unpublished work included in this meta-analysis.

Throughout all steps of the process (from literature search to study selection, effect size computation and coding of moderators and other study descriptors) authors were contacted to obtain missing information. In total, 67 authors were contacted for various reasons (e.g., missing data, clarification of method, etc.) and all of these authors were, at the same time, asked about potential unpublished work. Authors from one paper were sometimes contacted together, resulting in 51 e-mails being sent. We obtained 47 responses of 51 requests (92% response rate), and authors sent the requested data or information in 40 cases (85% success), which led to inclusion in 30 cases (78% inclusion). Studies were excluded if the information provided by the authors indicated that the study did not fit our inclusion or exclusion criteria (10 studies) or if the authors were not able to provide the requested data (7 studies), either because it was not collected or because it was no longer available.

Selection Criteria

The 2nd and 3rd steps of our study selection consisted of reviewing the abstracts and/or methods sections of the 5,770 eligible sources respectively, to verify if the sources satisfied the
inclusion criteria. In short (see below for additional detail), in Step 2, studies were rapidly examined to determine whether they could potentially be relevant to our meta-analysis. Then, in Step 3, the 693 remaining studies were examined more carefully to ensure that the experimental design met all our inclusion/exclusion criteria. Three authors and two research assistants were involved in this process (with the research assistants primarily playing a role in Step 2). In all, each study that was eventually included was thus processed by at least two persons. While most studies were easily categorized as meeting or failing to meet the inclusion requirements, those few studies that were less clear were put aside and discussed during group meetings of at least three of the authors until unanimous consensus was reached.

Step 2 - Quick scan: A study passed through Step 2 if it included a comparison between a group that engaged in action video play and a control group, on a measure of attentional, perceptual, or cognitive skill, or if the performance in one of these domains was measured before and after a video game training period with an action video game trained group being contrasted with a nonaction video game trained group. Studies that measured other cognitive skills, such as risk taking or delay discounting (e.g., Bailey, West, & Kuffel, 2013), and studies contrasting different types of experienced video game players, such as action and role playing in Krishnan, Kang, Sperling, and Srinivasan (2013), were thus excluded at this stage. Although these criteria were often easily verifiable by examining the abstracts, we generally also screened the methods sections to check for whether relevant background measures such as verbal IQ could nevertheless be extracted. For example, in the study by Bailey and colleagues (2013) introduced earlier, the authors mentioned that the participants also completed the useful-field-of-view and stop-signal tasks. Thus, while the risk taking data did not fit any of the cognitive domains investigated here, the UFOV and stop-signal data could still be included; these data were not available in the paper and were obtained from the thesis of Benoit Bediou (Bailey, 2012).

Step 3 - Thorough screening: After careful reading of the full text, a study (or where applicable specific individual experiments within the manuscript under consideration) was included if it satisfied all the five inclusion/exclusion criteria detailed below. Illustrative examples of excluded studies and the reason(s) for exclusion can be found in supplementary Table S1.

1. Effect size measure. The study measured performance in one (or more) of the 8 domains of cognition introduced earlier, in healthy participants, even if this measure was not the primary dependent measure in the study. Studies were included only if we could extract measures of mean and standard deviation or recover other information permitting the calculation of effect size.

2. Game genre and hours of practice criteria for cross-sectional studies. Our focus on action video games means that predominantly only first- and third-person shooter games were included. This definition excluded not only games that clearly belong to other genres, such as real-time strategy games (e.g., Starcraft, Rise of Nations), puzzle games (e.g., Tetris, Portal), or role-playing games (e.g., World of Warcraft), but also games that may have been classified as action games by some authors, such as fighting games (Tanaka et al., 2013), arcade games like Pac-Man, or the cognitive research interface Space Fortress (Wang et al., 2017). Importantly, this definition also excluded all studies conducted before the year 2000, because at that time the existing power and graphics limitations did not allow for the fast and complex dynamics characteristics of action games described earlier.

Studies comparing habitual action video game players to nonaction video game players were included if the participants in the AVGP group played at least 3 hours per week of action video games and had done so for the last 6 months, and participants in the NVGP group spent less than 1 hour per week playing specifically action video games, or played fewer than 3 hours per week of video games in general, across all genres. These criteria were used as they capture most of the studies in the field. Some laboratories, however, have systematically used more stringent criteria, like the Bavelier laboratory which has always required at least 5 hours per week of action video game for the AVGP group and no more than 1 hour per week of play in other genres for NVGP.

Studies were thus excluded either because the AVGP group did not play a minimum of 3 hours per week of action games exclusively, or if they did play more than 3 hours of video games, but where this total may have included games that did not belong to the action genre, such as when the games listed in the manuscript included role playing, strategy, sports, or fighting games (Adams, 2013, Experiment 2; Bialystok, 2006; Granck, Gorbet, & Sergio, 2010; Vallat, Lamb, & Annetta, 2013). In addition, studies in which the NVGPs played more than 1 hour of action games were also excluded (e.g., NVGPs played less than 2 hours of action games and less than 5 hours overall in Dobrowolski, Hanusz, Sobczyk, Skorko, & Wiatrow, 2015; NVGPs played less than 4 hours in Durlach, Kring, & Bowens, 2009). For Rupp, McConnell, and Smith (2016), only the groups of high-play gamers and nonvideo game players met our criteria for inclusion. The group of medium-play gamers was excluded because they played an average of 3 hours of action video games per week suggesting some individuals in this group may have played for fewer than 3 hours.

In another recent study (Unsworth et al., 2015), the authors correlated performance in a number of tasks with hours of action video game play, such that their sample included some AVGPs and NVGPs, but the majority of the individuals were intermediate players. Their data could therefore not be included as published given that our inclusion criteria focused on extreme groups and required some separation between the AVGP and NVGP group in terms of weekly gaming hours. However, thanks to the authors sharing their raw data with us, we were able to compute the effect sizes for the comparison of AVGPs and NVGPs that met our inclusion criteria.

Studies involving clinical populations were systematically excluded (R. W. Li et al., 2011), as were studies comparing action video game players to particular populations such as musicians or bilinguals (Bergstrom, Howard, & Howard, 2012; Bialystok, 2006). Another study was excluded because we could not obtain the data in format corresponding to our action video game players selection criteria (Collins & Freeman, 2014).

3. Game genre for experimental group in intervention studies. For studies involving a training intervention, the experimental game had to be an action game as defined above, that is a first or third person shooter game. There was no restriction regarding the type of platform (i.e., console, computer/laptop, or mobile device).

4. Game genre for control group in intervention studies. Training with an action game had to be compared with training on another, nonaction video game. Studies comparing performance before and after training with no control group or with just a
test-retest control group were excluded. As is standard in the field, all intervention studies entailed comparisons between a commercially available action and a commercially available nonaction video game. This is typical of studies in that domain as it ensures both experimental and control interventions used media known to be engaging and challenging to their users (Jacoby & Ahissar, 2013). Studies in which control training consisted exclusively in repeated practice of a version of the cognitive task used to measure performance changes, such as the UFOV trained group in Belchior et al. (2013), were thus excluded. In addition, two studies compared an action game to more than one nonaction video game (Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Oei & Patterson, 2013), in which case all the comparisons between the action and the variety of nonaction games were included. Finally, only studies in which the experimental and control training had matched schedules were included; for example, Boot, Champion, et al. (2013) was excluded because the amount of training was highly disparate across training groups.

5. Hours of practice criteria and schedule of testing for intervention studies. Only studies that included a pre- and a posttest were included. Given the focus of the field on long lasting impact of action video game on cognition, a minimum delay of 24 hours was required between the last video game training session and the first posttraining measure of performance. Such a criterion is typical of most studies and rules out the myriad effects that could arise from having just played a video game (e.g., as related to physiological arousal). Studies that used either short training or immediate posttests were thus excluded (e.g., Gallagher & Preswicht, 2012; Nelson & Strachan, 2009; Obana & Kozhevnikov, 2012; Rehbein, Kleimann, & Mossle, 2007; Sanchez, 2012), as well as studies that used only posttest measures (Grarne et al., 2010).

Durable changing cognitive skills requires repeated, distributed training; video games are no exception to this (Stafford & Dewar, 2014). Although most studies have used 10 hours or more of training distributed over two weeks, this criterion was relaxed to a minimal duration of 8 hours of training, distributed over a period of at least 8 days.

These combined criteria (delay of 24 hours for posttest and 8 hours of training distributed over 8 days) excluded a few studies: the ones using training on the scale of just a few minutes (Rehbein et al., 2007) to a few hours (e.g., 4 hours in Cherney, 2008; 4 hours on 4 different games in Sersale, 2005) as well as studies where training was highly massed across time (e.g., Gallagher & Preswicht, 2012 did 12 hours in 2 days; van Ravenzwaaij, Boekel, Forstmann, Ratcliff, & Wagemakers, 2014 did 10 hours in 5 days in Experiment 1, and 20 hours in 5 days in Experiment 2), to avoid confounding effects related to the known inefficiency of massed practice (Baddeley & Longman, 1978; Benjamin & Tullis, 2010; Stafford & Dewar, 2014; Vlach & Sandhofer, 2012).

Among the 82 records included, 32 (39%) included more than one study. Among these, nine records included exactly identical participants, 22 involved independent or partially independent groups of participants, and one included both. Forty-three (37%) of these studies measured performance in more than one domain of cognition. For our analyses, we used a hierarchical model in which the effect sizes are combined at the level of the 116 studies with independent or partially independent subjects over a total of 309 effects. Clusters are therefore defined as effect sizes measures being nested in studies, which allows both within-study and between-study sources of variances to be considered in the models. The hierarchical structure of effect sizes and moderators used in our analysis is described in further detail in the section on effect size computation.

Coding of Moderators and Other Study Descriptors

For each analysis (cross-sectional and intervention), we provide information relevant to the overall effect of action games on behavior collapsed across cognitive domains and all other moderators, as well as a moderator analysis that takes advantage of the methodological heterogeneity, or between-study variability, to provide a more detailed and finer-grained view of the factors modulating the main effect of action video game. To increase the consistency in how the different studies were coded, an initial coding of moderators and study descriptors was performed by Benoit Bediou, who thus had a more complete picture of the data in terms of quality and heterogeneity. Once all studies were coded, the entire dataset—that is all effect sizes, and their classification into cognitive domains, types of studies, type of effect, age group, and so forth—was cross-checked by a minimum of two other authors. We list below for each effect size the moderators coded and included in the analysis (see Table 1 for the number of effect sizes for each moderator level in cross-sectional studies, and Table 2 for intervention studies).

Cognitive Domain

We categorized all the computed effect sizes into eight cognitive domains, according to the task and conditions used to extract the relevant effect: (1) perception, (2) bottom-up attention, (3) top-down attention, (4) spatial cognition, (5) multitasking, (6) inhibition, (7) problem solving, and (8) verbal cognition. The number of effect sizes in each cognitive domain for correlational and intervention studies can be found respectively in Table 1 and Table 2.

The initial selection of relevant task conditions and their classification into cognitive domains described earlier, as well as further coding of the additional moderators considered in the analysis, was first performed by three authors, and then discussed with the rest of the authors until unanimous agreement was reached. For example, the full version of the UFOV task generally contains several conditions that are run in separate blocks, and that probe different skills. The single-task conditions (center task or peripheral task alone) were assigned to the ‘perceptual’ skill; the dual-task condition in the absence of distractors was assigned to the ‘task-switching/multitasking’ skill; and the dual-task condition in the presence of distractors was assigned to the ‘top-down’ attention’ skill. We recognize that there is a certain degree of arbitrariness in such classifications as clearly the latter condition also qualifies for the ‘multitasking’ skill; yet this task condition is typically utilized for the load it puts on top-down attention. In our classification, we aimed at preserving the intent of the measure in the original work. Each effect size was assigned to one and only one cognitive domain based on the main domain that task component is hypothesized to tap. The agreement between the three raters involved in the coding of this moderator was good (Cohen’s $\kappa = .68$; Fleiss $\kappa = .70$ and Krippendorff $\alpha = .70$; Cohen, 1960; Davies & Fleiss, 1982; Fleiss, 1971; Krippendorff, 1980).

Age group. We defined three discrete age ranges, capturing the most important differences in cognitive functioning across the
Table 1  
Cross-Sectional Meta-Analysis

<table>
<thead>
<tr>
<th>Moderator-Level</th>
<th>k</th>
<th>m</th>
<th>F</th>
<th>g</th>
<th>95% CI</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception</td>
<td>30</td>
<td>22</td>
<td>.775</td>
<td>[.564, .985]</td>
<td>16.6</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Top-down attention</td>
<td>71</td>
<td>48</td>
<td>.625</td>
<td>[.494, .756]</td>
<td>27.4</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Spatial cognition</td>
<td>27</td>
<td>19</td>
<td>.750</td>
<td>[.526, .975]</td>
<td>14.5</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Inhibition</td>
<td>11</td>
<td>9</td>
<td>.310</td>
<td>[.065, .556]</td>
<td>7.2</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Multi-tasking</td>
<td>22</td>
<td>17</td>
<td>.549</td>
<td>[.277, .821]</td>
<td>11.9</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Problem solving</td>
<td>7</td>
<td>4</td>
<td>.501</td>
<td>[−.017, 1.019]</td>
<td>2.4</td>
<td>.054</td>
<td></td>
</tr>
<tr>
<td>Verbal cognition</td>
<td>26</td>
<td>16</td>
<td>.297</td>
<td>[.032, .563]</td>
<td>7.7</td>
<td>.033*</td>
<td></td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>5</td>
<td>3</td>
<td>.324</td>
<td>[−.337, .985]</td>
<td>2.9</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>Younger adults</td>
<td>189</td>
<td>86</td>
<td>.598</td>
<td>[.498, .697]</td>
<td>33.6</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>DV type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>139</td>
<td>61</td>
<td>.582</td>
<td>[.467, .697]</td>
<td>31.2</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>55</td>
<td>41</td>
<td>.614</td>
<td>[.467, .76]</td>
<td>31.5</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Effect type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main</td>
<td>139</td>
<td>65</td>
<td>.620</td>
<td>[.515, .724]</td>
<td>33.4</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>55</td>
<td>35</td>
<td>.518</td>
<td>[.353, .683]</td>
<td>17.3</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bavelier</td>
<td>54</td>
<td>32</td>
<td>.800</td>
<td>[.551, 1.05]</td>
<td>19.5</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>140</td>
<td>57</td>
<td>.510</td>
<td>[.404, .616]</td>
<td>29.1</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Recruitment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overt</td>
<td>140</td>
<td>74</td>
<td>.624</td>
<td>[.504, .745]</td>
<td>33</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Covert</td>
<td>54</td>
<td>16</td>
<td>.504</td>
<td>[.329, .679]</td>
<td>7.2</td>
<td>&lt;.001***</td>
<td></td>
</tr>
</tbody>
</table>

Note: The effect sizes of each moderator level are shown; these analyses are based on robust variance estimates with a model including all moderators. k = number of effect sizes; m = number of clusters (F tests) or individual studies (t tests) for each moderator and each level, respectively. F = AHT-F test comparing the levels of a given moderator. g = effect size estimate (Hedges); 95% CI = 95% confidence interval; df = degrees of freedom; p = p value of AHT-F tests for moderator effects and t tests comparing each level against zero. For each moderator, the first row (in bold and italics) shows the result of the F tests (AHT Type) examining possible differences between the levels of each moderator.

Effect Size Computation

Effect sizes give the magnitude and the direction either of the difference between two groups (i.e., AVGPs vs. NVGPs) on a given measure, or of the difference in the magnitude of pretest to posttest changes between two treatments (i.e., training on an action vs. on a nonaction control game). Positive effect sizes reflect greater performance (or improvement) in AVGPs (or action-trained) compared with NVGPs (or control-trained) groups (i.e., more correct responses, fewer errors, or faster RTs).

We used the bias-corrected Hedges’ g as our main measure of effect sizes. This is equivalent to a Cohen’s d with an additional correction factor for small samples, and is thus more conservative.
than the classic Cohen’s $d$ measure (Rosenthal, Rosnow, & Rubin, 2000). For cross-sectional studies, the denominator was the within-group pooled standard deviation across the performance in AVGPs and in NVGPs groups. For intervention studies, the denominator was the pooled standard deviation across the posttest minus pretest difference in the action video game trained group and in the nonaction, control video game trained group. Note that when means and standard deviations of the posttest minus pretest difference were not available from the paper or from the authors, statistics related to the effect(s) of interest (means and standard deviations or standard errors, or statistics reflecting these differences, such as $F$ or $t$ tests, chi-square tests, or correlation coefficients depending on what was available) were converted into Hedges’ $g$ using the procedure described in Borenstein, Hedges, Higgins, and Rothstein (2009b). For example, if for a given task, the statistics were reported for both speed and accuracy without specifying one as being more relevant than the other, then we kept the effect size with the greatest absolute value, independent of its significance or direction. For difference effects, such as switch costs, we followed the same reasoning, with positive effects corresponding to greater performance or improvement in AVGPs versus NVGPs (e.g., smaller switch costs). Although for most effects the direction of an effect (positive or negative) was unarguable, a difficulty arose in coding the direction of flanker compatibility effects. Such effects (subtracting RT on trials with response compatible distractors from trials with response incompatible distractors) have been interpreted in two opposing ways in the literature. In the Attentional Network Test (ANT) and other similar Eriksen flanker/Posner cueing tasks, greater compatibility effects have been interpreted as reflecting a lack of attentional control (Rueda et al., 2004). The load theory of Lavie and colleagues (2004) conversely hypothesizes that larger compatibility effects are the direct result of greater attentional resources. This is based upon the fact that larger compatibility effects are typically found in conditions of low load (i.e., situations when more attentional resources are available). Thus, in this view, greater compatibility effects are considered a positive outcome. Given these opposing interpretations of the flanker compatibility effect (see Dye & Bavelier, 2010 for an in-depth discussion), we chose to follow the interpretation provided by the authors of the given

For each study that survived our inclusion/exclusion criteria, we first identified within each task the condition(s) reflecting the impact of action video games on one of the eight cognitive domains above. Because different conditions of the same task may measure different cognitive domains (e.g., the Useful Field of View task has a component related primarily to visual selective attention as well as a component related to multitasking ability), more than one effect size could initially be extracted per task. Effects from the same groups of participants were attributed the same cluster, even if reflecting different cognitive domains.

If for a given task, the statistics were reported for both speed and accuracy without specifying one as being more relevant than the other, then we kept the effect size with the greatest absolute value, independent of its significance or direction. For difference effects, such as switch costs, we followed the same reasoning, with positive effects corresponding to greater performance or improvement in AVGPs versus NVGPs (e.g., smaller switch costs). Although for most effects the direction of an effect (positive or negative) was unarguable, a difficulty arose in coding the direction of flanker compatibility effects. Such effects (subtracting RT on trials with response compatible distractors from trials with response incompatible distractors) have been interpreted in two opposing ways in the literature. In the Attentional Network Test (ANT) and other similar Eriksen flanker/Posner cueing tasks, greater compatibility effects have been interpreted as reflecting a lack of attentional control (Rueda et al., 2004). The load theory of Lavie and colleagues (2004) conversely hypothesizes that larger compatibility effects are the direct result of greater attentional resources. This is based upon the fact that larger compatibility effects are typically found in conditions of low load (i.e., situations when more attentional resources are available). Thus, in this view, greater compatibility effects are considered a positive outcome. Given these opposing interpretations of the flanker compatibility effect (see Dye & Bavelier, 2010 for an in-depth discussion), we chose to follow the interpretation provided by the authors of the given

### Table 2

Meta-Analysis of Intervention Studies

<table>
<thead>
<tr>
<th>Moderator–Level</th>
<th>k</th>
<th>m</th>
<th>$F$</th>
<th>$g$</th>
<th>95% CI</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive domain</td>
<td>90</td>
<td>22</td>
<td>.358</td>
<td>3.3</td>
<td>.827</td>
<td>4.0</td>
<td>0.04</td>
</tr>
<tr>
<td>Perception</td>
<td>10</td>
<td>8</td>
<td>.227</td>
<td>[-.112, .565]</td>
<td>5.7</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Top-down attention</td>
<td>35</td>
<td>13</td>
<td>.309</td>
<td>[.142, .477]</td>
<td>5.4</td>
<td>.005**</td>
<td></td>
</tr>
<tr>
<td>Spatial cognition</td>
<td>29</td>
<td>9</td>
<td>.448</td>
<td>[.122, .774]</td>
<td>4.9</td>
<td>.017*</td>
<td></td>
</tr>
<tr>
<td>Multi-tasking</td>
<td>7</td>
<td>5</td>
<td>.291</td>
<td>[-.567, 1.149]</td>
<td>3.9</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Verbal cognition</td>
<td>9</td>
<td>5</td>
<td>.532</td>
<td>[-.235, 1.298]</td>
<td>3.3</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>DV type</td>
<td>90</td>
<td>22</td>
<td>5.388</td>
<td>2.3</td>
<td>.128</td>
<td>3.0</td>
<td>.784</td>
</tr>
<tr>
<td>Accuracy</td>
<td>81</td>
<td>19</td>
<td>.347</td>
<td>[.164, .53]</td>
<td>6.6</td>
<td>.003**</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>9</td>
<td>6</td>
<td>.536</td>
<td>[.068, 1.004]</td>
<td>2.2</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Effect type</td>
<td>90</td>
<td>22</td>
<td>.900</td>
<td>5.3</td>
<td>.006**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main</td>
<td>81</td>
<td>18</td>
<td>.358</td>
<td>[.141, .576]</td>
<td>4.9</td>
<td>.008**</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>9</td>
<td>5</td>
<td>.431</td>
<td>[-.356, 1.218]</td>
<td>2.5</td>
<td>.165</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>90</td>
<td>22</td>
<td>19.992</td>
<td>5.3</td>
<td>.006**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bavelier</td>
<td>18</td>
<td>11</td>
<td>1.033</td>
<td>[.656, 1.41]</td>
<td>4.8</td>
<td>.001***</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>72</td>
<td>11</td>
<td>.199</td>
<td>[.073, .47]</td>
<td>3.8</td>
<td>.11</td>
<td></td>
</tr>
</tbody>
</table>

Note. The effect sizes of each moderator level are shown below. Analyses are based on robust variance estimates with a model including all moderators. Studies involving older adults were excluded ($k = 11, m = 2$). $k =$ number of effect sizes; $m =$ number of clusters ($F$ tests) or individual studies ($t$ tests) for each moderator and each level, respectively. $F =$ AHT-F test comparing the levels of a given moderator; $g =$ effect size estimate (Hedges); 95% CI = 95% confidence interval; df = degrees of freedom; $p =$ value of AHT-F tests for moderator effects and $t$ test comparing each level against zero. Significant effects with a degree of freedom (df) below 4 are likely to be underpowered and should not be trusted. For each moderator, the first row (in bold and italics) shows the result of the $F$ tests (AHT Type) Examining Possible Differences between the Levels of Each moderator.

### Table 2

Meta-Analysis of Intervention Studies

<table>
<thead>
<tr>
<th>Moderator–Level</th>
<th>k</th>
<th>m</th>
<th>$F$</th>
<th>$g$</th>
<th>95% CI</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive domain</td>
<td>90</td>
<td>22</td>
<td>.358</td>
<td>3.3</td>
<td>.827</td>
<td>4.0</td>
<td>0.04</td>
</tr>
<tr>
<td>Perception</td>
<td>10</td>
<td>8</td>
<td>.227</td>
<td>[-.112, .565]</td>
<td>5.7</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Top-down attention</td>
<td>35</td>
<td>13</td>
<td>.309</td>
<td>[.142, .477]</td>
<td>5.4</td>
<td>.005**</td>
<td></td>
</tr>
<tr>
<td>Spatial cognition</td>
<td>29</td>
<td>9</td>
<td>.448</td>
<td>[.122, .774]</td>
<td>4.9</td>
<td>.017*</td>
<td></td>
</tr>
<tr>
<td>Multi-tasking</td>
<td>7</td>
<td>5</td>
<td>.291</td>
<td>[-.567, 1.149]</td>
<td>3.9</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Verbal cognition</td>
<td>9</td>
<td>5</td>
<td>.532</td>
<td>[-.235, 1.298]</td>
<td>3.3</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>DV type</td>
<td>90</td>
<td>22</td>
<td>5.388</td>
<td>2.3</td>
<td>.128</td>
<td>3.0</td>
<td>.784</td>
</tr>
<tr>
<td>Accuracy</td>
<td>81</td>
<td>19</td>
<td>.347</td>
<td>[.164, .53]</td>
<td>6.6</td>
<td>.003**</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>9</td>
<td>6</td>
<td>.536</td>
<td>[.068, 1.004]</td>
<td>2.2</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Effect type</td>
<td>90</td>
<td>22</td>
<td>.900</td>
<td>5.3</td>
<td>.006**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main</td>
<td>81</td>
<td>18</td>
<td>.358</td>
<td>[.141, .576]</td>
<td>4.9</td>
<td>.008**</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>9</td>
<td>5</td>
<td>.431</td>
<td>[-.356, 1.218]</td>
<td>2.5</td>
<td>.165</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>90</td>
<td>22</td>
<td>19.992</td>
<td>5.3</td>
<td>.006**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bavelier</td>
<td>18</td>
<td>11</td>
<td>1.033</td>
<td>[.656, 1.41]</td>
<td>4.8</td>
<td>.001***</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>72</td>
<td>11</td>
<td>.199</td>
<td>[.073, .47]</td>
<td>3.8</td>
<td>.11</td>
<td></td>
</tr>
</tbody>
</table>

Note. The effect sizes of each moderator level are shown below. Analyses are based on robust variance estimates with a model including all moderators. Studies involving older adults were excluded ($k = 11, m = 2$). $k =$ number of effect sizes; $m =$ number of clusters ($F$ tests) or individual studies ($t$ tests) for each moderator and each level, respectively. $F =$ AHT-F test comparing the levels of a given moderator; $g =$ effect size estimate (Hedges); 95% CI = 95% confidence interval; df = degrees of freedom; $p =$ value of AHT-F tests for moderator effects and $t$ test comparing each level against zero. Significant effects with a degree of freedom (df) below 4 are likely to be underpowered and should not be trusted. For each moderator, the first row (in bold and italics) shows the result of the $F$ tests (AHT Type) Examining Possible Differences between the Levels of Each moderator.
papers (i.e., if the authors considered the direction of difference/change to be positive, it was coded as positive here and vice versa). Note that we also conducted additional analyses in which larger compatibility effects were always coded as either positive or negative and found that this choice did not influence the results (see supplementary materials, section 4.1, for separate analyses coding greater interference as better top-down attention vs. worse top-down attention).

To minimize the likelihood of errors, three steps were taken. First, two authors were involved in the 309 effect size computations. Second, whenever possible, effect sizes were computed using more than one method (e.g., F values and degrees of freedom vs. means and standard deviations) and then compared, to ensure cross validity of the different conversions into effect sizes. In cases where the computations were not identical, to be conservative, the smaller of the two values was utilized. Importantly, these differences never exceeded 0.1 in effect size magnitude. Finally, to assess intercoder reliability, 15 randomly chosen effect sizes (corresponding to 5% of the total effect sizes), were checked by both authors. Agreement was achieved in 100% of the cases suggesting that intercoder reliability was likely high. Conversions between effect sizes, were all performed using Comprehensive Meta-Analysis (CMA) software. All analyses were then conducted using the R packages metafor (Viechtbauer, 2010) and robumeta (Fisher, Tipton, & Hou, 2016).

Statistical Analysis

In all, our search yielded 309 effect sizes nested in 116 independent studies, with 198 effect sizes extracted from cross-sectional studies, and 111 effect sizes from intervention studies. Note that the final cross-sectional meta-analysis included 194 effect sizes nested in 89 studies drawn from 73 records, as moderator levels with less than 4 effect sizes were not considered (see below). For intervention studies after applying this same criterion, the meta-analysis including both old and young adults comprised 101 effect sizes nested in 24 studies drawn from 23 records. The main intervention meta-analysis focusing on only young adults included 90 effect sizes nested in 22 studies drawn from 21 records.

To account for the hierarchical structure of our data set, robust variance estimation (Hedges, Tipton, & Johnson, 2010) with hierarchical weights and small sample corrections (Tipton, 2015; Tipton & Pustejovsky, 2015) were used throughout. In these analyses, we clustered at the level of the studies. First, we computed the main effect of action video game play across all moderators separately for cross-sectional and intervention studies. This was systematically followed by a multiple moderator model within each. We note though that to run such a model, moderator levels with 4 effect sizes or less had to be excluded from the analyses as the small sample correction in such multiple moderator models is problematic when the model contains empty cells. This resulted in excluding one cognitive domain for the meta-analysis of cross-sectional studies (bottom-up attention) and three cognitive domains for that of intervention studies (bottom-up attention, inhibition, problem solving). None of the other moderators contained cells with 4 or fewer effect sizes (see k values in result tables). The presence of moderator effects were tested by means of small-sample adjusted F tests or t tests, using the Wald test function from the R package clubSandwich (Pustejovsky, 2016) or the t tests provided in the robumeta package (Fisher et al., 2016), which implements the small sample corrections developed in Tipton (2015) and Tipton and Pustejovsky (2015). An important feature of these small sample corrections is that the degrees of freedom depend not only on the number of studies, but also features of the covariates. In the tables, this results in different degrees of freedom (and thus power) for testing different moderator variables. Importantly, when the degrees of freedom for t tests are smaller than 4, the stated p value can be too small; for this reason, Tanner-Smith, Tipton, and Polanin (2016) suggest using a lower threshold for significance in these cases (e.g., p < .01 instead of p < .05), a strategy which we adopt throughout. For the interested reader, we provide comparison of the models using hierarchical, as we did in the present paper, versus correlated versus fixed (WLS) weights in Tables S2 and S3 for respectively the cross-sectional and intervention meta-analysis (see supplementary information, Section 4.2).

We also investigated possible concerns with publication bias. One reason for publication bias would be from selective reporting—the file drawer problem. To investigate this possibility, we aggregated the effect sizes to the study level and used mixed effects models and associated methods, including contour enhanced funnel plots (Peters, Sutton, Jones, Abrams, & Rushton, 2008), and Egger’s tests and trim-and-fill analyses (Duval & Tweedie, 2000). Although mixed effects models are less than ideal given the hierarchical structure of our data set, we note that there is not yet an accepted method in the field to correct for publication bias when using robust variance estimation models. Egger’s test and trim-and-fill aggregating effect sizes to the study level appears thus the most appropriate (Tanner-Smith, 2012). A second reason for publication bias could occur is if effect sizes estimated in smaller studies are larger than those estimated in larger studies. This correlation between the effect sizes and standard errors could occur for various reasons. To address this concern, we also present a new method known as the PET-PEESE - for Precision Effect Test (Sterne & Egger, 2005) and Precision-Effect Estimate with Standard Error (Stanley & Doucouliagos, 2014). For this method, we used the RVE approach. Here the focus is on the intercept, which provides an estimate of a study in which the standard error or variance is 0. If this corrected effect differs from the overall estimated effect, then it suggests that the overall effect is biased; as can be seen in Table S4 and S5, this method appears to have some serious limitations. To complement these analyses we also consider possible differences in an additional moderator ‘Publication Type’ coding published and unpublished studies separately, and extracting overall effect sizes separately for each type of study, for both cross-sectional and intervention studies.

Finally, our moderator analyses looked at the influence of each moderator while controlling for all the others. Such a multiple moderator approach documents the sources, if any, of heterogeneity in the data set, and thus provides key information as to how the main effect should be understood in the context of all existing moderators. For the analysis of cross-sectional studies, the multiple moderator model included cognitive domains (7 levels), age (children, younger adults), dependent measures (speed, accuracy), type of effect (main, difference), laboratory of origin (Bavelier, others) as well as the recruitment method (covert, overt). For intervention studies, the main analysis focused on younger adults and thus the multiple moderator model included cognitive domains.
(5 levels), dependent measures (speed, accuracy), type of effect (main, difference), laboratory of origin (Bavelier, others) as well as the single continuous moderator, training duration.

For graphical illustration, we used the moving constant technique developed by Johnson and Huedo-Medina (2011). The regression line and confidence interval are obtained by estimating the intercept of a model with training duration while holding other moderators constant at their means, after subtracting a moving constant to the current value of training duration. In other words, such figure plots estimated values across observed training durations from 8 to 50 hours, holding other moderator dimensions constant. This is essentially the equation of the best-fitting line, adding 95% CIs using RVE (see Figure 4 for additional comments).

**Results**

For each RVE analysis, we report the number of effect sizes (k), the number of clusters (m), and the degrees of freedom (df).

**Effect of Action Video Games in Cross-Sectional Studies**

The question addressed by cross-sectional studies is whether people who are habitual action video game players display different cognitive skills than nonaction video game players. The analyses included 7 cognitive domains as bottom-up attention had to be excluded based on too few effect sizes (see methods section): perception, top-down attention, spatial cognition, inhibition, task-switching/multitasking, problem solving and verbal cognition. An overall mean weighted effect size in the medium range (g = 0.55, 95% CI [0.42, 0.68], k = 194, m = 89, df = 24.6, p < .0001) was observed with superior performance being seen in action video game players as compared with nongamers.

**Moderator Analyses**

The data structure included six main moderators: Cognitive Domains (7 levels—as described above), Age (2 levels—children, young adults), Type of Dependent Variable (2 levels—RT, accuracy), Type of Effect (2 levels—main, difference), Laboratory of Origin (2 levels—Bavelier, Others), Recruitment (2 levels—Overt, Covert). Moderator patterns were analyzed using a model including all moderators to best handle the different sources of heterogeneity given the nested and complex structure of our dataset.

We first report the adjusted F tests that qualify for each moderator whether there exist differences among its levels (see Table 1, F tests in Bold). As can be seen, significant differences are found only for Laboratory of Origin, indicating statistically stronger effect of action video game play in the Bavelier laboratory studies as compared with studies carried by other groups. We then review the effect at each moderator level as revealed by the robust variance multiple moderator model.

**Cognitive domains moderator.** The impact of action video games varied across cognitive domains (as shown in Figure 1A and Table 1). The one cognitive domain where action video game play shows no significant effect given the current data set and analyses is problem solving. However, the low degrees of freedom in this analysis suggests that there are not yet enough studies to draw a firm conclusion. Small-to-medium effect sizes are found for inhibition (g = 0.31, 95% CI [0.07, 0.56], df = 7.2, p < .02) and verbal cognition (g = 0.30, 95% CI [0.03, 0.56], df = 7.7 p < .033); these differences are promising, and call for further studies in these domains. In contrast, large effects sizes attributable to habitual action video game play are found for perception (g = 0.78, 95% CI [0.56, 0.99], df = 16.6, p < .0001), top-down attention (g = 0.63, 95% CI [0.49, 0.76], df = 27.4, p < .0001), and spatial cognition (g = 0.75, 95% CI [0.53, 0.98], df = 14.5, p < .0001); the effect size for multitasking/task-switching is slightly smaller but in the upper medium range (g = 0.55, 95% CI [0.28, 0.82], df = 11.9, p < .001).

**Other categorical moderators.** The age moderator indicates a robust effect in young adults, but only a trend in children that should be interpreted with caution and may in part reflect the low number of effect sizes for children (k = 5). For all other moderators, we find significant effects indicating that the effect of habitual action video game play is found in both accuracy and speed measures, main effects and difference effects, the Bavelier laboratory or other laboratories as well as in both studies using overt and covert recruitment. These results are illustrated in Figure 1B.

**Publication bias.** The potential threat of publication bias is always a concern in meta-analysis, including both bias from the ‘file drawer’ problem and from small samples. Contour-enhanced funnel plots were generated (Peters, Sutton, Jones, Abrams, & Rushton, 2008); we then used Egger’s regression test to assess funnel plot asymmetry, as well as the Duval and Tweedie (2000) trim-and-fill analysis to correct for any observed asymmetry and to estimate a corrected effect size (these analyses were done at the level of numbers of clusters/studies m = 89). Egger’s test confirmed significant asymmetry (β = 2.37, SE = 0.29, Z = 8.09, p < .001). Trim-and-fill analysis imputed 29 additional studies with negative effect sizes of low precision to make the plot more symmetrical. Visual inspection of the imputed studies in Figure 2 indicates that most missing studies fall into the category of non-significant effects, in accordance with a publication bias issue. Yet, even with these additional studies, the main effect of habitual action video game play remains sizable with trim-and-fill correction (g = 0.48, 95% CI [0.36, 0.60], m = 118, p < .001) despite being 28% smaller than the Random/Mixed effect (g = 0.67, 95% CI [0.57, 0.76], m = 89, p < .001).

Additional tests of funnel plot asymmetry, and corrections including Egger’s test (Egger, Smith, Schneider, & Minder, 1997) and trim-and-fill analyses, along with the PET/PEESE were performed for each subpopulation of effects (i.e., separately for all moderators and levels), and are presented in Table S4. These subanalyses do indicate that some effects may be more susceptible to selective reporting bias than others. Indeed, while the trim-and-fill analysis suggest a 28% reduction of the overall aggregated effect of action video game in this cross-sectional meta-analysis, effect sizes by subgroup were reduced by a rather wide range between 0% (for a few levels of the ‘cognitive domain’ moderator as well as the children level of the ‘age’ moderator) to 34% (for the accuracy level of the ‘DV Type’ moderator).

Finally, a multiple moderator RVE model indicated no main effect of publication type (F = 2.53, k = 194, m = 89, df = 21.7, p = .13). For completeness, we report separately the effect size for published (g = 0.62, 95% CI [0.51, 0.73], df = 31, p < .001) and unpublished records (g = 0.43, 95% CI [0.22, 0.64], df = 16.1,
A key question for the field of cognitive training is whether the effect observed in habitual action video game players, just described above, can be induced through training. The analysis of intervention studies addresses the question of whether there is a causal impact of playing action video games (vs. playing other commercially available video games) on cognition. Among the 8 possible cognitive domains, 3 had 4 or fewer effect sizes, and were thus not included in the analyses leaving the 5 following cognitive domains: perception, top-down attention, spatial cognition, task-switching/multitasking, and verbal cognition.

**Effect of Action Video Games in Intervention Studies**

A key question for the field of cognitive training is whether the effect observed in habitual action video game players, just described above, can be induced through training. The analysis of intervention studies addresses the question of whether there is a causal impact of playing action video games (vs. playing other commercially available video games) on cognition. Among the 8 possible cognitive domains, 3 had 4 or fewer effect sizes, and were thus not included in the analyses leaving the 5 following cognitive domains: perception, top-down attention, spatial cognition, task-switching/multitasking, and verbal cognition.

![Figure 1](image.jpg)

*Figure 1.* Cross-sectional meta-analysis. Illustration of the effect sizes of action video game play in cross-sectional studies for each level of our 6 moderators (see Table 1 for full statistics). (A) Cognitive domain moderator; (B) Other 5 moderators.

$p < .001$), which are both positive and significant; we also note that the outcome of this new RVE model (Table S6) closely parallels that of the RVE presented above (see Table 1).

**Intervention studies including young and older adults.** The first analysis we carried out included intervention studies in both young and older adults ($k = 101, m = 24$). This analysis incorporated 6 different moderators, one of which was age. This analysis confirmed a positive effect of action video game play on cognition in young adults ($g = 0.40, 95\% \text{ CI } [0.21, 0.50], df = 7.0; p < .001$), but, if anything, a negative trend in older adults ($g = -0.36, 95\% \text{ CI } [-1.16, 0.43], df = 1.7; p = .16$). Although the difference between these two groups could not be confirmed, as indicated by overlapping CI and an AHT-Type F test with too few degrees of freedom to warrant any conclusions ($F = 20.0$,
Moderator analyses. The data structure included five main moderators: Cognitive Domains (5 levels—see Table 2), Type of Dependent Variable (2 levels—RT, accuracy), Type of Effect (2 levels—main, difference), Laboratory of Origin (2 levels—Bavelier, Others), and Training Duration (continuous variable bounded between 8 and 50 hours in the case of our data set). Moderator patterns were analyzed in a model including all moderators to best handle the different sources of heterogeneity given the complex structure of our dataset.

We first report the F tests that qualify for each moderator whether there exist differences among its levels (see Table 2). As can be seen, no significant difference is found across moderator levels except for the laboratory of origin moderator, indicating larger effects from Bavelier laboratory studies as compared with studies carried by other groups. For completeness, we qualify below the effect of each moderator level separately as revealed by the robust variance multiple moderator model. As a note of caution, we note that there are rather few clusters (studies) that test particular outcomes, making many tests of these moderators statistically underpowered (i.e., they have low degrees of freedom [df]).

Cognitive domains moderator. Although the impact of action video games interventions did not differ across the five cognitive domains considered, when looking at each domain separately, a similar pattern as in cross-sectional studies was observed (Figure 3A). The most promising impact of action game play were on the domains of spatial cognition (g = 0.45, 95% CI [0.12, 0.77], df = 4.9, p < .017) and top-down attention (g = 0.31, 95% CI [0.14, 0.48], df = 5.4, p < .005), followed by the domain of perception (g = 0.23, 95% CI [−0.11, 0.57], df = 5.7, p = .15). The impact of action video game play on the last two domains considered, multitasking (g = 0.29, 95% CI [−0.57, 1.15], df = 3.9, p = .40) and verbal cognition (g = 0.53, 95% CI [−0.24, 1.30], df = 3.3, p = .12) was not significant.

Other categorical moderators. The analyses of other moderators is illustrated in Figure 3B and indicate that the effect of habitual action video game play is significant in both accuracy (g = 0.35) and speed measures (g = 0.54), although the results concerning speed may be taken with caution given the low number of effect sizes (k = 9, m = 6) and the associated low number of degrees of freedom (df = 2.2), compared with accuracy (k = 81, m = 19, df = 6.6). A significant effect of action video game intervention is observed for main effects, but not for difference effects, which also suffer from a low number of effect sizes and associated low degrees of freedom (df = 2.5). A significant effect is found for studies carried out by the Bavelier laboratory (g = 1.0, k = 18, m = 11), but not by studies carried by other laboratories (g = 0.2). Here the number of effect sizes for other laboratories is actually relatively healthy (k = 72, m = 11), however the associated degrees of freedom is extremely low (df = 3.8) as most other laboratories have used large batteries of tasks covering several domains of cognition when carrying out one intervention study. Given our choice of clustering at the level of study, this design choice by most other laboratories results in a heavy penalization in terms of degrees of freedom.

Continuous moderator of training duration. There was no effect of training duration in the multiple moderator analysis reported so far (slope = −0.003, 95% CI [−0.02, 0.01], k = 90, m = 22, df = 10.61, p = .73). Two additional models were also estimated to better understand the role of training duration as a moderator. An effect of training duration was found in the single moderator RVE analysis (slope = 0.015, 95% CI [0.00, 0.03], k = 90, m = 22, df = 9.25, p = .03) as illustrated in Figure 4. We then contrasted the effect of training duration with or without laboratory of origin included in the list of multiple moderators. There is a weaker but still visible effect of training duration when all moderators, but laboratory of origin, were included. Yet the RVE analysis showed this effect to be nonsignificant (slope = 0.009, 95% CI [−0.01, 0.03], k = 90, m = 22, df = 11.08, p = .32) suggesting that it is fragile. Finally, the effect of training duration totally disappeared when laboratory of origin was included in the multiple moderators considered. This pattern of results held for both the moving constant technique as illustrated in Figure 4, and the RVE analysis. This pattern of results is in line with the fact that the Bavelier laboratory is the only one to have carried out training studies of more than 30 hours; but it also indicates remaining heterogeneity in the impact of the other moderators on training duration.

![Cross-sectional meta-analysis. Contour-enhanced funnel plot of cross-sectional studies effect sizes. The cluster-average effect sizes aggregated over the 89 clusters/studies are shown in black triangles; the 29 additional clusters that have to be imputed to achieve symmetry as per the trim-and-fill procedure are shown in white triangles. The black dashed line shows the estimated effect size for the Random/Mixed effect model (g = 0.67) and the gray dashed line the estimated effect size for the Trim & Fill analysis (g = 0.48), suggesting a 28% reduction in effect size as a result of publication bias correction. See the online article for the color version of this figure.](image-url)
Overall, this analysis shows that low dosage studies tend to report smaller effects, and highlight differences between other laboratories and Bavalier lab’s high dosage conditions.

Publication Bias

We generated contour-enhanced funnel plots (Peters et al., 2008) and tested for asymmetry using Egger’s test, which confirmed significant asymmetry ($\beta = 4.15, SE = 0.1, Z = 4.16, p < .001$). To correct for the asymmetry of the funnel plot, trim-and-fill analysis (Duval & Tweedie, 2000) imputed 7 additional studies with negative effect sizes of low precision to make the plot more symmetric (see Figure 5). With these additional studies, the main effect of action video game intervention remained sizable and significant ($g = 0.40, 95\% CI [0.18, 0.62], m = 29, p < .001$), albeit 32% smaller compared with the Random/Mixed effect ($g = 0.58, 95\% CI [0.38, 0.79], m = 22, p < .001$).

As with cross-sectional studies, tests and corrections of funnel plot asymmetry were performed at the level of subpopulation effects corresponding to each moderator and level (Table S5). These analyses confirm the overall result of a publication bias threat of about 30%, and further show that selective reporting bias is more important in some measures than others. Indeed, while the overall effect of intervention studies was reduced by 32% upon trim-and-fill correction, effects at the level of subpopulation were reduced upon correction by a wide range from 0% (for the ‘verbal cognition’ level of the Cognitive Domain moderator) to 47% (for the ‘difference’ level of the Effect type moderator).

Figure 3. Intervention meta-analysis. Illustration of the effect sizes of action video game play in intervention studies with young adults for each level of the 4 categorical moderators (see Table 2 for full statistics). (A) Cognitive domain moderator; (B) The 3 other categorical moderators.
Additionally, multiple moderator analyses using RVE ($k = 90, m = 22$) indicated a significant difference between published and unpublished studies ($F = 14.91, df = 8.5, p = .004$), with published studies ($g = 0.40, 95\% \text{ CI} [0.17, 0.72], df = 5.3, p = .006$) showing effects in the opposite direction from unpublished studies ($g = -0.52, 95\% \text{ CI} [-0.03, -0.49], df = 8.5, p = .033$). This negative effect is in line with the fact that all 4 unpublished effect sizes were extracted from one record using Tetris as the active control game. Of these 4 measures, 3 concerned visuospatial cognition and were negative indicating greater enhancement on visuospatial skill from playing Tetris than an action video game, whereas the one measure of attentional control was positive. This pattern of results can be understood given Tetris play has been documented to selectively enhance visuospatial cognition (Adams, 2013; Uttal et al., 2013). For completeness, we report the complete outcome of this new RVE model in Table S7 of the supplementary information.

**Discussion**

The present meta-analyses shed light on the available literature on action video game play and its impact on cognition. This work departs from previous meta-analyses that have considered the impact of video game play in general by recognizing that not all video games have the same impact on cognition. Critically, the meta-analysis of intervention studies contrast action video game-trained groups with active control groups also required to play a commercially available video game, highlighting the importance of carefully considering game genre when assessing the impact of video game play on cognition. We also present a meta-analysis of cross-sectional studies that documents the potential societal impact of action video game play by contrasting self-selected action video game players with individuals who self-identified as nonaction video game players with limited video game exposure across all genres. These two meta-analyses significantly extend previously published meta-analyses considering the impact of action video gameplay.
especially those with training duration of 30 hours which have, to this point, only been performed in the Bavelier laboratory (see Figure 4 panel 3). We review below in detail the different points raised by each of the two presented meta-analyses.

Methodological Considerations—Strengths and Weaknesses

Robust variance estimate models. Throughout these meta-analyses, we used robust variance estimation with hierarchical effect weights. RVE allowed us to capture the hierarchical clustering of our data without requiring substantive assumptions about the correlation structure underlying effects. In addition to using RVE, we focused our analyses on models controlling for several moderators at the same time. Using these models—as opposed to the more common subgroup or single moderator models—allowed us to better isolate sources of heterogeneity (see also Johnson & Huedo-Medina, 2011).

The main effects of action video games documented by the present analyses are robust, whether in cross-sectional ($df = 24.8$) or in intervention ($df = 5.6$) studies. However, many of the moderator effects have rather small degrees of freedom (often less than 4)—indicating a lack of statistical power that can only be remedied with future research in these areas.

Small sample sizes. A possible weakness of the action video game literature, especially for intervention studies, is the small sample size used in most studies (between 13 and 45 participants in intervention studies). This suggests that intervention studies to-date are underpowered, particularly given the effect sizes found in the present meta-analysis. For example, assuming a correlation of .5 between pre- and post-tests, with two groups of $n = 15$ each, an effect size of 0.45 standard deviations would be required of the test to have 80% power. For an effect size of a third of a standard deviation (the average estimated here), two groups of $n = 29$ each would be required to achieve 80% power.

Possible publication bias concerns. Throughout these meta-analyses, special care was taken to locate and retrieve unpublished work related to the question of interest with, for example, 15 unpublished records of 73 being included in the cross-sectional meta-analysis. Yet, only 1 unpublished record of 21 could be located for intervention studies, despite using one and same search to identify these records. Given the concerns around publication bias, we conducted extensive publication bias analyses, focusing on both threats due to the ‘file-drawer’ problem (i.e., small studies with nonsignificant findings not published) and threats attributable to small-study bias (i.e., small studies tending to be conducted less well than large studies). Conducting these analyses was not straightforward, however, because our data have a hierarchical structure, yet standard methods for publication bias of both types assume that all effect sizes are independent. Additionally, making sense of the results of these analyses was also difficult, since the two types of bias do not necessarily have to point in the same direction, and since methods for small-study bias are very new and still being tested in the meta-analysis literature. Our approach, therefore, was to conduct many different analyses, including analyses by subgroups, always separately for cross-sectional and intervention studies. All of these results are provided in the supplemental information (see SI section 4.2 & 4.4). Throughout the text, we have attempted to make sense of these results, and hope that readers will take these cautions into account when interpreting results from this study.

The most consistent trend across all of the publication bias analyses was that the effect sizes estimated using the currently available data are likely larger than the true average effects. The trim-and-fill analyses suggest that the corrected effects are about 30% smaller than those estimated based on available data, whether for cross-sectional (28% reduction, for a corrected $g = 0.40$) or intervention studies (32% reduction, for a corrected $g = 0.23$; see Figures S5 and S6, respectively). In addition, they highlight greater potential for selective reporting bias in some measures than others, with a range for the correction between 0% and 34% reduction for cross-sectional studies and between 0% and 47% reduction for intervention studies. The small-study bias analyses (PET and
PEESE) suggest that these biases may be even larger. However, the fact that these analyses sometimes result in estimated effects in the opposite direction is perplexing, given the limited range of sample sizes found in these data. Thus, although the main result (i.e., potential threat of publication bias) from these analyses is shared with the other analyses conducted, the outcomes may also suggest limitations with the PET and PEESE approaches to publication bias detection and correction.

Does Action Video Game Play Impact Cognitive Processing?

The purpose of these meta-analyses was to summarize the literature examining the effect of playing action video games on cognition, both from the point of view of habitual game play and of intervention studies using action video games as their experimental training. Such a synopsis is particularly timely given that several research teams have started the process of translating the base research in this domain into practical use and that this domain has been the subject of several recent critiques whose main arguments could be effectively addressed via meta-analysis. Habitual action video game play was associated with more than a half standard deviation advantage across all measured domains of cognition (g = 0.55). In other words, individuals who choose to play action video games regularly exhibit better cognitive skills than those who play little to no video games. Furthermore, intervention studies using as their active training an action video game led to about a third of a standard deviation advantage across the domains of cognition considered (g = 0.34).

In all, our meta-analysis thus indicates a medium effect size of habitual action video game play on cognition, and a small-to-medium effect size of intervention studies directly testing the effect of action video game on enhancing cognition. This latter result is of note given the inclusion of only active control intervention studies, which are known to set a higher bar as compared with no control, or only test-retest control groups. These findings are consistent with the recent work of Powers et al. (Powers & Brooks, 2014; Powers et al., 2013). In examining the effect of action/violent games, Powers et al. (2013) reported Cohen’s d effect sizes of $d = 0.62$, 95% CI [0.52, 0.72] for cross-sectional studies based on 196 comparisons, and $d = 0.22$, 95% CI [0.13, 0.30] for intervention studies, based on 135 comparisons. Along the same line, Powers and Brooks (2014) focusing on intervention studies examining the impact of first person shooter games reported a Cohen’s $d$ effect size of $d = 0.23$, 95% CI [0.07, 0.39], based on 61 comparisons.

Does Action Video Game Play Impact All Domains of Cognition Equally?

Given the positive association between action gaming and enhancements in cognition (broadly construed) that were observed above, we further investigated whether significant enhancements were noted across individual domains of cognition and if so, whether the magnitude of the enhancements were equivalent across these domains. This question is of both theoretical relevance (e.g., most theories of learning/generalization suggest that the impact of a training regimen will be seen most clearly in domains that are loaded upon by the training) and practical relevance (e.g., because many research groups are using off-the-shelf action video games for translational purposes in the attempt to address problems with unique aspects of cognition—such as low-level vision or attention—it is critical to confirm that action gaming does indeed result in benefits in these domains).

In the analysis of cross-sectional studies, of the seven cognitive domains with sufficient numbers of studies to be included in this analysis, six were found to be positively impacted by action video game play, albeit to different extents. The three cognitive domains showing the most robust impact were perception ($g = 0.77$), spatial cognition ($g = 0.75$) and top-down attention ($g = 0.62$), with top-down attention showing the smallest confidence intervals and the most observations. A medium effect was observed for multitasking/task-switching ($g = 0.55$). The weakest effects were seen for inhibition ($g = 0.31$) and for verbal cognition ($g = 0.30$). The fact that verbal cognition is significant was perhaps the most surprising result here. Indeed, while action video game play clearly loads on perception, top-down attention, spatial cognition, multitasking and inhibition, it is not typically associated with verbal skills. Of note, most verbal cognition effect sizes included in the present meta-analysis were derived from verbal short-term or verbal working memory tasks. It is possible that such verbal tasks load on the type of cognitive flexibility that may also be at play in top-down attention, spatial cognition, and multitasking/task-switching enhancements. The possibility that enhanced cognitive flexibility after action video game play may help in the verbal domain has been recently explored in dyslexic Italian children. It was found that playing the action-packed minigames of Raving Rabbids led to improvements in reading speed that was not found after playing Raving Rabbids minigames that instead load on motor reactivity (Franceschini et al., 2013, 2017).

Although significant enhancements were observed in six of the seven domains of consideration, the observed variety of outcomes across cognitive domains suggests that action video game play is unlikely to equally impact all cognitive domains. It remains, though, that the cognitive domains with the fewest observations tended to show the weakest effects calling for more studies in not only multitasking/task-switching, inhibition and verbal cognition, but also problem solving. For now, we can safely conclude that being a habitual action video game player benefits cognition at large, with the most robust positive effects on perception, spatial cognition and top-down attention.

Critically, the results seen in the analysis of cross-sectional studies above were largely mirrored by the analysis of intervention studies, suggesting a causal role of action gaming in generating those effects. Among the five cognitive domains that could be interrogated, spatial cognition and top-down attention were clearly benefitted by action video game training, with these being followed by perception in terms of positive impact. This pattern of overall results, whereby spatial cognition, top-down attention and perception are the most promising domains, is a near direct match to what was observed in cross-sectional studies. Clearly though the field will benefit from a greater number of training studies, as, for instance, there are as yet too few studies available to draw firm conclusions as to the causal role of action gaming in enhancing multitasking and verbal cognition.

The similarity of the results seen in cross-sectional studies and those observed in intervention studies speaks to the continued value of cross-sectional work in this field. As in any other fields,
cross-sectional designs are unequivocally associated with potentially problematic confounds such as selection bias which are simply inherent to cross-sectional studies. Yet, other confounds such as participant expectation effects can be minimized via proper methodology. Indeed, if participants are recruited without knowing that their action-gaming status is of experimental interest, or in other words recruited covertly, participants will not be in a position to form expectations as to how their video game status may alter their behavior. In addition cross-sectional methodology has a number of distinct advantages. Cross-sectional studies are relatively fast to implement, are much cheaper than intervention studies, and are a common-sense first step before running a time-consuming and costly intervention study. It would seem premature to launch into a full randomized control trial if a null effect is obtained in a cross-sectional studies. A failed cross-sectional study can instead be an opportunity to revise the sensitivity of the tests used, the type of video game features to consider, the domains of cognition to be tested, or potential individual difference factors that could affect the impact of experience, before investing in a long-term intervention study.

**Does Action Video Game Play Impact All Ages Equally?**

In the same way that there is interest in whether action gaming impacts all cognitive domains equally, there is also interest in whether action gaming impacts individuals of all ages equally. Although the available data are not ideal with respect to addressing this question (in that there were no intervention studies in children and no cross-sectional studies in elderly individuals), the data that are available do begin to speak to the issue.

In cross-sectional work, studies involving child participants made up 2.5% of the weighted effect sizes, with all children studies focusing on top-down attention. Despite these caveats, a positive, though nonsignificant, trend was observed in children ($g = 0.33$, 95% CI $[-0.34, 0.99]$), $df = 2.9$). Such a finding is generally consistent with data showing that younger children tend to show substantial neuroplasticity. This positive result, however, should not be interpreted as a call for children to indulge in action video game play (particularly given the fact that, while there is a critical distinction between “action” and “violent” games, it is nonetheless unequivocally the case that many action games are violent). Rather, it is both a call for more studies to be carried out in child populations using nonviolent action video games, such as is seen in some of the minigames found in Rayman’s Raving Rabbids or in Splatoon, as well as a call for the industry to develop age-appropriate narratives that incorporate the game play mechanisms of action video games in a way that is appealing to children, especially girls.

In terms of intervention studies, a meta-analysis including both young and older adults indicated numerically opposite effects of action video game play in young and older adults. While the small number of effect sizes and the associated large confidence intervals in older adults indicates it is premature to conclude that there is different impact of action video game play across these age groups, there are both theoretical and practical considerations that have led us to focus solely on intervention studies for young adults. First, a recent meta-analysis by Wang et al. (2016) reported a larger impact of video game play in young adults as compared with older adults, calling for caution when mixing these two age groups. Second, there are theoretical reasons why the effect of action video game-based interventions in older adults may be different from that in young adults. Indeed, such studies are likely to violate some of the main principles of learning. For example, action video game play, as defined here (mostly commercial first or third person shooter games), presents a substantial mismatch between the difficulty of the game play and the diminished perceptual, cognitive, and motor abilities that characterize the older age group. In short, commercially available action video games are specifically produced for young adults. Designers therefore take into account the typical abilities of young adults when determining game characteristics. Because of this fact, these games are likely to be too challenging and thus engender frustration and helplessness, rather than engagement and attention, in older adults. This is consistent with the findings of McDermott and collaborators who attempted to utilize a commercially available action video game (Medal of Honor: Heroes 2 on the Nintendo Wii) to train older participants (McDermott, 2013). The authors found that the game could initially only be played by the older adults in the easiest game mode. In this mode, the player was moved through the space by the computer, rather than the player navigating him/herself. As such, the player was only responsible for aiming and shooting (which the gaming world describes as being “on rails”). This situation negates the need to effectively monitor the entire screen, to decide on action plans, to multitask, and so forth. This in turn reduces the load placed upon these processes and through this minimizes the extent to which a positive impact of the gaming experience may be realized.

A second and potentially associated issue that arises when attempting to train older adults with action video games is related to compliance. For instance, Boot and colleagues (2013) found much lower compliance in the group of older adults who were asked to train on an action game (individuals in this group played an average of 10 hours in total), as compared with the group of individuals who were asked to train on a control nonaction game (individuals in this group largely completed the requested 60 hours of training; note that this study was not included in the present meta-analysis for the very reason that the hours of training were not equivalent between the action-trained and control-trained groups).

This set of results also illustrates the difficulties of translating the current results established on healthy and young adults to other, especially clinical, populations. First, it is critical to remember that learning with video games follows the same main principles as those documented in the field of learning (Stafford & Dewar, 2014). Thus, any action video game-based intervention needs to tailor the game play so that it matches the skill level and possible growth of the trainee. In short, the fact that commercially available action video games enhance top-down attention in young adults does not imply that the very same games will do so when administered to older adults. Instead, although the game mechanics should be conserved, the game pacing and load should be adapted to the trainee. Second, when discussing video games that are as complex as action video games, it is critical to remember that players may vary in their game strategy. For example, individuals with attention-related problems may be expected to benefit from their action game play given the robust impact of action game play on top-down attention. Yet, although many individuals diagnosed
with ADD/ADHD report playing such games, no studies have ever documented a positive impact of commercially available action video games on individuals diagnosed with ADD/ADHD. Although only observational, it is our experience that action video game players diagnosed with ADD/ADHD, when tested in our laboratories, struggle with our computer-based assessments of top-down attention. It is very possible that these individuals approach the action game play with different strategies that limit the benefits they can reap from such game play. In particular, individuals diagnosed with ADD/ADHD may enjoy a more reactive game play, whereby they throw themselves in the middle of the action reaping the reward of that rapid interactive experience, but in doing so fail to develop the proactive play that is more characteristic of healthy individuals. A challenge when translating this work for clinical application is to develop action-packed video game that limits reactive play and rather foster proactive planning.

Are the Action Gaming Benefits True Benefits in Processing or Are They Related to Strategy Shifts (e.g., Speed–Accuracy Trade-Offs)?

One consistent question that arises related to the impact of action gaming on cognition is the extent to which the benefits reflect true enhancements in processing versus strategy shifts of some sort, with speed–accuracy trade-offs being one possible strategy commonly put forward. To examine this possibility, we contrasted results in studies where the primary dependent measure was accuracy with studies where the primary dependent measure was RT. If the benefits arise exclusively due to a speed–accuracy trade-off, enhancements should be noted only in those tasks that emphasized speed and not those that emphasized accuracy. In contrast to this hypothesis though, in cross-sectional studies, the impact of action video game play on cognition appeared equally strong whether speed ($g = 0.61$, 95% CI [0.47, 0.76], $df = 31.5$) or accuracy ($g = 0.58$, 95% CI [0.47, 0.70], $df = 31.2$) measures were considered. This pattern was mirrored in the intervention studies where both accuracy ($g = 0.35$, 95% CI [0.16, 0.53], $df = 6.6$) and speed ($g = 0.54$, 95% CI [0.07, 1.00], $df = 2.2$) appear to benefit from action video game training (although the effect on speed remains to be confirmed given the low number of studies). These results reinforce the view that speed–accuracy trade-offs are not a likely result of action game play training. This point is particularly relevant for those research teams using action games for rehabilitative or job-training purposes, where true changes in processing are needed.

We acknowledge though that our analysis is not optimal, and that better tests of speed–accuracy trade-off require that both measures be available in all (or at least most) studies, which was not the case here (see Donnelly, Brooks, & Hornet, 2015 for an example of this approach).

Is There Evidence for the Expectation/Placebo Hypothesis?

Recent critiques of the action video game literature published over the past few years (Boot et al., 2011; Boot & Simons, 2012; Boot, Simons, et al., 2013; Kristjansson, 2013) have argued for some form of the expectation hypothesis—the possibility that the reported impact of action video games is attributable to participants’ expectations about the outcome(s) of the study. We ran a number of different analyses to address the possibility that through expectations some of the reported effects may be more attributable to a placebo effect than a causal effect of action gaming per se. In terms of cross-sectional effects, the most extreme version of the expectation hypothesis predicts that positive effects of action video gaming should be found only in studies where participants are made aware that their action video game status is of experimental interest. Meanwhile, no effect should be found in studies where participants are unaware that gaming is a factor of interest. Indeed, if the participants do not know that their gaming habits are of experimental interest, they cannot purposefully alter their behavior to match the experimental hypotheses. In contrast to the expectation hypothesis however, medium-to-large effect sizes were observed for both overt ($g = 0.62$, CI [0.50, 0.75]) and covert ($g = 0.50$, CI [0.33, 0.68]) recruitment studies.

And while the outcome above convincingly demonstrates that habitual action video game play impact cannot be simply explained away by participants’ expectations, we further examined the question by asking whether effects were only seen for hypotheses that should be easier for participants to intuit (e.g., main effect hypotheses such as, “action gamers should respond faster” or “action gamers should be more accurate”) and not seen for hypotheses that should be more difficult to participants to intuit (e.g., difference effect hypotheses such as, “action gamers should be disproportionately faster in switch trials as compared with non-switch trials”) as would be suggested by the expectation hypothesis. Again though, in contrast to the expected results derived from the expectation/placebo hypothesis, a robust impact of action video game experience was observed for both main effects ($g = 0.62$) and difference effects ($g = 0.52$). That habitual action video game play can influence difference effects speaks against the hypothesis that the source of action game player advantages relies entirely on participants’ expectations.

Finally, in terms of intervention studies, although recruitment method is clearly not of concern, it may nonetheless be possible for participants to form expectations based upon their assigned training game in intervention studies. As such, we also examined whether effects in training studies were only observed for easy to intuit hypotheses (main effects) and not for difficult to intuit hypotheses (difference effects). Again though, in contrast to the expectation hypothesis, numerically the effect of action video game training was sizable for both main ($g = 0.36$, $df = 4.9$) and difference effects ($g = 0.43$, $df = 2.5$) effects. Although the small number of clusters for difference effects prevents us from drawing final conclusions, this pattern of results is not consistent with the view that action video game training acts through expectation effects. The present results do not support the most extreme version of the expectation/placebo hypothesis, which would hold that ALL of the benefits associated with action gaming experience arise via expectation/placebo effects. Yet, they do not rule out a less extreme version of the expectation hypothesis. Indeed, rather than positing that expectations fully explain the previously observed differences between AVGPs and NVGPs, it could instead be the case that the presence of expectations exaggerates differences between the groups. In other words, under this form of the expectation hypothesis, differences in performance do exist between AVGPs and NVGPs that are unrelated to expectations; however, the presence of expectations may serve to increase the magnitude of this effect. The current analysis provides some support for this form of the hypothesis. Although the observed effect was not significantly larger in overt recruitment than in covert recruitment studies ($F = 1.61$, $df = 13.2$, $p = .226$), it was certainly numer-
ically higher in overt than in covert recruitment studies. This points toward a possible role for expectation in increasing the size of the difference in performance that is observed between AVGPs and NVGPs. We note, though, that the main versus difference analyses are more equivocal with main > difference in cross-sectional studies and difference > main in intervention studies.

**Examining the Effect of Laboratory of Origin**

In cross-sectional studies, laboratory moderated the impact of action video game play \( (F = 4.7, df = 26.8, p = .04) \) with a significant larger impact in Bavelier studies \( (g = 0.80) \) than in other laboratories studies \( (g = 0.51) \), even after controlling for differences in domains and other moderators. Importantly, the impact of habitual action video game play was still highly significant when considering studies of other laboratories on their own. Several factors may contribute to this effect, including cognitive domain studied, type of recruitment, and participant selection criteria. Indeed, the Bavelier laboratory has a disproportionate number of studies on perception and top-down attention—domains with some of the largest effect sizes, and has mostly conducted overt recruitment studies (when this line of research begun, action video games were unanimously believed to be mind-numbing). The Bavelier laboratory has also used stricter selection criteria than the rest of the literature for participant inclusion (see further discussion below). The relative contribution of each of these factors remains difficult to assess.

Similarly, in intervention studies, effects documented by the Bavelier laboratory were significantly greater than that documented for those from other laboratories \( (F = 20.0, df = 5.3, p < .006) \). Exactly why the results for the Bavelier lab are larger, however, is unclear from the current data. This is because the effects of laboratory and training duration are nearly confused: more than 90% of the studies conducted by other labs used a training duration of under 30 hours, whereas more than 80% of the studies conducted by Bavelier lab used a training duration more than 30 hours. This makes it difficult to tease apart the effects of laboratory versus training duration, without making extrapolations. To illustrate this, in Figure 4 (Panel 3), we include estimated relationships separately for Bavelier and other labs. This shows that the effects in prior models for training duration (panels 1 and 2) were driven by the Bavelier labs and their overall longer training durations; in comparison, for the other labs, with shorter durations, the effect is not significantly different than zero. One implication of this finding is that it would be useful for other groups to invest in training studies of longer durations.

**Going Forward**

Although the data clearly indicate a significant and positive impact of action video gaming on cognition, there remain a number of key issues going forward.

1. **Defining action video games: Now and going forward.**

   The present meta-analysis is the first to focus exclusively on action video games and to compare playing action games to playing other commercially available video game genres. What is termed action video game in the context of research is, however, only a subset of what is labeled the “action video game genre” in the popular media (see e.g., [https://en.wikipedia.org/wiki/Action_game](https://en.wikipedia.org/wiki/Action_game)). As discussed above, the present meta-analysis focuses nearly exclusively on first and/or third shooter games, and thus does not include sports, fighting, maze, platforms, rhythm, or real-time strategy games. In fact, in some of the studies included here, these other game types served as control video games.

   This distinction matters because training studies illustrate that not all video games have the same impact (Cohen et al., 2007; Powers & Brooks, 2014; Powers et al., 2013). The results documented here only hold primarily for first and/or third person shooter games, which were termed “action video games” by Green and Bavelier in 2003, before many of today’s game genres became mainstream. As the video game industry has evolved over the past decade and a half, some totally new game genres have emerged (e.g., rhythm, music, real-time strategy). In other cases, multiple preexisting genres have merged to create new hybrid genres (e.g., action-RPG, action-adventure). Of particular relevance is that action components are becoming increasingly common in genres that previously had very little action—for example, fantasy, role-playing, and strategy games—and thus one would expect that games with these action components, even if they are from classically nonaction genres, would nonetheless still produce benefits to cognitive skills (Dale & Green, 2017). Indeed, the results of a recent study indicated that 40 hours of training on the real-time strategy game StarCraft resulted in enhanced cognitive flexibility as compared with the control video game The Sims 2 (Glass et al., 2013). In light of this state of affairs, any meta-analysis considering cross sections such as action games, non-action games, or non-specific game genres, will merge video games with highly heterogeneous mechanics and cognitive impact; as a result, such meta-analyses are likely to be at best uninformative and at worse misleading.

   In all, the direction that video game genres have taken over the past decade (in particular, borrowing popular features across genres) has made it less and less appropriate to study the impact of a given game genre on cognition as genre is an increasingly poor indicator of the core cognitive processes employed when playing the games. Instead, a system that classifies games via specific game mechanics and dynamics that are more or less likely to foster enhancements in cognition is needed. This would not only aid in the classification of players in cross-sectional work, but also in choosing experimental and control games in intervention studies. Most importantly, this also represents the type of foundational knowledge needed for building games for impact.

   The existing body of work points to a number of key game features that predict such positive impact, above and beyond factors such as fun, time on task and immersion, which are present in both action games and the commercial video games used as control games in the training studies reviewed above. These hypothesized key game features include pacing (making decisions under time constraints), divided attention, a constant need to dynamically alternate between focused and divided attention, sufficient variety in contexts and tasks to avoid automatization, and environments that have enough structure to foster prediction and limit reactive behaviors, among others. Research is ongoing to test the relative impact of these different components in inducing cognitive enhancements, keeping in mind that a confluence of multiple components is likely key.

2. **Who counts as action video gamers: Now and going forward.**

   The selection criteria in the present meta-analysis were relatively
loose compared with those used in the Bavelier laboratory over the past 15 years. For example, the present meta-analysis included all studies with an AVGP group that played 3 hours or more of action games per week, whereas our group has always required at least 5 hours per week during the past year, and more recently has allowed 3+ hours per week during the past year if participants reported playing 5+ of action video games before the past year. In addition, the Bavelier lab has always excluded potential AVGPs who, in addition to their action gaming, report playing 3 hours or more of any other game genres (thus ensuring we were able to draw conclusions about the impact of action games specifically). Similarly, NVGPs in Bavelier studies have been required to play at most 1 hour per week during the past year of action first/third shooter games, sports, RPG, driving or adventure video games and at most 3 hours of any remaining video game categories combined. Critically, those individuals who report playing more than one hour per week before the past year of action first/third shooter games, sports, RPG, driving or adventure video games are also excluded.

The present meta-analysis on the contrary utilized much more lenient criteria. Here, to be included, NVGPs needed to play at most 1 hour of action video games during the past year or at most 3 hours across all game genres. This potentially allowed the inclusion of participants who were avid players of other genres or who had been avid action video game players in their past. This leniency was partially necessitated by the fact that the required level of detail for the stricter criteria we utilize in our group is almost always lacking in the methods section of papers, and many authors did not have such information when contacted. This may explain in part why the magnitude of the effects reported by Bavelier and colleagues is somewhat greater than that reported by all other groups in cross-sectional studies.

We note that the Bavelier laboratory shares part of the responsibility in not clearly outlining all of the inclusion/exclusion criteria in methods sections. As well, we can see how some of our choices of terms may have been misleading to others, as exemplified by the mismatch between our restricted use of the term action video game to first or third shooter games and the 2015 Wikipedia definition of the action video game genre. We provide in the supplemental materials the 2017 version of our video game questionnaire and the selection criteria we have used in the laboratory. It should be clear that the game categories listed are motivated by research considerations and not by industry classifications. Yet, examples of games and our labels for game categories have evolved over the years in concert with the changing landscape of video games.

The issue of how to screen participants as to their gaming habits is a delicate one. The Bavelier lab’s current video game play questionnaire is not without weaknesses. It was designed based upon the recommendations of the General Household survey in the U.K. (Office of Population Censuses & Surveys, 1996). The goal of the questionnaire has primarily been to identify extremes of behavior—that is individuals who mostly play just a large amount of action video games as well as individuals who do very little gaming overall. When used in this way, it produces robust results as demonstrated here. It should be noted though that the two targeted populations are now relatively rare, representing between 5%-10% of the student population each, a fact that might explain in part the rather low Ns found in this literature. Given the aim of characterizing the impact that playing various video games has on the brain and behavior, it would seem on the face to be more desirable to assess each individual’s complete set of gaming habits (i.e., the amount they play each of every possible genre).

However, it is unclear whether a full assessment of gaming habits is presently possible. First, the currently utilized questionnaire is ill-suited for classifying intermediate players, as it lacks the necessary granularity for those individuals who happen to play several game genres regularly. Relying on nearly 1000 participants worth of empirical game play questionnaire, Green et al. (2017) have recently established that the exact hours reported by participants are not accurate. In particular, individuals who play multiple types of games tend to show substantial and highly nonlinear mis-estimations of their absolute gaming times. This is a critical issue as such multigenre players are an already sizable and likely to continue growing portion of the population. Thus although the questionnaire approach is able to reliably identify participants with extremes game play in one genre, it is not suitable if the aim is to build complex regression models that attempt to take into account differences in gaming across many different possible game genres, as done in Unsworth, Devilly, and Ward (2007). As recently commented by Green et al. (2017), such an approach is destined for failure, as intermediate players are unable to accurately indicate their playing time when engaged in several game genres regularly.

Second, beyond the issue with the survey are the issues with attempting to classify many different genres in the first place. As noted above, genre boundaries are increasingly indistinct, making categorization at a finer level of fidelity extremely tenuous (e.g., should one put the game “Skyrim” into the “RPG” category, the “action” category, or create a new “action-RPG” category—as is done in, for instance, the Wikipedia page for the game). Adding even more difficulty is the fact that many games today permit multiple play styles and thus a game such as Skyrim can be played as either a “pure RPG,” “pure action game,” or some mixture thereof depending on how the individual player approaches the game (Dale & Green, 2017).

A third issue that is becoming increasingly key in terms of participant classification is that of past experience. Many of today’s college students were born before the start of the so-called 6th generation video game console era (i.e., the Playstation II, the Xbox, the Nintendo GameCube, etc.) and thus have potentially been exposed to action video games for their entire lives. Because in experimental work we, and others, have seen that the effects of action games are quite long lasting (at least on the scale of several years), this means that a college student who had played no action games over the past year, but who played avidly all through high school would very likely pattern more like an action game player than a nonaction game player. This is why the Bavelier laboratory includes experience before the last year, but this was not realistic for the present meta-analysis to incorporate. From this point of view, cohort studies would be highly valuable to track video game impact on cognition as video game accessibility and genres continue to evolve.

In all, this state of affairs calls for more in-depth reporting in methods sections of exact participant selection criteria. Furthermore, selection surveys should attempt to gain a fuller measure of participants’ gaming, including their full history of game play. And although this is a worthy goal it will be a hard goal to reach.
This issue is clearly exemplified in other fields facing the same issue, such as estimating alcohol consumption or sleep habits in health science. Simple survey methods are known to be quite unreliable at probing long past behaviors, with methods such as daily diaries being preferred (though these are less helpful when attempting to assess the previous game play habits of individuals now in college). In the case of video games, quantitative measures are also potentially available via online game accounts and these appear highly promising in properly assessing present and past consumption. Its implementation is, however, complicated by the wide variety of existing platforms and the fact that some of this data can be proprietary.

Finally, another phenomenon that researchers will need to take into account is the greater eco-system of new technologies and the fast pace with which it is evolving. The advent of smart phones, iPads and multifunction devices has led to an increase in media multitasking. Although hailed by society as a way to increase productivity, a seminal 2009 study suggested high media multitaskers have weak cognitive control (Ophir, Nass, & Wagner, 2009). Our own work recently confirmed the different impact of action video game play and media multitasking on top-down attention, multitasking and spatial cognition (Cardoso-Leite et al., 2015). Interestingly, individuals who qualified both as action video game players and high media multitaskers exhibited the same low cognitive control as high media multitaskers. Thus, not only do different media experiences have different demands, and thus different impacts, but they also interact in a nonlinear and not necessarily easily predictable way. This state of affairs reinforces the importance of carefully considering the exact technology used by an individual.

3. True experiments: Key design issues in interventions.

The value of active controls. The intervention meta-analysis documents an increment of about a third of a standard deviation across all cognitive domains in action video game trainees as compared with control video game trainees. Indeed, a rather unique feature of the present meta-analysis is that we included only intervention studies with an active control group that played a commercially available, immersive video game as a control game, in an effort to equate the known engagement delivered by the action video games used as experimental intervention games.

Another distinguishing feature of this work is its focus on the long-lasting impact of action video games play, as enforced by including only studies with at least a 24-hr delay between the end of training and posttesting. Intervention studies in other domains such as for example the aggression literature allow for the measurement of exclusively short-lived impact by posttesting participants only a few minutes after the end of training. In contrast to these short-term effects, long-lasting brain plasticity is notoriously difficult to induce in adults, which means studies in the field call for rather long training durations (8–50 hours).

The fact that significant effects were found for intervention (i.e., training) studies, and not just cross-sectional (i.e., habitual/novice) comparisons, indicates that there is a causal relationship between playing action video games and enhancements in cognition in healthy young adults. These results are all the more noteworthy in that the present meta-analysis considered only intervention studies that contrasted action video game play to active control group(s) in which control participants also played similarly engaging and challenging commercially available video games which, unlike a simple placebo, has active ingredients of its own. As noted by Uttal and colleagues (2013, p. 12), “a high-performing control group can depress the overall effect size reported.”

Yet, we would argue that the use of control group(s) where participants are active and challenged is necessary for a more useful assessment of the efficacy of any intervention, be it action video game play or others. Indeed, by contrasting action video games to other commercial video games, we can rule out possible confounds ranging from those related to simply being exposed to an immersive form of media such as a video game, to feeling aroused by a new and challenging set of experiences, to being confronted with the rich reward structure that is inherent in commercial video games, in addition to the more standard concerns about test-retest improvements and Hawthorne-like effects.

Living up to the standard of double-blind randomized controlled studies. Another weakness of the intervention studies included in the current analysis is that experimenters were rarely blind to the assignment of the participants. Keeping the experimenter who will administer the tests separate from those who will oversee the training should be preferred, and may be easier to achieve if the participants train at home. Few of the included studies indicated that this was done, potentially due to the separate methodological concerns that arise with home training (e.g., as related to overall compliance, consistency of training conditions, etc.). Unless such a design is planned at the time of funding, maintaining entirely separate staff for testing and training throughout the intervention duration can be quite taxing for a single laboratory. Future studies should, however, strive to do so. Anecdotaly, we note that because the participants necessarily cannot be blinded to their condition (i.e., they obviously know the activity they are being asked to perform), it is thus important to inform participants not to comment on their training with testing staff to avoid participants themselves giving cues and potentially biasing the testing staff.

Further, unlike the experimenters who can potentially be blinded, participants cannot be blinded to their intervention. Thus, participants may formulate many different expectations as to the hypothesized impact of their assigned video game. Under such circumstances, it seems important to frame the study as the evaluation of two potentially active video games and keep the control video game as engaging and as challenging as possible. In the present meta-analysis, only studies that used commercially available video games as active controls were included, guaranteeing high engagement and challenging play in the control game. The use of commercial games as control experiences should also at least partially limit expectation effects. Indeed, although participants may know that the research they are participating in is about video games, they are unlikely to have the technical knowledge to intuit how each different video game is hypothesized by the research team to alter behavior (particularly those trainee participants that are typically not well versed in video games).

Although only rarely done, more and more laboratories are systematically questioning participants about their expectations at the end of studies to be in a position to directly assess differences in expectation across groups (e.g., Blacker & Curby, 2013). Unfortunately, the value of such probes remain unclear (as there is evidence in the field that participants may be unwilling to indicate their beliefs about study goals, e.g., Nichols & Maner, 2008).

Furthermore, if participants, upon being questioned, do indicate...
some beliefs regarding the study hypotheses, there is no effective way to know whether these beliefs were present throughout the study, or if they were in fact produced by the probe question itself. Indeed, by measuring expectations, we may either change the expectations, or, possibly more troubling, create expectations. Reaching common agreement about the structure and interpretation of such poststudy expectation questionnaire should be a valuable next step for the field.

**Time course of training and duration of the effect.** Beyond the content of the training itself, future work is also required to fully map the time course of training effects. Although the studies included in the present analysis used between 8 to 50 hours of training, the exact amount of training necessary to induce long-lasting changes remains unclear. Furthermore, the necessary training length may depend on the cognitive domain. For example, although training periods on the order of 10–20 hours appear sufficient to observe changes in top-down attention in young adults, such doses may not lead to changes in perception, with most successful perception studies using training durations of 30 to 50 hours.

In terms of practical application, while one may observe changes that are statistically significant after 10 hours of training, one should not expect that 10 hours of training will produce changes of the magnitude that are required to have practical relevance in everyday life. Instead, changes of a size that is practically relevant may require substantially longer total amounts of training, distributed in small doses over periods of weeks or months. As documented by Stafford and Dewar (2014), video game training seems to obey the same rules as essentially all learning interventions, in particular that distributed practice provides the best learning outcomes. Unsurprisingly then, those studies that have used highly massed practice (e.g., individual training sessions that last for four straight hours) have seen little success (Gallagher & Preswicht, 2012; van Ravenzwaaij et al., 2014).

Finally, the amount of training required may very well vary during the life span. Given the current state of knowledge, it may be time for the field to consider using longer training durations than have been typically employed, as well as to carry out more systematic dose-response curve studies. The latter may be a practically challenging proposition as repeated testing should be avoided when evaluating transfer effects (Green et al., 2014). Different groups of participants who have undergone different intervention durations will have to be compared, with the durations spanning not just days and weeks, but possibly months and years (Jaeggi, Berman, & Jonides, 2009).

4. **Other challenges for the future.** A final important avenue in future studies will be to uncover how, at the neural level, action video game play impacts various skills. An increasing number of studies have begun to address the structural (Basak, Voss, Erickson, Boot, & Kramer, 2011; Erickson et al., 2010; Kühn, Gleich, Lorenz, Lindenberger, & Gallinat, 2014; Kühn et al., 2011; Sagi et al., 2012; Tanaka et al., 2013; Vo et al., 2011) and functional (Bavelier, Achtman, Mani, & Föcker, 2012; Lee et al., 2012; Mishra, Zinni, Bavelier, & Hillyard, 2011; Prakash et al., 2012; Voss et al., 2012) neural correlates of video game play. Unfortunately, not all video games equally impact behavior and thus the underlying neural changes must also be different. Only a handful of cross-sectional studies to date have examined the neural correlates of improved attentional control specifically in action game players. These studies have indicated changes in the fronto-parietal network (known to mediate attentional control) in action game players (Bavelier, Achtman, et al., 2012; Föcker, Cole, & Bavelier, 2014; Mishra et al., 2011; Krishnan et al. 2013) with the one intervention study that specifically utilized action video games providing convergent evidence for changes in this network (Wu et al., 2012).

Although there is mounting evidence for alterations of the fronto-parietal network of attentional control in action video game players, the type of synaptic remodeling needed to give rise to the behavioral benefits linked with action video game play is bound to also be observed in areas mediating task performance. Using neural models of the behavioral impact documented, we have shown how changes in network connectivity in those neuronal networks known to mediate performance on the task of interest may account for the behavioral benefits noted (see Bejjanki et al., 2014; Green, Pouget, & Bavelier, 2010). Although it has been proposed that changes in top-down attention may help guide the changes in network connectivity, a direct link between these processes remains to be established. This may be best achieved through both cognitive and neural studies, as the neural mechanisms underlying learning and transfer still largely elude us (Sasaki, Nanez, & Watanabe, 2010).

5. **Contextualizing the current results: Potential adverse effects of video game play.** The interested reader may wonder how the reported effects here link to a number of recent debates about the potential negative effects associated with video game play. These include the impact of violent video games on a variety of negative outcomes, including aggression, affect and social skills as well as the potential for video game play to become addictive.

With respect to the influence of violent video games, as stated in the introduction, action and violent video games are not synonymous. A game may utilize “action” mechanics in the absence of any violent content—as is the case, for example, of cooperative shooter games such as Splatoon, or some of the child friendly shooter minigames within Rayman’s Raving Rabbids. It is also equally possible to have violent content in games without any action characteristics as is the case in many turn-based role playing games (e.g., Final Fantasy VII). Thus, these two literatures, while asking parallel questions in terms of games and their impact on behavior, have evolved independently of each other, as they focus on different games and different behavioral domains.

These differences also extend to methodology. For obvious ethical reasons, intervention studies using violent video games to test whether there are causal long-term negative outcomes are not acceptable. As a result, intervention studies in the violent video game literature are not to be equated with the intervention studies in the present work. The type of intervention studies conducted in the field of cognition and the type of intervention studies conducted in the field of social behavior use fundamentally different methodology. In the violent video game literature, most intervention studies consist of short episodes of
violent video game play (between 10 and 40 min) followed by an immediate evaluation of their impact on aggression, empathy, affect or social skills. A key distinction is that these effects are known to be transient, lasting on the order of just a few minutes or tens of minutes before participants return to their baseline behavioral pattern. That the expected effects are transient is the very reason these intervention studies are ethically acceptable. In the case of action video games, such transient effects on cognition have been seldom studied (one intervention study documented increased understanding of a lesson on plate tectonics after 25 min of action game play Sanchez, 2012). Yet, by and large, these short-term, fleeting effects have not been the focus of the literature on action video games. And indeed, such designs do not align with the translational goals of the action video game literature. That is, whether we consider patient rehabilitation or educational applications, the enhanced cognition effect reported with action video game will be of practical significance if and only if it is sustained over time (otherwise it will not benefit everyday life). An effect dependent on a short-lived state induced by the video game play itself would be of much reduced practical use. Given these key design differences, it makes little sense to directly compare the impact of intervention studies in these two literatures.

Both literatures, meanwhile, do utilize correlational designs. However, the rationale for these studies is rather different. In the action video game literature, the only correlational designs found are cross-sectional studies, which mostly act as a first guide to identify where to invest in time-consuming and high cost intervention studies, which require tens of hours of training. In the violent video game literature, cross-sectional and longitudinal studies are the key sources of information to assess whether violent video game play leads to long-lasting negative outcomes. Indeed, in the absence of sustained intervention studies testing long-lasting effects of violent video game on negative outcomes, only analyses of populations that have chosen to subject themselves to violent video game can inform us about the long-lasting impact of violent video game exposure. Accordingly, while cross-sectional designs have been used, longitudinal designs are the gold standard in the violent video game literature when it comes to assessing long-term causal impact. In contrast, the gold standard in the action video game literature is long-duration training studies with impact being assessed days, weeks or months after the end of the intervention.

In the future, it would be interesting to better assess the short-term impact of action video game play on cognition so as to examine the possible link between the existence of a short-term impact and the long-term cognitive enhancement(s) documented here. There is evidence that heightened arousal, greater vigilance, immersion in the flow state results in enhanced performance—an effect probably best documented by the (in)famous “Mozart effect” (Pietschnig, Voracek, & Formann, 2010). Yet the link between such transient boosts in cognition and long-term plasticity in the cognitive domain remains largely elusive. We note that this is not only a question for cognition, but similarly so for affective and social behaviors, where although it is the case that several hypotheses assume a build-up of transient negative outcomes into a long-lasting effect (Anderson & Bushman, 2002; Bandura, 1986), this link has never been directly put to the test.

Another growing concern is the potential for video game play to be addictive. “Internet Gaming Disorder” (IGD), as it is technically referred to, would be placed in the same basic category as behavioral disorders, such as pathological gambling. IGD, however, is not yet recognized yet as an official diagnosis in the latest version of the DSM (it is instead listed as a diagnosis in need of further research; American Psychiatric Association, 2013). In this young but quickly evolving field, much of the research to-date remains focused on basics such as defining symptoms, testing diagnostic criteria, assessing prevalence, outlining patterns of comorbidity and evaluating clinical relevance (D. A. Gentile, 2009; Király, Griffiths, & Demetrovics, 2015; Przybylski, Weinstein, & Murayama, 2017; Rehbein, Kliem, Baier, Mossle, & Petry, 2015; van Rooij et al., 2014). In most cases, this work has aggregated over all game genres, rather than examining whether certain genres are more or less associated with IGD. This in turn makes it difficult to evaluate the literature on IGD in the context of the current metaanalysis, that has focused specifically on the impact of the action video game genre. The little research on IGD that has been conducted in this vein though has indicated that while action video game play is indeed associated with an elevated risk of IGD (as compared with, for instance, no gaming whatsoever), there are a number of other game genres, such as real-time strategy games and role-playing games, that are associated with significantly higher levels of risk (Eichenbaum, Kattner, Bradford, Gentile, Choo, et al., 2015; Eichenbaum, Kattner, Bradford, Gentile, & Green, 2015; Lemmens & Hendriks, 2016). At the neural level, those subsystems involved in addiction are tightly coupled with those implicated in various forms of learning, including cognitive learning. Whether video games happen to be outstanding teaching tools (D. A. Gentile, Groves, & Gentile, 2014) because of such coupling will be an important future direction for research (Huys, Tobler, Hasler, & Flagel, 2014; Hyman, Malenka, & Nestler, 2006). Indeed, the potential for IGD should be taken into careful consideration by those utilizing, not just action video games, but any type of video games as potential cognitive enhancement interventions.

Conclusion

Overall, the present meta-analysis confirms a medium size impact of habitual action video game play on cognition and a small-to-medium size effect of training young adults with action video games on a few cognitive domains. Yet, much work remains, in particular in leveraging key action mechanics to create action games appropriate across the life span; in carrying out training studies with improved methodology and in particular larger sample sizes and of longer duration (an expensive proposition); in understanding the potential role of expectations in cognitive enhancements and how they may be put to good use rather than considered as a nuisance variable; and in elucidating the appropriate dosage in training regimens, recognizing that different cognitive domains may require different amount of practice to alter. Future studies exploring how action video game play not only impacts performance, but also how it alters
learning curves across various cognitive domains may be particularly informative in this respect.

References

References to studies included in the meta-analysis are preceded by an asterisk.


Received February 12, 2015
Revision received September 12, 2017
Accepted September 21, 2017

---

**Psychological Bulletin Call for Papers: Replication and Reproducibility: Questions Asked and Answered via Research Synthesis**

The primary mission of *Psychological Bulletin* is to contribute a cohesive, authoritative, theory-based, and complete synthesis of scientific evidence in the field of psychology.

The editorial team is currently interested in the contribution of meta-analysis and systematic reviews to answering questions about the replication and reproducibility of psychological findings. This call is designed to encourage authors to consider the degree to which multi-study primary research papers, systematic replications efforts (e.g., Open Science Framework and Many Labs Replications), and meta-analysis can contribute to substantial psychological knowledge.

The editors of *Psychological Bulletin* announce interest in publishing Replication and Reproducibility: Questions Asked and Answered via Research Synthesis. Papers may be published as a special issue or as a series of special sections. Manuscripts will be evaluated and may be published as they come in but must be submitted by July 1, 2018 for consideration for the special sections/issue.

We invite authors to submit single papers that address a clearly defined question related to replication and reproducibility of psychological findings using systematic-review or meta-analytic perspectives. Purely methodological papers, such as tests or criticisms of particular approaches, or commentaries on replication and reproducibility, are not appropriate for *Psychological Bulletin*. Primary-level research reports are also outside of the interest domain of *Psychological Bulletin*. Examples of possible manuscripts include but are not limited to meta-analyzing replication efforts, comparing evidence from different replication methods, synthesizing models of replication determinants, meta-analyzing predictors of questionable research practices and retraction, and comparing methods to gauge publication bias.

For guidance in proposing a submission, interested authors may contact the Editor, Dolores Albarracín (dalbarra@illinois.edu), or Associate Editor, Blair T. Johnson (blair.t.johnson@uconn.edu). Alternatively, authors are free to submit unsolicited manuscripts and should indicate that the submission targets this call in the cover letter.