

Methods for a fully online automated cognitive training study on Prolific

Freya Joessel^{a,b,*}, Sylvie Denking^{a,b}, Paul-Emile Joessel^c, C. Shawn Green^d,
Daphne Bavelier^{a,b,*}

^a Faculté de Psychologie et Sciences de L'Education, Boulevard du Pont d'Arve, 40, (FPSE), Université de Genève, 1205, Geneva, Switzerland

^b Campus Biotech, Chemin des Mines, 9, 1202, Geneva, Switzerland

^c Schoffelgasse 4, 8001, Zurich, Switzerland

^d Department of Psychology, University of Wisconsin-Madison, 1202 W Johnson St, Madison, WI, 53706, United States

ARTICLE INFO

Keywords:

Cognitive training
Intervention studies
Online assessment
Methods

ABSTRACT

Intervention studies are the current gold standard when investigating the causal link between an intervention (e.g., physical activity, cognitive training, meditation, action video games) and its impact on cognitive functions. Such studies are resource intensive, especially when conducted to the latest standards in the field. Recently, it has been noted that the development of online tools to conduct such studies may significantly reduce resource demands, and thus allow more of these acutely necessary studies to be carried out. Here we present a series of tools to conduct intervention studies in a fully online fashion such that participants may go through the entire experimental pipeline without any contact with the experimenters. In particular this included Prolific for participant recruitment and management, the implementation of a pseudo-randomized group assignment procedure such that groups are matched at pre-test, and the development of various dashboards for experimenters and participants to follow their progression throughout the pipeline. These tools were implemented in a 12-h mechanistic cognitive training study where participants completed the training and pre- and post-test assessments remotely over multiple weeks. This new digital pipeline allowed us to limit the resource demands, implement strong masking practices, recruit a sample more diverse than the usual WEIRD in-laboratory samples, and complete the study in less time than usually needed.

1. Introduction

Over the past several decades, a massive increase in our understanding of the basic scientific principles underlying neuroplasticity, paired with a concomitant increase in our knowledge of how to utilize such principles in behavioral training paradigms (Deveau et al., 2015; Raviv et al., 2022), has provided optimism that tools for cognitive enhancement may become routinely deployed. This in turn has resulted in significant interest in methods to most convincingly demonstrate the efficacy of such tools.

Intervention studies, which are the gold standard when it comes to establishing causal effects of all forms of behavioral training throughout

the psychological sciences, including in cognitive training, follow similar broad procedures as many randomized controlled trials in the medical domain (Friedman et al., 2015; Houle, 2015). More precisely, in the cognitive training domain (Green et al., 2019), a group of participants is first tested on a series of cognitive measures (pre-test/baseline). They are then typically randomly assigned to one of two types of interventions: (i) cognitively demanding training (active group) or (ii) some type of control experience. In methodologically stronger studies, the control experience involves training that is matched with the active training along as many dimensions as possible, while still having significantly reduced cognitive demands (active control).² In less methodologically strong studies, the control experience might instead

* Corresponding authors at: Faculté de Psychologie et Sciences de L'Education, Boulevard du Pont d'Arve, 40, (FPSE), Université de Genève, 1205, Geneva, Switzerland.

E-mail addresses: joessel@wisc.edu (F. Joessel), sylvie.denking@unige.ch (S. Denking), cshawn.green@wisc.edu (C.S. Green), daphne.bavelier@unige.ch (D. Bavelier).

¹ Present address: Department of Psychology, University of Wisconsin-Madison, 1202 W Johnson St, Madison, WI, 53706, United States.

² We recognize that in fields where a treatment already exists, the strongest control could be a "treatment-as-usual" group (e.g., for medical interventions). In the field of cognitive training, though, there is no commonly accepted intervention that could be considered to be "treatment-as-usual". As such, in this field, the "standard" treatment is just business-as-usual (Green et al., 2019).

simply be no training at all (such a group-type goes by multiple labels depending on the field, including passive control, test-retest control, no-contact control, or business-as-usual control). After pre-testing, participants then undergo their given training experiences. In most work in the field, this has involved between a few hours (Ainsworth et al., 2013; Lilienthal et al., 2013; Moreau et al., 2017; Prins et al., 2011) to tens of hours of training or more (Bostock et al., 2019; Schmiedek, 2010; Zhang et al., 2021), with larger effects typically being observed with longer durations of training (Bediou et al., 2023; Jaeggi et al., 2008; Ludyga et al., 2020; Osman et al., 2023; Schwaighofer et al., 2015; Verhaeghen, 2021). Finally, following training, participants are tested again (post-test) on the same set of tasks as at pre-test. The critical question is then whether the participants in the active group showed larger improvements from pre-test to post-test than the control group(s) on the cognitive tasks of interest.

Using this kind of intervention paradigm, results suggest that it may be possible for cognitive training to enhance a broad range of cognitive abilities. For example, different kinds of computer-based training paradigms that challenge executive functions and/or attentional control have been shown to lead to improvements in untrained aspects of working memory (Anguera et al., 2012; Berger et al., 2020; Brehmer et al., 2012; Jaeggi et al., 2008; Jaeggi, Buschkuhl, et al., 2010; Schweizer et al., 2011), visuo-spatial skills (Adams et al., 2016; Feng et al., 2007; Stepankova et al., 2014), fluid intelligence (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010; Jaušovec & Jaušovec, 2012; Schmiedek, 2010; Schweizer et al., 2011), speed of processing (Ball et al., 2013; Rebok et al., 2014), visual attention (Belchior et al., 2013; Green & Bavelier, 2003; Strobach et al., 2012; Wu & Spence, 2013; Zhang et al., 2021), visual working memory (Blacker & Curby, 2013; Novak & Tassell, 2015; Oei & Patterson, 2013), and cognitive flexibility (Green et al., 2012; Hutchinson et al., 2016; Strobach et al., 2012). This is not limited to computerized cognitive training but extends to practicing focused attention meditation (in person or via apps), shown to improve working memory and inhibition (see Verhaeghen, 2021 for a meta-analysis) or practicing physical exercise, which has been linked to better inhibitory control (Chuang et al., 2015; Maillot et al., 2012; Stroth et al., 2010), working memory (Moreau et al., 2015), and executive functions (Fedewa & Ahn, 2011; Hillman et al., 2005; Prakash et al., 2015; Sibley & Etnier, 2003). While the studies above have largely focused on enhancing executive functions in healthy participants, other studies examine possible impact beyond executive functions and beyond healthy populations. Promising results include reducing ADHD symptoms in children (Bigorra et al., 2016; Klingberg et al., 2005, 2002; Mishra et al., 2016), improving language and math abilities (Franceschini et al., 2013; Libertus et al., 2017; Loosli et al., 2012; Novak & Tassell, 2015; Pasqualotto et al., 2022; Wang et al., 2019), enhancing vision in those with and without vision-impairments (Green & Bavelier, 2007; R. Li et al., 2009; Nyquist et al., 2016; R. W. Li et al., 2011; Vedamurthy et al., 2015), decreasing stress and depression while increasing psychological well-being (Hale et al., 2021; Marquez et al., 2020; Netz et al., 2005; Verhaeghen, 2017) and finally augmenting quality of life in older adults (Brehmer et al., 2012; Gavelin et al., 2021; Hou et al., 2020; Kelly et al., 2014; Nguyen et al., 2019; Rebok et al., 2014).

However, while a substantial number of both empirical papers and meta-analyses of the empirical literature have supported the idea that various cognitive capacities could be augmented via some forms of computerized behavioral training, it is worth noting that other intervention studies with similar approaches as well as some meta-analyses have failed to find such effects (Abdin et al., 2018; Brehmer et al., 2012; Chooi & Thompson, 2012; Kelly et al., 2014; Melby-Lervåg et al., 2016; Moreno-Peral et al., 2022; Owens et al., 2013; Redick et al., 2013; Sala et al., 2019; Sink et al., 2015; Thompson et al., 2013; Verhaeghen, 2021; Von Bastian et al., 2013; Young et al., 2015). Importantly, to resolve these discrepancies, essentially all reviews and meta-analyses have identified a similar set of seven limitations across the field of

cognitive training (Green et al., 2019).

Prevalence of studies without proper control training conditions. As noted by a number of meta-analyses focused on various types of behavioral interventions to enhance cognitive function, it is not particularly unusual for such designs to fail to include an active control group (working memory training: 20 out of 30 (Sala et al., 2019), 52/106 (Melby-Lervåg et al., 2016), 57/112 (Weicker et al., 2016), 12/24 (Au et al., 2015); Action Video Games: 22/33 (Sala et al., 2018); aerobic exercise: 22/80 (Ludyga et al., 2020), commercially-available computerized training: 8/21 (Tetlow & Edwards, 2017), meditation: 26/46 (Verhaeghen, 2021)). While business-as-usual control groups allow researchers to control for test-retest effects, they are not an appropriate controls for other factors that may contribute to pre- to post-test improvements such as motivation (Greene & Miller, 1996; Honicke & Broadbent, 2016; Walker et al., 2006), expectations (Boot et al., 2013; Denking et al., 2021; Foroughi et al., 2016; Parong et al., 2022), or experimenter effects (Mayo, 1993). Thus, in the absence of an appropriate active control group, it is not possible to unambiguously attribute pre- to post-test effects to the intervention and its presumed active components.

Small sample size studies (i.e. underpowered studies). Given the expected effect sizes in cognitive training (about Hedge's $G = 0.3$), power analyses at a power of 0.8 yield a minimum of 25 participants per group, and 41 with a power of 0.95 (computed with G*Power, Faul et al., 2009). In recent meta-analyses of interventions to improve cognition, 66 % of included studies had fewer than 25 participants per group and thus were likely underpowered (working memory training: 25 out of 30 (Sala et al., 2019), 139 out of 216 (Melby-Lervåg et al., 2016), 87 out of 112 (Weicker et al., 2016); Action Video Games: 22 out of 22 (Bediou et al., 2018), 24 out of 33 (Sala et al., 2018); aerobic exercise: 21 out of 51 (Ludyga et al., 2020), commercially-available computerized training: 7 out of 21 (Tetlow & Edwards, 2017), meditation: 27 out of 46 (Verhaeghen, 2021)).

Paucity of studied with longer training durations. While 10 h of training have sometimes been reported to be sufficient to induce plastic changes in cognitive abilities following training (Feng et al., 2007; Haimov & Shatil, 2013; Jaeggi et al., 2008; Lilienthal et al., 2013; Moreau et al., 2015; Schmidt et al., 2015; Stepankova et al., 2014; Vance et al., 2007; Wu & Spence, 2013), most empirical evidence (as well as essentially all theory) points toward a dose-response effect between training duration and size of training effects (Bediou et al., 2018; Chen et al., 2020; Jaeggi et al., 2008; Osman et al., 2023; Stepankova et al., 2014; Verhaeghen, 2021; Weicker et al., 2016). In recent meta-analyses of interventions to improve cognition, 48 % of included studies had a training duration of 10 h or less which may explain the mixed results observed so far in the literature (working memory training: 25 out of 30 (Sala et al., 2019), 60 out of 106 (Melby-Lervåg et al., 2016), 57 out of 112 (Weicker et al., 2016); Action Video Games: 9 out of 22 (Bediou et al., 2018), 16 out of 33 (Sala et al., 2018); aerobic exercise: 10 out of 51 (Ludyga et al., 2020), meditation: 24 out of 41 (Verhaeghen, 2021)).

Non-systematic use of masking practices³. The preferred design when evaluating the impact of an intervention are intervention studies where both the participant and the experimenter are unaware of the condition to which the participant is assigned. In the case of cognitive training however, it is not possible to fully mask condition to the participants as they will necessarily be cognizant of what their assigned training entails (i.e., as compared to a drug study, which could use two pills that look identical but only one of them contains the active substance). In this field, the focus is thus on masking the participant to the experimenter intent rather than to the intervention content per se, as well as masking

³ We note that this has historically been called "blinding". However, because this practice does not involve literally making the participants or experimenter unable to see (and thus could be considered ableist), we use terms such as "masking" or "unaware," which are more accurate descriptions of the intention of experimental procedures (see Morris et al., 2007).

experimenters to the condition of the participant. As discussed in Green et al. (2019), this can potentially be achieved by presenting all arms of the intervention study to participants as active (noting that this is not necessarily always trivial given that the content of various experiences may naturally produce different expectations in participants) and by masking the experimenters involved in participants' skill evaluation.

Homogeneity of sample populations. A major concern in the field of psychology is the over-reliance of behavioral research on WEIRD samples (White, Educated, Industrialized, Rich, and Democratic) that are not necessarily representative of the average human population (Henrich et al., 2010). As such, many results from behavioral psychology and neuroscience may not be reproducible when the study includes a more heterogeneous sample (Dotson & Duarte, 2020). Including more heterogeneous samples would provide the much needed opportunity to develop more complete and nuanced theories of cognitive mechanisms (Bryan et al., 2021).

Potential mismatch of groups at pre-test. While random group assignment is often considered the gold standard for intervention studies, purely random group assignments may lead to differences between group at baseline/pre-test. This possibility becomes even more likely when sample size is small (Bruhn & McKenzie, 2023). Unfortunately, there is no statistical way to fully and convincingly control for pre-test differences between groups once present (Green et al., 2013; Miller & Chapman, 2001). Methods such as stratification (also known as blocking or pairing) (Adelman, 1969; Feldt, 1958) or minimization (Pocock & Simon, 1975; Taves, 1974), can be used to minimize variance at pre-test between the groups of the intervention. However, stratification methods usually handle matching across more than a couple of variables, or matching on variables for which the distribution is not well-known poorly. Stratification and minimization methods usually require that several or even all the participants be recruited before they can begin the experiment, and some methods (such as pairing) will lead to some participants being discarded from the final sample as no match could be found for them. These methods can thus be impractical for studies as resource-intensive as intervention studies.

Lack of systematic business-as-usual group as a baseline. Intervention studies testing cognitive training interventions should include a business-as-usual control group in addition to an active control group(s) (Green et al., 2019). For example in the case of mechanistic studies where the goal is to investigate which features of an intervention leads to observed cognitive enhancements, a business-as-usual control group helps in the interpretation if the result is an absence of difference in cognitive changes between the experimental and active control group (i. e., to arbitrate whether “neither group” improved beyond test-retest effects or “both groups” improved by the same amount). In efficacy studies where the goal of the control groups is to control for external confounds (e.g. expectations, development, etc.), such a business-as-usual control group allows researchers to disentangle cases where the control groups may not have controlled well enough for these external confounds. In addition, a business-as-usual control group also allows researchers to ensure that interventions do not cause any harm to the target population, especially when the intervention replaces other crucial activities for development or health. Finally, including a business-as-usual control group in all intervention studies allows for stronger meta-analyses. Indeed, with a business-as-usual control, all intervention studies would share that similar point of comparison, and thus effect-sizes could be accurately compared.

Addressing the issues listed above comes with a host of obvious costs – in terms of funds, staff hours, participant hours, etc. These are no small concerns, as intervention studies already require substantial amounts of staff time and financial resources. For example, adding an active control condition (compared to a business-as-usual control) means doubling both the staff in charge of supervising participants during training and the participant payments (since now twice as many participants are completing training hours that need to be compensated). Enrolling more participants, or increasing the duration of training, both mean either

having a study run for longer (thus running the risk of polluting the sample with cohort effects), or supervising more participants simultaneously to keep the study within a sensible time frame. Recruiting more diverse participants, such as reaching out to non-WEIRD communities, also comes at a cost of having the lab staff scout outside of the usual convenience samples to ensure greater diversity of the subject pool. Finally, better masking practices whereby the experimenters testing participants are unaware of the participant group assignment doubles the staff needed to support the study: one half of the team will be in charge of monitoring participants during training, allowing the other half to remain unaware of group assignment as they take care of pre-test and post-test assessments. Meeting these new methodological demands therefore comes at a cost and runs the risk of being too burdensome for some laboratories or communities.

The present work considers the application of new methodological advances in online data collection to help alleviate the limitations considered above. In particular, we focus on studies conducted on samples of healthy young adults, and we recognize that how these limitations can be best addressed will depend on the specific needs of the target population (e.g. older adults, patients, kids, etc...). While online data collection in some domains is not new (e.g., market research/questionnaire studies have enrolled participants in this way since the early 90s (Krosnick, 1999; Rand, 2012; Van Selm & Jankowski, 2006), it certainly was not a typical approach in most of the social sciences prior to the past decade or so (Sassenberg & Ditrich, 2019). For instance, in four large social psychology journals, the proportion of studies with data collected online increased from 10 % in 2009 to 50 % of the studies in 2018, with a large percentage of online studies relying on self-report questionnaires (50 % in 2009 up to more than 65 % in 2018) rather than the computer-based task assessments required to objectively measure constructs such as perceptual, attentional or other cognitive skills. However, the development of Python (e.g., PsychoPy: Peirce et al., 2019) and Javascript (e.g., PsychoPy's counterpart PsychoJs, and jsPsych: De Leeuw et al., 2023) libraries with reliable display durations for stimuli and reaction time measurements (Anwyl-Irvine et al., 2021; Bridges et al., 2020; Hilbig, 2016) has allowed psychologists and neuroscientists to develop robust cognitive assessments that can be deployed online. This effort has also been supported by the advent of platforms such as Pavlovia (pavlovia.org), Gorilla (Anwyl-Irvine et al., 2020), or Millisecond (Inquisit 5, 2016) that handle all the backend necessary to host cognitive tasks and online participant recruitment platforms, such as Prolific and MTurk, that allow researchers to reach large numbers of participants even beyond their own university or surrounding community. However, most online studies are limited to a single session of data collection (which may include a series of tasks, some questionnaires, or a combination of both), and to date no fully online intervention study (i.e., without any contact between experimenter participants) has been completed to our knowledge.

Designing fully-online intervention studies requires the development of seamless experimental pipelines from the point of view of the participant. In other words, assuming that a participant does not face any technical issues, they should not have to interact with an experimenter at any time during the experiment, and the experiment should feel like a coherent complete experience as termed in the field of user experience design (Nielsen, 1994; Plass et al., 2019). Such a paradigm would remove the limits imposed by the number of experimenters available at any given time to supervise training/assessment, and with proper tools, the demand in human resources needed to manage the same number of participants would be much lower than with typical supervised assessment. Participants would thus be able to complete their sessions independently of experimenter availability (freeing precious human-hours) and hardware availability (such that several participants can complete the tasks simultaneously). In addition, providing seamless transitions between the different parts of the experiment (for example recruitment, pre-test assessments, training times or post-test assessments) would ensure that the participant does not feel any barrier to

their progression in the study. Such barriers include having to re-enter an infrequent subject code they may have forgotten, going to an obscure link, or waiting on an experimenter to unlock the next step. On the researcher side, this may be achieved by creating a user-friendly automatic pipeline in which the experimenter inputs are limited to a minimum.

There currently exists no online platform that allows for a completely automatic implementation of a full intervention study with a training period several weeks long and multi-day pre- and post-test assessments. In this paper, we present the technical methods we used to conduct a fully online, nearly fully automatic 12-h intervention study as a tutorial on how to implement such complex intervention studies. The protocol of the study involved a pre-test assessment where participants completed questionnaires and cognitive tasks for about an hour (pre-test), after which they were assigned to one of five groups depending on their sex, age, and visual attention abilities at pre-test. As the design of this study best fit the “Basic science” category from Green et al. (2019) - where various arms are used to examine specific mechanics - this study included one active intervention group, three other active control groups controlling for different aspects of the training, and a business-as-usual group controlling for test-retest effects (i.e. rather than the simpler two-arm: experimental and active control group efficacy design). The participants then trained for 12 h (distributed over a 4 to 6 weeks period) on a gamified cognitive task embedded within a custom-made video game (Training), and finally, they completed questionnaires and cognitive tasks over 4 days (post-test).

Below we detail how to use Prolific for intervention studies, how to manage multi-day pre-test and/or post-test assessments, and present the tools used by the researchers and the participants at each step of the experiment to create a seamless experience for the participant with no contact with the experimenters. The tools that were implemented for each of the successive steps of the study, in each case from the point of view of the participant and then of the experimenter.

2. Methods

As part of the study, participants (typical healthy young adults) successively completed four parts: recruitment, pre-test, training, and post-test. Each part was further divided into daily sessions. Given our focus here was on general methodological advances, we do not fully describe the specific tasks or training paradigms below. Instead, this study should be treated as a standard pre-test → training → post-test design. For the participant, recruitment involved responding to an ad on Prolific, responding to a filtering question, and agreeing to the

consent form. Pre-test was composed of a set of pre-test questionnaires and a battery of cognitive tasks; with recruitment and pre-test all being completed in one session. The training consisted of completing the 12 h on the training game within the allotted time period of 4–6 weeks; this occurred entirely outside of Prolific. Finally, the post-test brought participants back within Prolific and involved completing questionnaires as well as a battery of tasks spread over four days, with one session per day. Fig. 1 provides a succinct overview of the study design. This study was reviewed and approved by the Commission Cantonale d’Ethique et de Recherche in Geneva, Switzerland, with the approval number 2018-01669, dated January 10th, 2019.

Participants were paid for any and all time they spent in the study, even if they chose to drop out early, following standard practices and research ethics in the field of behavioral intervention studies. In addition, and also in line with the incentivization often practiced in long intervention studies, participants received a bonus payment upon completing all their assigned stages of the study. Participants were informed of this payment structure in the ad for the study, and it was repeated during the consent process. To automatize the pipeline, we ensured that (1) the enrollment in the study, (2) the questionnaires and cognitive batteries, (3) the training games, and (4) the transitions between the various steps of the pipeline could all be completed without researcher intervention. Here, a step refers to each component of the pipeline as presented in Figs. 2, 3, and 4.

Given the constraints that using Prolific imposed on our design, we will first present how we used Prolific for such a multi-day intervention study before explaining how each step of our design was implemented online.

2.1. How to use Prolific for studies with many sessions

When a participant enters a study on Prolific, Prolific expects the participant to signal completion of the study within a time window consistent with the duration of the study advertised by the experimenter. In addition, Prolific does not allow for compensation lower than the minimal wage (set at 6.00€/hour at the time our study was running), meaning that for an intervention study spanning several weeks, one cannot just open a study on Prolific that the participant would enter at pre-test and close after completion of post-test.

As such, the study needed to be divided into “sessions” that could each be completed in one sitting, each one of them being linked to an individual “Prolific study” to be marked as completed when participants finished the corresponding session. To avoid any ambiguity, in the context of the methods, we will refer to the studies created on Prolific

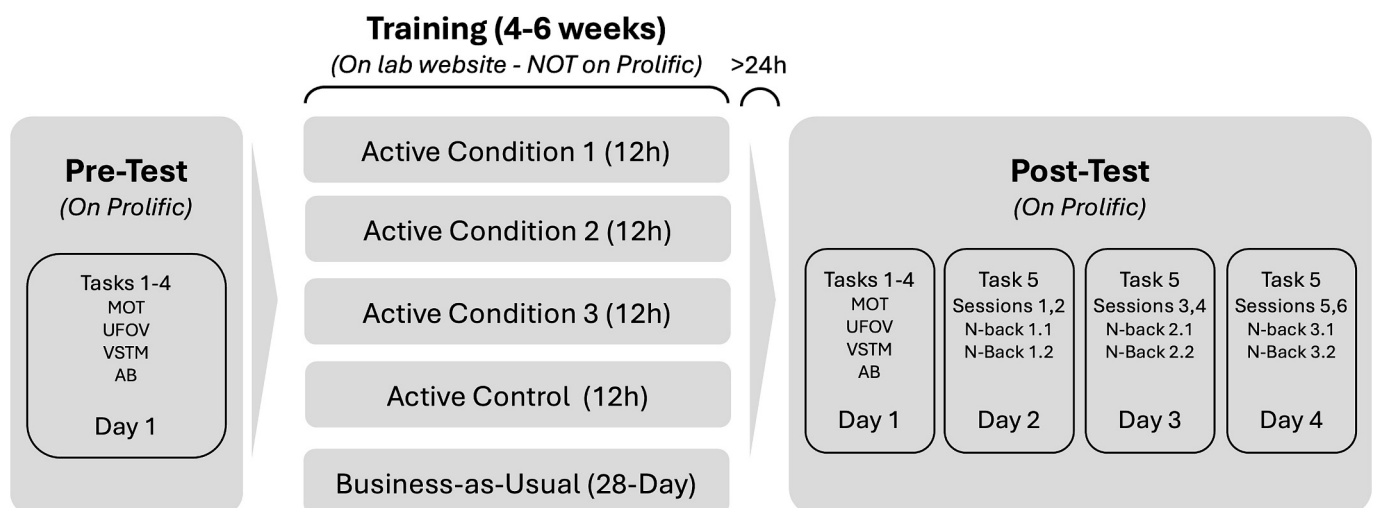


Fig. 1. Overview of the design of the study. Abbreviations: MOT (Multiple Object Tracking task), UFOV (Useful Field of View task), VSTM (Visual Short-Term Memory task), AB (Attentional Blink task), N-back (N-Back task).

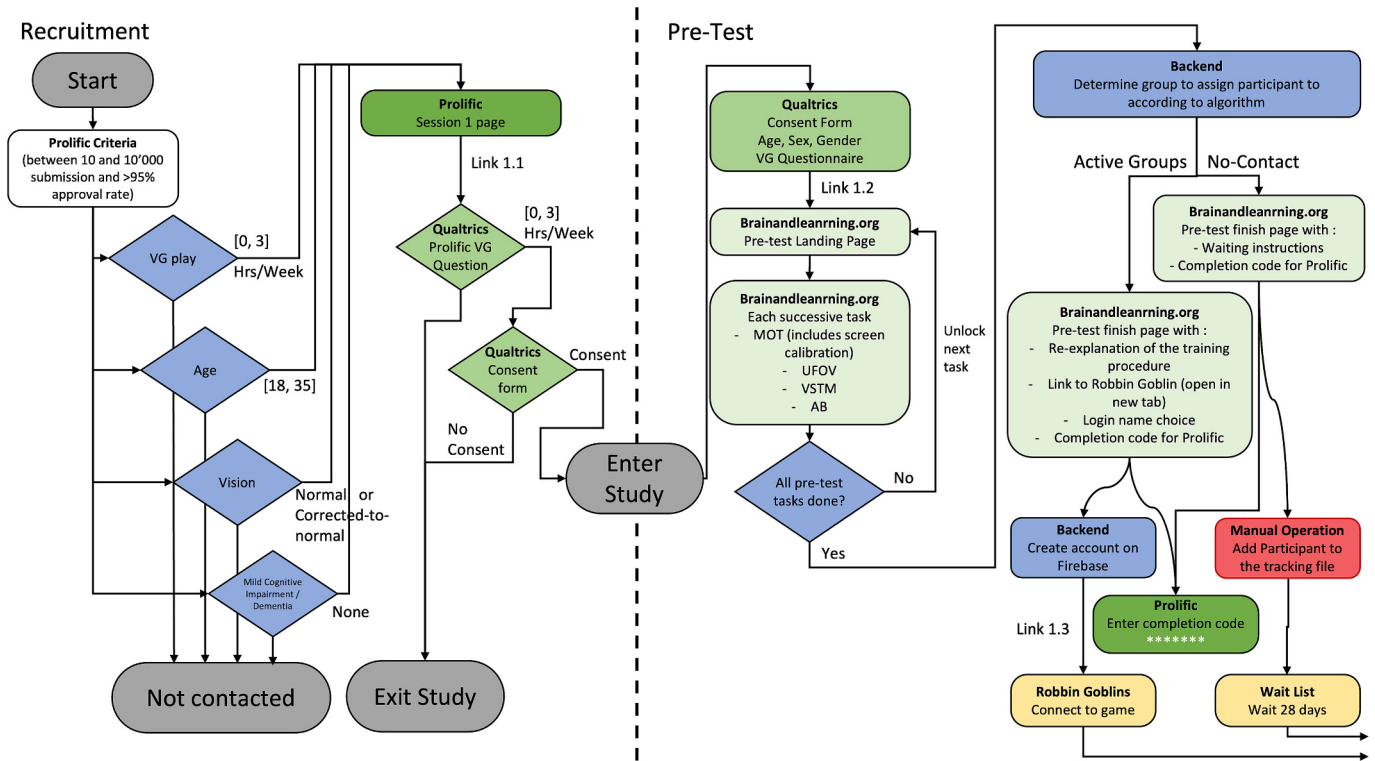


Fig. 2. Workflow of the steps in the recruitment and pre-test parts of our experimental pipeline. Green boxes and diamonds correspond to steps where the participants interacted with our pipeline; the location of this interaction is given in the title of each box (front end development). Different shades of green represent different websites, dark green for Prolific, medium green for Qualtrics, and light green for our lab website. Blue boxes (backend) represent decisions/actions points that were automatically handled by our pipeline (i.e. back-end development). Red boxes are decisions and actions that needed to be taken by one of the experimenters, all other steps were fully automated. Yellow boxes represent steps pertaining to the training game or business-as-usual group. Grey boxes represent entry/exit points to the study. Abbreviations: VG (Video Game), MOT (Multiple Object Tracking task), UFOV (Useful Field of View task), VSTM (Visual Short-Term Memory task), AB (Attentional Blink task). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that participants can register to as “Prolific studies”, and we will refer to the ensemble of tasks a participant was asked to complete on a given day as “session”. In our case, we thus created 5 different Prolific studies, one for the pre-test session and four for the 4 sessions of post-test (one per day). The training sessions were not carried within Prolific in order to allow the participants to follow their own schedule as long as they completed the training within the allotted time window (see also Fig. 1).

To recruit participants, the pre-test session was advertised on Prolific to eligible participants. The advertisement for the pre-test session detailed the whole procedure: pre-test, training, post-test, the duration of each session and each part, the expected compensation for completing each part (including the bonus reward for overall study completion), and the schedule that participants were expected to follow. However, since from the point of view of Prolific, this session only concerned the pre-test, we marked it as completed for the participants who completed the pre-test assessments.

For the post-test which included 4 sessions and one-session-per-day structure, we used the “allow list” feature of Prolific studies. This allowed us to open access to a given session only to the participants who had completed the previous session the day before at the latest. Since this procedure could not be done automatically on Prolific, an experimenter connected daily to Prolific and added the IDs of the participants who had completed a session the previous day to the “allow list” of the Prolific study for the next session.

Finally, regarding training, we built an online pipeline in which participants could complete their daily training separate from Prolific as using Prolific would have required too many experimenter manipulations and would remove a lot of the flexibility in the training schedule. Participants, at the end of the pre-test, were instructed that during their training period they should just come back to the webpage where the

game was hosted and do so until completion of the training. Upon training completion, they were instructed that the Prolific study for the first day of post-test would be open to them after a 24-h wait period.

2.2. Keeping track of the participant's “identity” throughout the pipeline

Our study used tools and frameworks that were not natively linked to each other, either because they did not live on the same server (e.g., Prolific, Qualtrics, our own server) or because they were designed years apart by different programmers (for example the cognitive tasks and the training game). To keep track of the participant's data on each of these various tools we used various methods to pass the participant's de-identified Prolific ID from one tool to the next.

2.2.1. On Prolific (March–August 2022)

We took advantage of Prolific's feature where the participant's Prolific ID can be embedded within the URL of the study (*Prolific_PID* parameter). This ID can thus be passed from Prolific to the next step such as to Qualtrics (links 1.1 in Figs. 2 and 3.1 in Fig. 4) or to our lab server (link 3.2 in Fig. 4).

2.2.2. On Qualtrics

Here we used the embedded data to capture the ID of the participant from the URL and store it while the participant completed the questionnaires. At the end of the questionnaires, the participants were directed to the next steps of our pipeline by creating a link that automatically included their ID (links 1.2 in Fig. 1 and 3.2 in Fig. 4).

2.2.3. On our server

Here we captured the Prolific ID passed to the server in the URL using

Training workflow

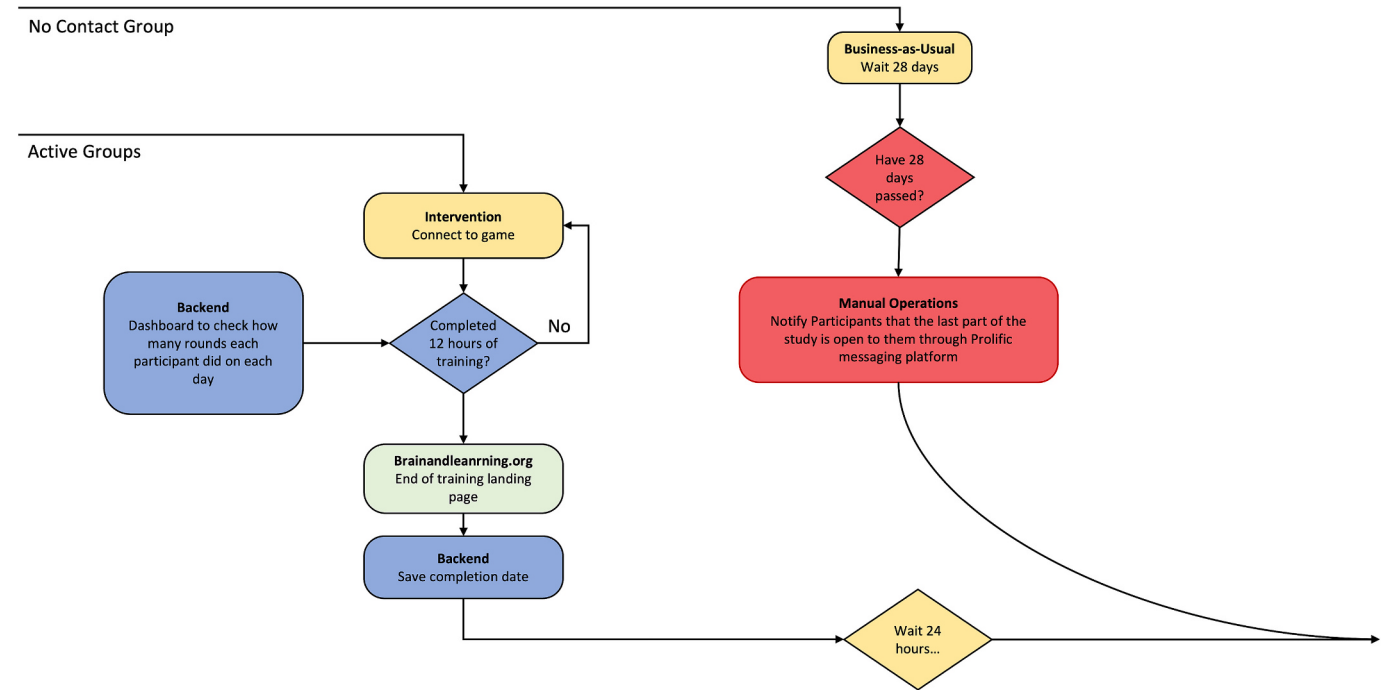


Fig. 3. Workflow of the steps in the training part of our experimental pipeline. Green boxes correspond to steps where the participants interacted with our pipeline. Blue boxes (backend) represent decisions/actions points that were automatically handled by our pipeline. Red boxes are decisions and actions that needed to be taken by one of the experimenters, all other steps were fully automated. Yellow boxes represent steps pertaining to the training game or business-as-usual group. Grey boxes represent entry/exit points to the study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

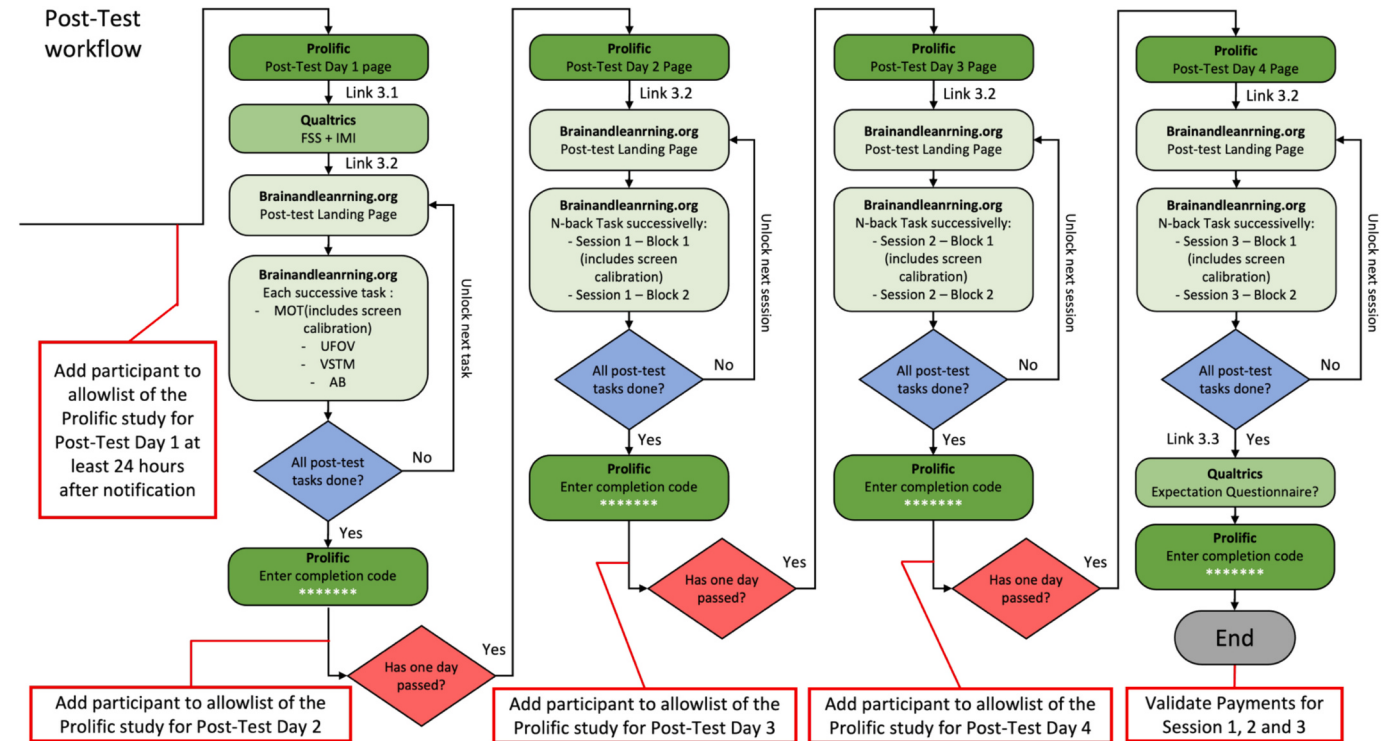


Fig. 4. Workflow of the steps in the four sessions of the post-test part of our experimental pipeline. Green boxes and diamonds correspond to steps where the participants interact with our pipeline. Blue boxes (backend) represent decisions/actions points that were automatically handled by our pipeline. Red boxes and text surrounded in thick red line are decisions and actions that needed to be taken by one of the experimenters, all other steps were fully automated. Grey boxes represent entry/exit points to the study. Abbreviations: FSS (Flow State Scale), IMI (Intrinsic Motivation Inventory), MOT (Multiple Object Tracking task), UFOV (Useful Field of View task), VSTM (Visual Short-Term Memory task), AB (Attentional Blink task). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

PHP and stored in our database to track the progress of each participant through the pipeline such that participants could pick up the tasks where they last left off.

By passing the Prolific ID of participants through the URL, the participants never re-entered their Prolific ID (a chain of 24 random alphanumeric characters), which minimized the risks of mismatch in databases and eased the flow of the experiment by diminishing the number of participant actions as they transitioned from one activity to the next.

2.3. Recruitment and consent

The handling of participant recruitment and consent is first described from the point of view of the participants and then from that of the experimenters, detailing the tools that were developed to keep track of participants as they entered the study.

2.3.1. Participant side

Recruitment was done similarly to most online studies using web-based platforms whereby participants who matched our Prolific criteria responded to an advertisement on the platform. The advertisement detailed the time commitments and schedule of the study in hopes that only participants who could truly commit to the study would sign up for it. Given the nature of the study, we also paid special attention write the ad in a way that would keep participants as masked as possible to the true purpose of the study.

As Prolific rules prevent researchers from excluding participants from a study based on criteria other than the ones that Prolific provides, our inclusion criteria on Prolific enforced that our study was advertised only to participants who had self-reported playing video games less than 3 h/week on their Prolific profile. Recognizing that this habit might have changed since the moment the participant answered this question for their Prolific profile, the participants were asked that same question right after agreeing to take part in the study via Prolific. Participants whose response did not fit our criterion (≤ 3 h/week of game play) were asked to exit the study and “return their submission” on Prolific (i.e., voluntarily leave the study on Prolific). The others were directed to a page with the consent form from our university's IRB, from where they were directed to the first step of the pre-test after consenting.

2.3.2. Experimenter side

From the experimenters' point of view, a participant table grouped the data necessary for all the parts of the experiments that were not pre-test and post-test. This table included the Prolific ID of the participant, their username to connect to the training intervention program, the information for the allocation algorithm (age, sex, and MOT score, see Section 2.5 Randomization of Group), an inclusion Boolean for the randomization algorithm, their assigned group, and the date at which they finished the training game (to set up the 24 h wait period between training and post-test for the active groups).

2.4. Pre-test and post-test tasks

Because the methods implemented to handle pre-test and post-test are similar to each other, both of these steps will be described in this section. The methods described here are agnostic of the specific tasks implemented in the study. We describe the experience from the participants' point of view first, and then from the experimenters' point of view.

2.4.1. Participant side

One key aspect of the user experience that we strived for was to make sure that at all times the participant was aware of their progress within the experimental pipeline, their current task, what they had already done, and which steps they had yet to complete. Given that we wanted participants to be autonomous within our pipeline (i.e., that they could

complete it without researcher intervention), we created dashboards for our pre-test and post-test assessments (see Fig. 5 for a screenshot of the post-test dashboard) that the participants accessed when they started each successive session. These dashboards detailed the deadline by which participants should complete each task, what to do once all tasks were completed, and importantly, the link to each successive task was unlocked only when the previous task was completed, ensuring that all the participants completed the tasks in the same order. In addition, even if we emphasized that the pre- and post-test sessions should be completed in one seating, the dashboard was set up such that the participants could come back to the dashboard and easily pick up where they left off, limiting attrition due to participants being unable to complete their session because of scheduling conflict or unexpected events (e.g. technical issues).

The dashboard was particularly essential for our post-test, as one of the tasks required the participants to complete three 40-min sessions over the course of three separate days spread over no more than 10 days. In this case, we indicated on the dashboard when each session could be completed at the latest such that the participants could plan accordingly to complete all the sessions within the allotted time. Regarding the cognitive tasks themselves, we used tasks coded in Javascript using the Canvas API (Yung et al., 2015) or a custom, more flexible, framework. All tasks were coded such that the current experimental stage (pre-test or post-test) could be passed as input through the URL so that the data would be saved in the right table, and the participant's Prolific ID was recovered from the current PHP session, and identifiable by the experimenters. Note that the tables were designed such that masked experimenters could not become unmasked. Finally, since the participant was expected to perform these tasks without supervision, we introduced each task with a detailed tutorial (as presented in Yung et al., 2015), with the controls repeated regularly throughout the task.

2.4.2. Experimenter side

While most of the management of the pre-test and post-test batteries was automated, one manual action was still required to manage the multi-day post-test battery. Namely, each day, we needed to open the successive session to the participants who completed the previous session the day (at least 24 h) before. To do so, we used the *allowlist* function of Prolific studies that allows the experimenter to open the study to a specific set of participants identified by their Prolific ID, and new participants may be added after the Prolific study was started. Unfortunately, at the time of this study, new Prolific IDs could only be added manually to this *allowlist*, such that there was no way to automatically add the Prolific IDs of participants who had completed session N-1 to the *allowlist* of the Prolific study for session N. To remedy this, we had one experimenter connect to Prolific every day at a specific time in the morning to add to the *allowlists* of the Prolific studies for sessions 2, 3, and 4 the Prolific IDs of the participants who had completed sessions 1, 2, 3 of our four-day post-test assessment respectively.

Within each session of the pre-test and post-test batteries however, we automatically kept track of the participant's progress such that the participant could only do the task in the order we had specified, and they were at all times informed of when they would be able to complete the following task. Technically, this was done using two tables in our SQL database that were accessed through phpMyAdmin, the *Pre-Test Table* and the *Post-Test Table* (see Fig. 6 for a snapshot of the Post-Test Table). These tables were set up such that they could serve as a dashboard “as is”. They shared a similar structure, but were adapted to the specifics of our pre- and post-test structure. In both tables, we recorded the Prolific ID, and the dates at which the participant started the pre-test, the post-test, and completed each of the individual tasks. For pre-test and day 1 of post-test, task completion was recorded in the following order: Multiple Object Tracking (MOT), Useful Field of View (UFOV), Attentional Blink (AB) and Visual Short-Term Memory (VSTM). For days 2 to 4 of post-test, we recorded 2 sessions of the adaptive N-back task every day. The content of the tasks themselves are beyond the scope of this article,

Part 3 - Dashboard

This is the dashboard from which you can access the tasks to complete for each of the 4 total sessions of part 3. Note that you may only complete one session per day and the last session should be completed **within 10 days** of completion of the first session (so plan accordingly!).

Each session is divided into different tasks. After each task, you will be redirected to this page. Instructions will be given at the beginning of each task, with a small practice to help you get familiar with the task. **Each task should take about 15-20 minutes.** You may take a short break between the tasks (preferably just a few minutes), but **when you start a task, we ask you to fully complete it!** In other words, once you start a task, do not reload the page of the task until you are directed back to this dashboard, otherwise the data we receive for it will be unusable and it may impair the payment process with Prolific. And if possible, please complete all the tasks in a session at the same time (i.e., it's better if you don't do one task in the morning, then the next after lunch, then the rest in the evening).

Note that you can come back to this dashboard any time you want using the link in the address bar of your browser (Please save it in your bookmarks!), but you need to go through Prolific first when you actually want to complete the tasks. At the end of each testing session, you will receive a completion code to enter in the Prolific page.

Session 1 consists of 4 tasks measuring different aspect of cognitive functions (similar to those you did in part 1):

- Cognition 1
- Cognition 2
- Cognition 3
- Cognition 4

Session 2, 3 and 4 consist of 2 sessions of a new cognitive task. At the end of session 4, you will have a short questionnaire to complete.

Once you have completed a session, you will receive a completion code to enter in Prolific. You will then be granted access to the next session. **Please note that there may be a few hours before you see the study listed in the list of studies available to you, and remember you can only complete only one session per day!**

To the extent possible, we would like you to complete the first session of part 3 at about the same time of day as you completed part 1 at the start of the study - that is 14:04 GMT. Please be aware that your time zone might not be GMT, adjust the time accordingly!

If this were to prove too inconvenient for you, then it is more important to actually complete the sessions than to respect these timings.

Session 1

Cognition 1

completed

Cognition 2

completed

Cognition 3

Start

Cognition 4

First finish the previous task

Session 2

Finish the previous session and wait for a day to unlock this session.

Session 3

Finish the previous session and wait for a day to unlock this session.

Session 4

Finish the previous session and wait for a day to unlock this session.

For any questions or concerns that you may have during this session, you can contact us directly through the Prolific message platform.

Fig. 5. Dashboard for Post-Test. In this example, the participant has completed the first two tasks of the first post-test session so the link to the third task is open. The “start” link goes to the following: https://brainandlearning.org/study_name/task?stage=post-test (notice the stage parameter so that the data are saved in the right database), and the link to the fourth task is currently unavailable. Sessions 2, 3, and 4 are currently locked and will automatically be unlocked 24 h after the last task of the previous day is finished. Here, the name of the tasks are hidden to prevent participant bias and expectations as much as possible.

		Session 1					Session 2		Session 3		Session 4	
id	subject	startedAt	mot	ufov	vstm	ab	nback-1.1	nback-1.2	nback-2.1	nback-2.2	nback-3.1	nback-3.2
17		2022-05-01 11:51:31	2022-05-01 12:18:33	2022-05-01 12:28:23	2022-05-01 10:41:19	2022-05-01 10:52:38	2022-05-02 08:51:55	2022-05-02 09:12:30	2022-05-03 11:24:03	2022-05-03 11:43:17	2022-05-04 13:34:29	2022-05-04 13:53:40
18		2022-05-02 21:42:51	2022-05-02 21:03:26	2022-05-02 21:08:57	2022-05-02 20:18:30	2022-05-02 20:27:43	2022-05-03 10:54:31	2022-05-03 11:29:42	2022-05-04 14:09:19	2022-05-04 14:26:52	2022-05-05 14:48:32	2022-05-05 15:11:14
19		2022-05-06 11:28:25	2022-05-06 12:40:55	2022-05-06 12:46:59	2022-05-06 09:55:59	2022-05-06 10:06:03	NULL	NULL	NULL	NULL	NULL	NULL
20		2022-05-07 13:09:34	2022-05-07 12:24:48	2022-05-07 12:29:08	2022-05-07 11:35:15	2022-05-07 11:43:34	2022-05-09 13:11:09	2022-05-09 13:37:36	2022-05-11 21:19:37	2022-05-11 21:38:09	2022-05-12 10:48:41	2022-05-12 11:07:51
21		2022-05-09 23:21:31	2022-05-09 23:35:33	2022-05-09 23:41:43	2022-05-09 21:49:04	2022-05-09 21:57:44	NULL	NULL	NULL	NULL	NULL	NULL
22		2022-05-10 16:34:04	2022-05-10 09:51:10	2022-05-10 10:00:36	2022-05-10 15:17:04	2022-05-10 15:36:58	2022-05-17 17:32:26	2022-05-17 17:55:47	2022-05-18 22:50:16	2022-05-18 23:09:36	2022-05-20 13:35:53	2022-05-20 14:04:17
23		2022-05-10 19:55:11	2022-05-10 20:09:50	2022-05-10 20:18:09	2022-05-10 18:26:34	2022-05-10 18:37:00	2022-05-11 15:48:02	2022-05-11 16:19:40	2022-05-12 19:12:16	2022-05-12 19:40:14	2022-05-15 10:39:44	2022-05-15 11:01:47
24		2022-05-14 14:37:54	2022-05-14 14:52:56	2022-05-14 14:59:04	2022-05-14 13:09:13	2022-05-14 13:19:42	2022-05-17 07:54:03	2022-05-17 08:14:23	2022-05-18 10:33:45	2022-05-18 11:02:25	2022-05-19 13:03:47	2022-05-19 13:31:12
25		2022-05-16 12:53:28	2022-05-16 13:05:17	2022-05-16 13:11:43	2022-05-16 11:19:47	2022-05-16 11:29:19	2022-05-18 09:23:29	2022-05-18 09:59:49	2022-05-19 11:23:31	2022-05-19 11:42:25	2022-05-20 13:03:06	2022-05-20 13:35:40
26		2022-05-19 18:08:25	2022-05-19 11:29:09	2022-05-19 12:33:14	2022-05-19 17:46:11	2022-05-19 18:06:34	2022-05-20 16:38:35	2022-05-20 16:58:38	2022-05-21 18:38:15	2022-05-21 19:12:24	2022-05-23 16:15:52	2022-05-23 17:14:37
27		2022-05-22 12:40:24	2022-05-22 11:55:17	2022-05-22 12:01:17	2022-05-22 11:11:36	2022-05-22 11:21:10	2022-05-23 16:33:09	2022-05-23 16:52:14	2022-05-24 17:04:02	2022-05-24 17:29:33	2022-05-25 13:49:28	2022-05-25 14:07:56
28		2022-05-23 14:25:31	2022-05-23 16:16:59	2022-05-23 16:30:02	2022-05-23 14:41:15	2022-05-23 14:54:09	2022-05-24 13:01:13	2022-05-24 13:27:15	2022-05-25 13:55:38	2022-05-25 14:18:31	2022-05-26 13:39:41	2022-05-26 14:01:50

Fig. 6. Snapshot of the table for the post-test. In this example, the participants with id 19 and 21 completed all the tasks for session 1 of the post-test, but had not completed the tasks for sessions 2, 3 and 4 yet. All other participants had completed all their sessions. Abbreviations: MOT (Multiple Object Tracking task), UFOV (Useful Field of View task), VSTM (Visual Short-Term Memory task), AB (Attentional Blink task), nback (N-Back task).

but a detailed description of the tasks can be found in Joessel (2022). We also recorded the date when the participant finished the pre-test. Additionally, for the post-test, we recorded the dates at which the participants completed each block of the N-back learning task. If a task (or task block) had not been completed yet by a participant, the corresponding cell's value was "NULL". The links to pre-test and post-test tasks were automatically made available to the participant based on the dates in this table such that only the button to the first task starting from the left that had a NULL value was displayed. In the case of post-test, buttons to begin the task for the next session were displayed only if 24 h had passed since the participants had finished the previous task. When the participant could complete a task, they saw a green "Start" button leading them to the task, otherwise a task indicating the status of the task was displayed.

2.5. Randomization of group assignment

Our study aimed to match participants across groups on sex, age, and pre-test performance on one of our pre-test attention tasks (MOT task). For the sex variable, we enforced at recruitment the same number of participants in each sex group by opening two identical studies on Prolific for each wave of recruitment, one targeted only at male participants, and one targeted only at female participants, both with the exact same target number of participants. We only opened a new wave of recruitment when both studies were filled with participants. Given that we did not wait for participants in a given wave of recruitment to be done with the study before opening a new wave of recruitment, we also adjusted the target sample sizes in the subsequent Prolific studies to take into account the number of male and female dropouts in previous waves. For the age and pre-test MOT performance, we used the variance minimization algorithm of Sella et al. (2021). This was done separately for the male and female samples, such that males were matched with males and females matched with females (this is likely better than matching participants across the whole sample directly if distributions by sex are different). In this procedure, each new participant was assigned to the group that minimized the total variance across groups and variables (here age and pre-test performance on the MOT). Our implementation of the algorithm was similar to that of Sella et al. (2021), with the addition that when a participants dropped out of the study (and as such would not be included in the final sample), we also excluded them from the sample used by the algorithm so that they did not influence the assignment of subsequent participants and the randomization procedure only ever considered participants that had not dropped out. While this choice has possible drawbacks for Intention-To-Treat analyses (i.e., if one group sees a greater drop rate than others), it has the advantage of guaranteeing well-matched group at pre-test for analyses. We also implemented the algorithm in JavaScript directly on our

server, such that as soon as the participants finished the pre-test battery, their sex, age, and pre-test performance on the MOT were saved in a table on our server and the group assignment was immediately computed for that participant (instead of having a researcher do that by hand). Thus, from the point of view of participants, at the end of pre-test, they were seamlessly redirected to the instructions relevant for their assigned training group.

2.6. Creating the account for the training game

2.6.1. Participant side

Upon completing the pre-test, participants were automatically redirected to the instruction page for their assigned training group as per the continuous attribution randomization algorithm described above. The study design called for 4 active training groups and one business-as-usual group (see Section 2). The participants in all the active groups were redirected to the same page on our website with instructions regarding the training session schedule they should adhere to and a text-based tutorial for their assigned training game. At this stage, participants were cued to provide an in-game name for their training game. If their in-game-name matched our criteria (alphanumeric characters only, minimum 6 characters long, and not already taken by another user), the participants clicked a button asking them to assert that they properly read and understood their assigned training instructions. Clicking that button revealed the link to the game which had been hidden until then. In the back-end, clicking this button created their account in the training game (see "experimenter" session below), such that the participants could start training on their assigned game right away if they wanted to.

The participants in the business-as-usual group were redirected to a page telling them that they would be re-contacted after a 4-week wait period, when the post-test study on Prolific would be opened for them, and they would be informed about it through the Prolific messaging platform.

2.6.2. Experimenter side

For the in-game name, Prolific IDs were not used, as participants had to connect to the game every day. While the game was set up to remember the participant's credentials using a cookie, some participants might choose to erase their internet data, or play in incognito mode altogether, requiring them to enter their in-game name every day to access the game. As such, we did not want them to use their Prolific ID (a chain of 24 random alphanumeric characters) as it may cause participants to quit the study out of sheer annoyance with the complexity of the ID. Once the participants provided an easy-to-remember in-game name (that we asked them to write down), we stored this in-game name in the participant table on our server to be able to link the task data with the in-game data.

Given the type of data produced by the game, the real-time database, Firebase, was used where data was stored as JSON objects; this database was not stored on our server. The data was structured with one main database where the condition of each participant was stored as key-value pair, with the in-game name as the key and the condition name as value, and four additional databases, one for each training group, that contained the data from the training game. Creating an account required two steps. First, upon creation of the account, the participant was automatically added as a user in the authentication list of identifiers, with their authentication email and password being arbitrarily assigned strings that included their in-game-name (in-game names thus had to be at least 6 characters long). Next, an entry was added in the main database with the in-game name as key, and condition name as value. This was handled automatically from our server using the firebase API.

2.7. Training game

2.7.1. Participant side

The participants were instructed to connect to the game webpage every day to complete their assigned game session for the day. To complete the training, the participants were required to have played 12 h of the game. However, to limit attrition and given the constraints of daily life, the duration of the training, and the fact that this study was conducted with minimal experimenter contact, the schedule allowed for potential missed days and maximum flexibility. The game play was divided into 10-min blocks, and the participants could complete at most 3 blocks per day. Thus, to complete the 72 required blocks, a minimum of 24 days (3.5 weeks) were required. The design allowed participants to take up to 42 days (6 weeks) from the day they connected to the game for the first time to complete the training. After each block, the participants were given the number of blocks that they had completed so far out of the 72 required so that they could keep track of their progression. In addition, the participants did not need to complete all three 10-min blocks in one seating, so they only needed to find a couple of 10-min time periods in their day to complete their daily requirements rather than a single 30-min slot which can be more challenging.

During each block, we also provided the participants with a timer counting down from 10 min as well the number of blocks completed so far. The participants could complete their three daily blocks in more than one seating, but they would have to restart a block if they left before the end of it. Finally, a block was only considered complete when the participants actually reached the end of the 10 min. The number of available blocks to complete was reset every day at 2 AM.

2.7.2. Experimenter side

2.7.2.1. Keeping track of the participants' progress during the training game. To meet our goal of supervising several tens of participants simultaneously with only one experimenter, we designed a dashboard where the progress of each participant was easily accessible (see Fig. 7). The dashboard was essentially a table where each column represented a day and each line a participant. The cells of the table were then colored according to the number of blocks of the task the participants had completed that day. Green was used for 2 or 3 blocks completed (on schedule or better), orange was used when only one block was completed, and red if no block had been completed that day. In addition, the last column showed the total number of minutes completed by the participant so far (which translated in number of blocks by dividing this number by 10) as well as the number of minutes they should have completed to be on schedule (which we defined as 20 min, or 2 blocks, per day, every day). If the number of minutes completed so far was lower than 80 % of the number of minutes they should have spent on the training since they had started the game, that cell was filled in pink instead of the usual grey. In addition, if a participant had not completed any block over the last two days (current day not included), the border of that cell was a thick red line instead of the usual thin black line. This allowed us to efficiently track which participants were starting to fall behind such that we could contact them to ensure that there were not experiencing technical issues, that they correctly understood the schedule they should adhere to, or simply to remind them to train.

This experimenter dashboard was operated fully online from the server, such that the databases from Firebase were downloaded on our

Subj. ID	2021/12/17	2021/12/18	2021/12/19	...	2022/01/12	2022/01/13	2022/01/14
Prolific_1							500/450
Prolific_2							250/390
Prolific_3							390/420
Prolific_4							220/450
...							

- Bolded borders are “frozen” (like in Excel), in other words the first row and columns are always visible
- Subjects are ordered by oldest to newest (arranged by starting date)
- Squares are color-coded as follows :
 - 0 minutes done : red
 - Less than 20 minutes : orange
 - 20 minutes or more : green
 - Before training started : empty (white)
 - Today : gray
- In the today cell, the following information are shown :
 - Number of minutes completed (overall)
 - Number of minutes to be on track, i.e. $20 \times (\text{number of days since start})$
- The today cell follows the following formatting
 - After 5 days, if the number of minutes done is lower than 80% of the number of minutes to be on track, color pink
 - If the last two days are missed (i.e. red), make the border thick red.

Fig. 7. Experimenter dashboard to keep track of the participants' progress throughout the training phase.

own server each time the experimenter connected to the dashboard webpage. This avoided the need to manually download the databases for each game condition on the experimenter's computer each time they wanted to check the participants' progress (i.e. at least daily, but sometimes multiple times a day).

2.7.2.2. Keeping track of the participants in the business-as-usual group.

For the business-as-usual group, we created an Excel spreadsheet where we reported the time at which the participant had finished the pre-test, a column with the day at which they should be recontacted and be added to the *allowlist* of the Prolific study for the first session of post-test, and a column that counted down to that date (using the TODAY() function in Excel). To make tracking easier, the cells in the countdown column were highlighted in red when it reached zero. Just like for the dashboard for the game data, this sheet was checked daily by the unmasked experimenter.

2.8. One-day wait time between game and post-test

2.8.1. Participant

Because we were interested in long-term effects of our intervention game on cognition, and not short-term effects such as arousal, we required a minimum of 24 h between the end of the training game and the first session of post-test. At the end of the training game, the participants were directed to a page where they were asked to enter their prolific ID. Then they were instructed that they needed to wait 24 h before the next step of the study (i.e. post-test) and that the session would be open to them the next day. They had a maximum of 6 days to start the post-test once the session was open to them.

2.8.2. Experimenter

The experimenter recorded the date at which the participants finished the training in the participant table. Every day, an experimenter would check the participant table and add the participants that had completed the training game the day before to the *allowlist* of the Prolific study for day 1 of the post-test.

2.9. Experimenters' roles and masking

Our final 12-h training study included 259 participants and was run by just three experimenters over a period of 125 days (~4 months). One experimenter was in charge of recruitment, assisting participants during pre-test and post-test when needed, and opening the sessions for pre-test and post-test (see Section 2.4.2). This experimenter was masked to the group assignment of the participant. The second experimenter oversaw participants during the training phase and was thus unmasked to their experimental condition. Their main responsibility included responding to participants' questions regarding the training and contacting participants falling behind the intended training schedule to address potential issues preventing them from playing the game. This experimenter was also in charge of monitoring when the participants in each experimental group were done with their training/wait period, so that the masked experimenter could open the Prolific study for the first day of training for them. To limit any experimenter bias, the experimenters followed pre-defined scripts such that all participants who were contacted received similar e-mails. Finally, the third experimenter was tasked with solving technical issues that may arise with participants and was not in contact with the participants.

3. Results

3.1. Participant retention

To assess barriers to participant retention throughout the study, we tabulated how many participants completed each of the different stages

of the study. Fig. 8 shows the number of participants that completed the following stages and the corresponding retention/adherence rate: (1) *Prolific gaming habit question*: Recruited participants were asked again for their gaming habits using exactly the same question as when they had filled in their Prolific profile, only 262 out of 619 (42 %) who initially showed interest in the study reported at this stage gaming habits that did match what they had previously reported on their Prolific profile, despite us using the exact same video game habit question. (2) *Consent form agreement* - 258 people out of 262 (98 %) consented to our study. (3) *Pre-test* - 166 out of 258 participants (64 %) continued with the study after reading the instructions which covered the specifics of pre-test, training - which clearly indicated the full 12 h of time commitment, and post-test. (4) *Training* - 141 from 166 (85 %) completed the training and finally (5) *post-test* - 128 out of 141 (91 %) completed the post-test.

Thus, out of the 620 participants who expressed initial interest in the study, 259 consented to it and 128 completed it fully (adherence rate = 49 % - see also Fig. 9 for the full CONSORT diagram).

3.2. Participant feedback

After consent, the participants who failed to complete a stage in the mandated time were contacted to make sure that they were not facing any technical issue and check whether they may want to drop-out of the study. In the latter case, they were then prompted with the question on Qualtrics: "In case you have opted out of the study, could you let us know the reasons why?" Their responses were sorted in four categories: (i) no reason provided, (ii) schedule conflict, (iii) technical issues, and (iv) poor experience with game and tasks. Out of the 131 participants who opted out, below are their distribution per category of reason (see also in Fig. 10).

- *No reason (NONE)*. Most participants (74 % - 96 participants) did not reply to our message. Most (79) of these participants dropped out of the study at the end of pre-test. During training and post-test, 17 participants did not respond to our message.
- *Scheduling conflicts (Time)*. 18 % (23 participants) mentioned realizing they could not keep up with the schedule of the training ("Schedule"), citing exams, vacation or work-related scheduling conflicts. Before starting the training, 11 participants stated scheduling conflicts that prevented them from continuing with the training; 10 cited conflicts during the training, and 2 during post-test.
- *Technical reasons*. 4 % (5 participants) opted out because of technical issues, all of which were due to participant realizing that their hardware was incompatible with the requirements of the study (we could not make the game work for four of them, and for the last one, one post-test task did not load).
- *Their experience with the games and tasks (UX)*. 4 % (5 participants) opted out because of their experience with either our assessments or training. 2 participants reported that the MOT-part of the training game was too repetitive and they did not want to spend 12 h on it, 1 reported that they found the MOT-part of the training game too hard, and finally 1 reported that could not do one of our post-test task because of a learning disability.
- *Other*. Finally, 1 % (1 participant) opted out because they found the study pay rate too low given its extended duration and the high difficulty of the training game.

3.3. Sample diversity

Country of residence was available from the Prolific profile of 232 participants out of the 258 who consented. Fig. 11 shows the distribution of countries for those 232 participants, as well as the distribution of countries for the 128 participants who completed the study. Not surprisingly, most of participants resided in the UK (home country of Prolific), with South Africa and Portugal coming in second and third

N at each stage of the experiment
Total N at recruitment = 619

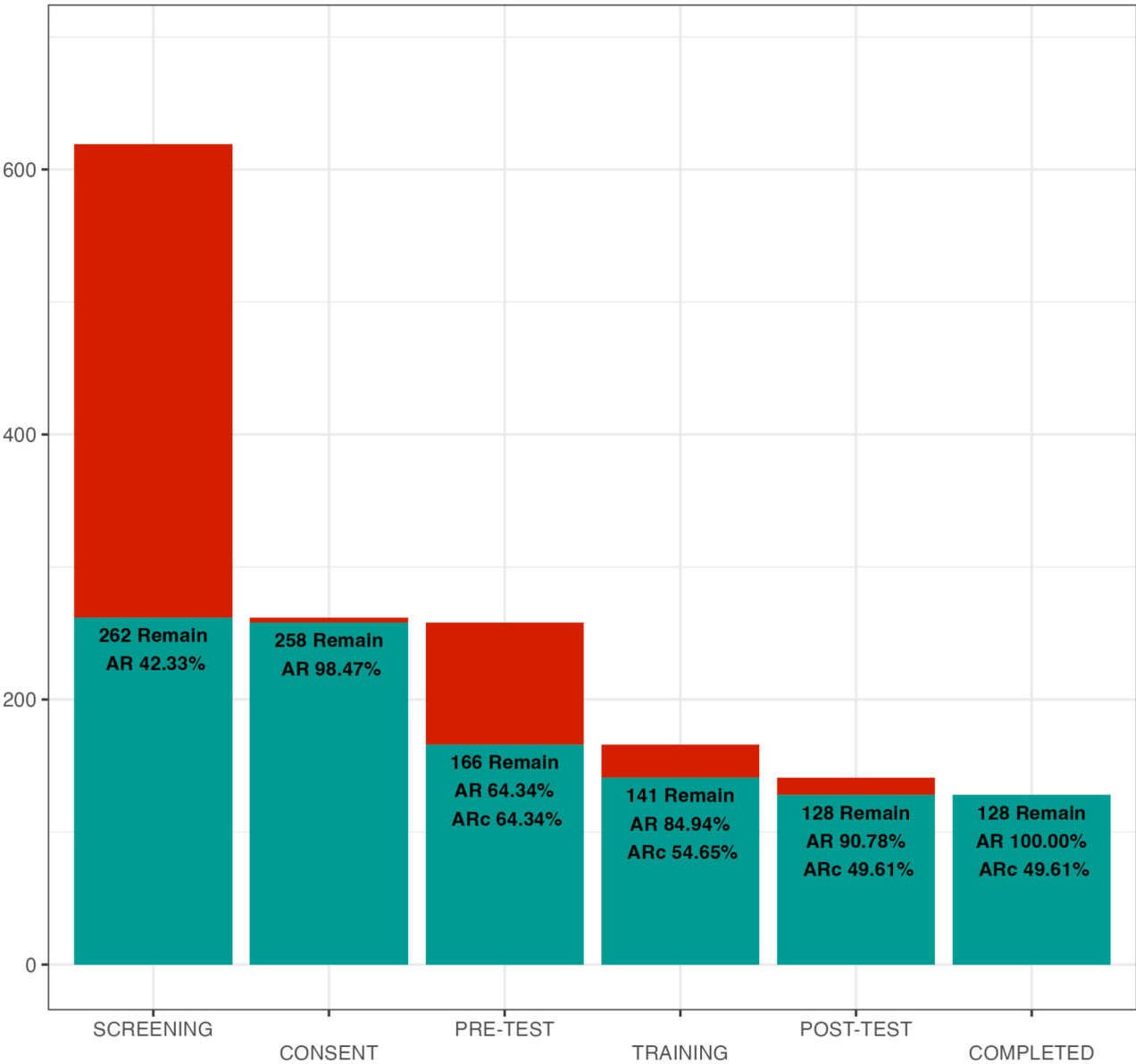


Fig. 8. Retention profile across the various stages of the experiment. The percentage of participants of each stage who moved on to the next stage is reported as AR, and the overall adherence rate at each stage out of participants who consented is denoted as ARc.

respectively (Fig. 11). We also had several participants from the Americas (Mexico and the US mostly). Of note, the distribution of country of residence of the participants who completed the study matches that of participants who consented to the study.

In addition, we also investigated the student and employment status of the participants. This information was available through the prolific profile for 200 participants, 120 of whom were students (58 %). For the 88 non-student participants, 80 participants provided employment status, 65 (81 %) had at least a part-time job, 15 (19 %) were unemployed.

4. Discussion

The demand for more, better powered, intervention studies on more diverse populations is a major driver for developing computerized training and assessments. We present here a set of tools allowing for the deployment of a fully online training study with a near-fully automated

pipeline, so as to reduce the personnel demands, cost, and time needed to complete such training studies. In our case, we were able to recruit 259 participants over a period of only 4 months, 128 of which fully completed our 5 h of assessment and 12-h training intervention, taking between 4.5 and 9 weeks to do so. This was done with a team composed of a lab technician that dedicated 1 h per day to this experiment, one Master's student who dedicated between 10 and 30 min per day, and a PhD student who was available only for technical support. All told there were approximately 130 work hours expended in total across all research-team personnel on the study.

For comparison, a hybrid intervention study previously conducted in the lab, with 9 h of in-lab assessments and 45 h of at-home training (over 9 weeks), took a total of eight months of cumulative duration to complete for a recruited sample size of 69 participants, 52 of whom completed the study (Zhang et al., 2021). The team for this study included five Master's students and one post-doctoral fellow, working on

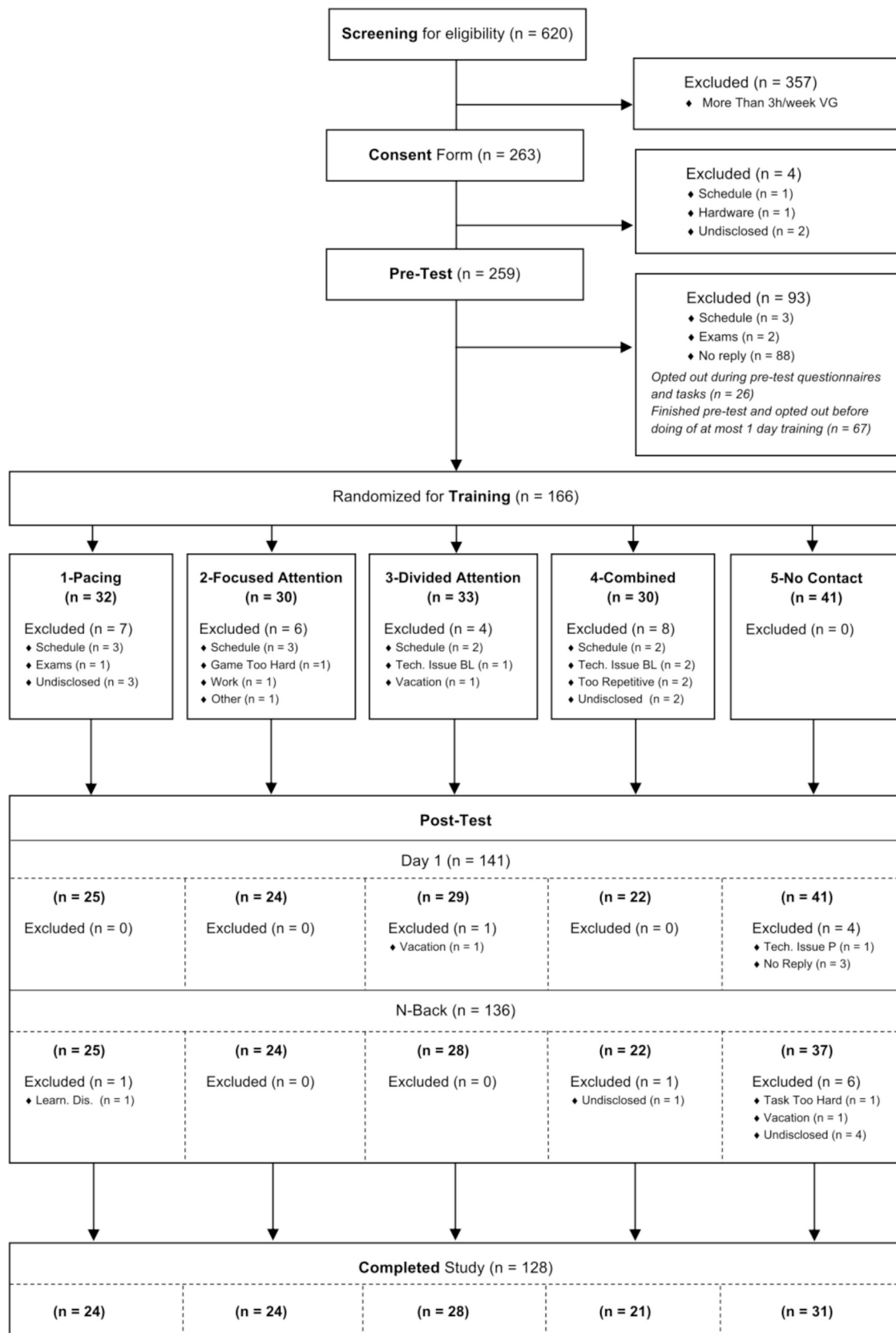


Fig. 9. CONSORT diagram of the study.

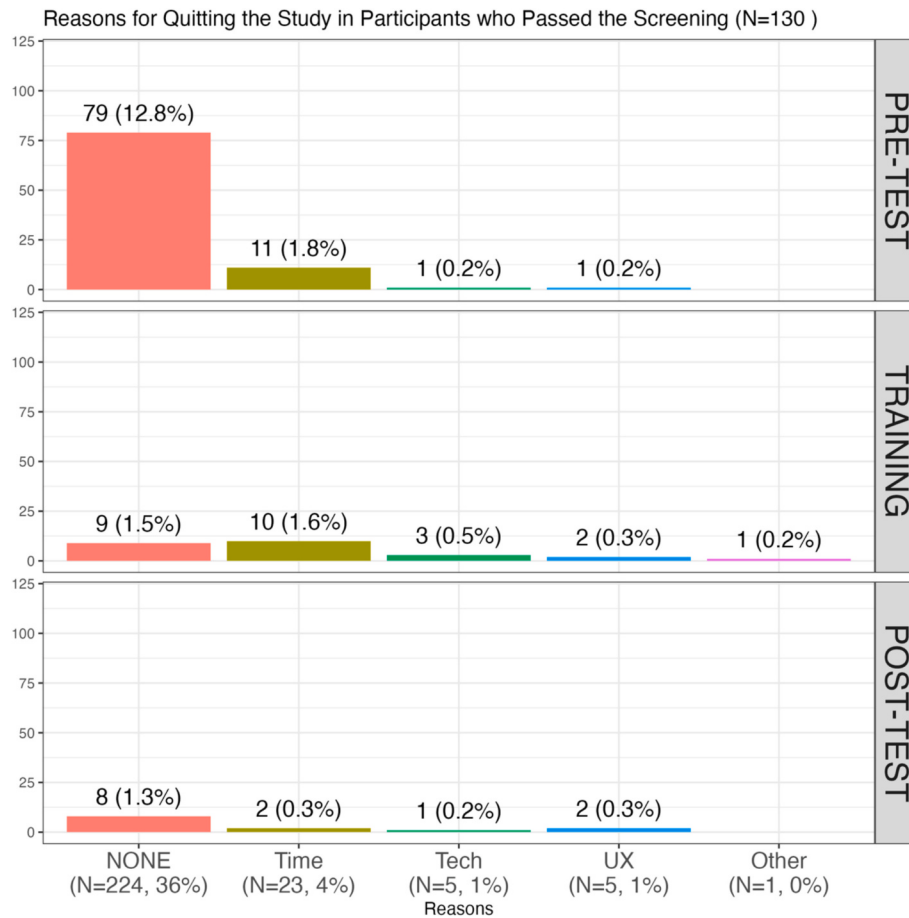


Fig. 10. Number of participants per category of reasons for opting out of the studies.

average 6.5 h/week each. In total this already reflects an order of magnitude greater personnel time compared to the currently described system (~1380 work hours in total across all research-team personnel).

Even after considering the longer assessment and training durations and the larger staff size employed in Zhang et al. (2021), it would still have taken around 5 times as long and around 20 times more work hours to reach the same number of participants who completed the study as we did with the pipeline described here. Still, given the large differences across intervention studies, more studies comparing effectiveness and cost-effectiveness of in-person v. remote intervention studies are needed.

We were also able to recruit more broadly across the world. The continent of origin of the 128 participants who completed the study was in Europe (88), North America (20), Africa (18), Oceania (1), and Asia (1). Of note, participants from global South nations did not show increased opt-out rate, which was a concern as additional technical difficulties may have arisen due to poorer internet infrastructures than in global North nations (speedtest.net, 2021). Finally, our pipeline allowed us to easily implement strong masking practices, such that the experimenter aware of the participant conditions, and the one unaware of them, had access to completely separate sets of tools. While more work clearly needs to be done to include non-WEIRD participants, this study already constitutes an advance compared to most studies which enroll mostly Global North Undergraduates.

Regarding adherence rate, without the direct supervision of an experimenter, our approach to maximize adherence was multifold and involved careful decisions at each stage of the process, from recruitment and consent, to in-progress touchpoints, to final payments. For instance, we were exceptionally careful during the consent process to fully describe the time requirements of the study to ensure that participants

came into the study fully aware of those requirements. Upon enrolling, participants encountered a dashboard that allowed them to monitor their progress at each timepoint. The computerized training program was designed such that participants could complete the training to accommodate their schedule as flexibly as possible; to that end, the daily requirements could be completed in independent chunks of 10 min. Furthermore, participants who failed to complete their daily sessions for 2 days in a row were contacted by our team to let them know they might run behind their training schedule, and to make sure they had not encountered any technical difficulties. Finally, participants were given a completion bonus payment for finishing the full study. However, because attrition in a study of this type is simply unavoidable, to ensure that our data allowed for rigorous and reproducible results, we ensured that our groups remained balanced even with attrition by using an algorithm that forced a similar number of participants in each group while matching the groups for age, sex, and performance at pre-test on one of the main DV of interest.

Now, the question is whether we were successful in incentivizing participants to complete tens of hours of a cognitive intervention without them dropping out of the study due to either boredom or frustration, which have been reported as side-effects in many cognitive training interventions. Overall, the adherence rate of the study from the moment the participants consented until the end of the study was 128/259 that is 49 %. This is rather low compared to the adherence rate for video game-based intervention studies conducted so far in our laboratories which has been of about 85 %. In some cases, this rate was boosted up to nearly 100 % with proper planning at onboarding (Feng et al., 2007; Green & Bavelier, 2006; Zhang et al., 2021). However, our low number most likely suffers from the fact that most these previous studies we compare ourselves to used commercial-grade and highly successful

Country of Residence of the Participants, Overall. N = 232/258

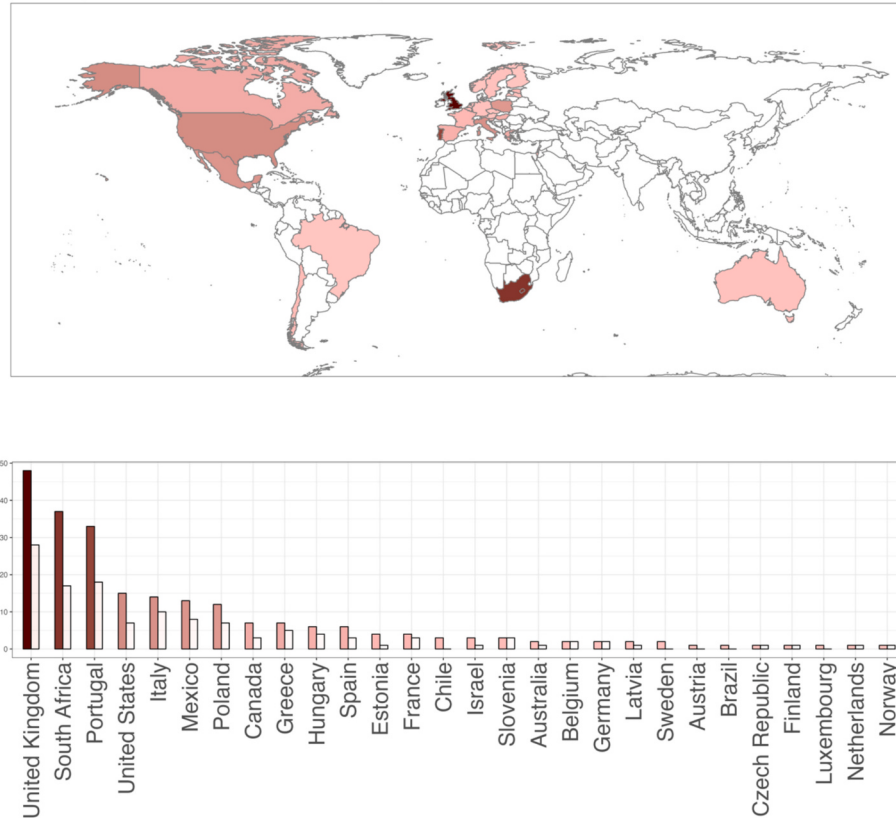


Fig. 11. Distribution of the country of residence for the participants who entered the study (colored bars) and who completed the study (white bars). Data was available for 232 out of 258 participants.

video games as training interventions (such as *Call of Duty* or *Zoo Tycoon*). Our interventions, while “gamified,” were simply not designed to compete with the quality of professional games. Indeed, intervention studies for cognitive training, usually using gamified tasks during training, have observed much lower adherence rates. For example, when training on a gamified version of the N-back task, adherence rates were 70 % for in-lab studies and 60 % for remote studies with experimenter supervision (Collins et al., 2022). Note that our study did not have such social component. Regarding remote training, in a study using *Cogmed* as intervention for 25 training sessions over 5 to 6 weeks (a schedule comparable to ours), Mawjee et al. (2017) reported an adherence rate of 75 % when the participants completed 15-min sessions and a 47 % adherence rate for 45-min sessions. Finally, an intervention study using a gamified version of the N-back task, over a similar schedule as ours but with experimenter supervision during pre-test and post-test, recorded an adherence rate of 46 % (Jaeggi et al., 2014). As such, it seems that our approach to limit attrition worked as expected given the constraints of this paradigm.

The present study also highlights key study stages where many subjects could not be retained. The first one is the video game screening question, where participants were asked about their gaming habits using the exact same question they had previously answered on Prolific. Our study called for only enrolling participants who fit our pre-set video game-based experience criteria on that very question. Surprisingly, more than half of the people that could access our recruitment information reported different gaming habits than what they had in their profile on Prolific. This is likely due to habit changes since they had initially answered that question in their Prolific profile months, if not years, before. This situation flags a crucial weakness when relying on pre-existing Prolific profile questions, especially for variables that may change over time such as media consumption. This also poses a

challenge to recruiting populations whose size is rather limited. For example, while Prolific currently (October 2023) boasts 130,000 active users (Prolific, 2023), the total number of participants that fit our Prolific-based video game criteria of less than 3 h of weekly video game play was comparatively quite low at 2851 (Males: 1289, Females: 1552, Undisclosed: 10). If that low number may be then substantially cut because of shifts in the initially reported variable, larger studies may well run out of participants who meet inclusion/exclusion criteria.

The second critical stage where participants could not be retained is when they opted out is at the end of pre-test. 29 % or 68/259 of the sample who passed the screening, either never created an account for the training part of the study or after briefly viewing the training game did not continue with the training. Here, we have identified a few weaknesses of the present study that should be improved for future similar studies. The first is the relatively large, and possibly daunting amount of information given to participants once they had finished the pre-test and as the training was introduced. Indeed, before onboarding participants on their assigned intervention, we needed to make sure they fully understood the play schedule they would have to follow. In addition, we needed to clearly state that exclusion of the study would occur should they not be able to complete their 12 h of training within this demanding schedule. This entailed providing several examples to fit the different situations the participants may be in. In the absence of researcher contact, this part was designed to be extremely comprehensive as experience has shown that any imprecision will lead to some participants not understanding the protocol. Unfortunately, this also resulted in what is often termed a “wall of text”, a situation known to lead to disinterest. To reduce the indigestibility of this wall of text, future studies may benefit from conveying the same information through pictures and schematics to introduce the different possible training regimens that the participants may be assigned to and the

training schedule they may follow (for example 30 min per day for only 4 weeks, 20 min per day over 6 weeks, or the fact that they can “compensate” for off days by doing 30 min per day on the next few days). Another relevant addition might be to also provide clear and concise instruction video(s) as some participants may prefer watching a video to reading text. This would also have the additional benefit of “humanizing” the experimenters to the participants, something that may go a long way in reducing the mental barriers to contacting experimenters when participants have any doubt or questions, given the pipeline as deployed did not call for direct contact between experimenters and participants. These videos may however create biases that need to be considered when designing the experiment, for example because of gender, age, ethnicity/race, social status, perceived professionalism (Aslaksen et al., 2007; Kállai et al., 2004; Marx & Goff, 2005; McCallum & Peterson, 2015; Modic-Stanke & Ivanec, 2016; Nichols & Maner, 2008).

A second weakness is that some participants may have entered the pre-test study with the intention to collect the reward for the pre-test, with no intention of completing the whole study. Given the way Prolific operates, it is difficult to conceive of ways to identify such participants or of actions that may discourage such behaviors. In addition, the reality of such intervention studies is that it can be daunting for participants to complete multiple hours of a task that presents limited interest. The end of pre-test being the first time participants experience the training task, this point in time may very well be when they might decide not to pursue the reward for the full training. Thus, it appears that this increased drop-out rate at pre-test is an inherent part of the costs of running studies on Prolific that needs to be considered when calculating the resources needed to run the study and evaluating sample sizes at recruitment. From this point of view, it should be clear that researchers should plan to initially recruit around twice as many participants as their protocol suggests they will eventually need to address the questions at hand.

A third weakness is that an experimenter had to connect to Prolific every day to check on participants that had completed the first session of post-test so as to add them to the *allowlist* of the Prolific Study for the next session. This was the most time-consuming experimenter task in our study (about one hour every day). Also, because of the way our server interacted with Firebase, it was necessary to download the full database on our server every time we connected to the experimenter training dashboard (Fig. 7). This meant that loading the dashboard took longer and longer as we neared toward the end of the experiment, wasting time and computational resources. The ability to either connect remotely to the database or limit the download to only new data would solve this issue. In addition, a major hurdle faced by remote interventions like ours is participant compliance with their assigned tasks, especially during training which can be somewhat boring. In our study, to check whether participants tried to “cheat” by mindlessly going through the training task instead of actively engaging with it, we would have needed to download and locally analyze the training data, which was not sustainable for regular checks. An ideal solution would be to show the number of trials completed for each round of the training game on the experimenter training dashboard (Fig. 7). Another improvement would be to automatically save the date at which the training ended in a table, and remove the step where participants manually enter their Prolific ID on a webpage to signal that they have completed the training. Finally, we used an Excel sheet shared between the experimenters in contact with the participants and the experimenter in charge of technical support. While this was an acceptable solution, there exist much better and more adapted tools now to handle the management of technical issues (e.g. Github, Gitlab, Youtrack), which should be considered in future experiments.

Fourth, on the design side, we purposefully built the game such that it could be used by the participants and experimenters with as few requirements as possible. On the participant's side, the game was accessible through a browser as we did not want participants to have to

download any program on their computer. This proved to be a good choice as some participants reported being concerned at the idea of having to download a program whose origin they did not know, and some even going as far as requesting playing our intervention game in incognito mode (which prevents the game from saving data on the participant's computer). We were not prepared for participants wanting to play in incognito mode so some features of our game did not work for them as smoothly; the most important one being the auto-login, such that those participants had to re-enter their in-game name every day to connect to the training game. In addition, while we designed our pipeline to be as accessible as possible to any young adult, including non-WEIRD ones, there are other challenges outside the design of the experiment that might prevent someone from enrolling in the study. Indeed, such studies demand availability as the time invested by participants is significant (e.g. more than 10 h of training, plus all the testing time); they also demand resources such as a stable internet connection and a somewhat recent computer. Furthermore, some populations may have skepticism or concerns about participation in such online, remote intervention studies, especially video game-based ones. On the experimenter's side, the game was coded in a way to be agnostic to the server it is hosted on, which gave it good transferability from one server to another as only the game files and a working server are required to put the game online. This feature is convenient when copying the game for a new experiment. Unfortunately, it also means that the game does not have access to resources on the host server, such as databases or functions that are not part of the game package. This includes the subject database where the ID of the participants and progress data unrelated to pre or post-test were stored (see Section 2.3.1). This would have been convenient to (i) automatically store the in-game name chosen by the participant, and (ii) to automatically add the date at which the participant finished their training to the table, instead of relying on the participant entering their Prolific ID in a dedicated page, which may lead to human error. A related point to consider is that in order to abide by more and more stringent protection of privacy laws, it might be advisable to not use the Prolific ID to identify the participants throughout the study, but rather assign them another random ID immediately upon inclusion in the study, with that ID being used throughout the study. The correspondence between Prolific ID and this new random ID should then be saved on a secure database.

Fifth, while the absence of contact with experimenters is a strength for basic science (i.e., in studies where the goal is to isolate the mechanism of a given intervention), it could be a weakness in settings where the goal is to maximize the impact of the intervention (such as clinical practice). Indeed, there are multiple data points suggesting that interactions with a researcher and/or caregivers will lead to positive outcomes (Levoy et al., 2022; Parong et al., 2023; Riedl & Schüssler, 2017). The absence of contact with the experimenter also limits the range of populations with whom this procedure may be used, especially clinical populations that might not be able to use a computer fully unsupervised. As such, when designing a study, care should be taken to choose the design and procedure(s) that best match the population and the goals of the study. In other words, the procedure presented in this paper may not be best-suited for a clinical study without adjustments that take into account the clinical goals and population.

In all though, while there are certainly improvements that can be made to the process and pipeline, that presented here already allowed us to overcome many of the largest limitations that have been identified in previous research and thus can serve as a springboard for larger, longer, and more methodologically rigorous behavioral interventions into the future.

CRedit authorship contribution statement

Freyja Joessel: Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft.

Sylvie Denkinger: Visualization, Investigation, Writing – review & editing. **Paul-Emile Joessel:** Software, Methodology, Writing – review & editing. **C. Shawn Green:** Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization, Writing – review & editing, Writing – original draft. **Daphne Bavelier:** Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization, Writing – review & editing, Writing – original draft.

Declaration of competing interest

The author declare no conflict of interest.

Acknowledgements

We thank the Office of Naval Research grant numbers N00014-20-1-2074 and N00014-22-1-2559 to DB, the EU Synergy grant number 810580 BrainPlay to DB, and the Office of Naval Research grant N00014-22-1-2283 to C.S.G. that supported this work. We also thank Hugo Marone who helped with the data collection, and we thank all the participants who took the time to participate in this study.

Data availability

The data presented in this article and codes to generate the figures and analyses are available on OSF (osf.io/3w5jy). The code base supporting the intervention pipeline is available on gitlab (gitlab.com/NegativePotato/fully-online-rct).

References

- Abdin, S., Welch, R. K., Byron-Daniel, J., & Meyrick, J. (2018). The effectiveness of physical activity interventions in improving well-being across office-based workplace settings: A systematic review. *Public Health*, 160, 70–76. <https://doi.org/10.1016/j.puhe.2018.03.029>
- Adams, D. M., Pilegard, C., & Mayer, R. E. (2016). Evaluating the cognitive consequences of playing *portal* for a short duration. *Journal of Educational Computing Research*, 54(2), 173–195. <https://doi.org/10.1177/0735633115620431>
- Addelman, S. (1969). The generalized randomized block design. *The American Statistician*, 23(4), 35–36. <https://doi.org/10.1080/00031305.1969.10481868>
- Ainsworth, B., Eddershaw, R., Meron, D., Baldwin, D. S., & Garner, M. (2013). The effect of focused attention and open monitoring meditation on attention network function in healthy volunteers. *Psychiatry Research*, 210(3), 1226–1231. <https://doi.org/10.1016/j.psychres.2013.09.002>
- Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuhl, M., Benson, B. L., Jennett, S., ... Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, 228(1), 107–115. <https://doi.org/10.1016/j.bbr.2011.11.040>
- Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, 53(4), 1407–1425. <https://doi.org/10.3758/s13428-020-01501-5>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Aslaksen, P. M., Myrbaek, I. N., Hoifødt, R. S., & Flaten, M. A. (2007). The effect of experimenter gender on autonomic and subjective responses to pain stimuli. *Pain*, 129(3), 260–268. <https://doi.org/10.1016/j.pain.2006.10.011>
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 22(2), 366–377. <https://doi.org/10.3758/s13423-014-0699-x>
- Ball, K. K., Ross, L. A., Roth, D. L., & Edwards, J. D. (2013). Speed of processing training in the ACTIVE study: How much is needed and who benefits? *Journal of Aging and Health*, 25(8 suppl), 65S–84S. <https://doi.org/10.1177/0898264312470167>
- Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144(1), 77–110. <https://doi.org/10.1037/bul0000130>
- Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023). Effects of action video game play on cognitive skills: A meta-analysis. *Technology, Mind, and Behavior*, 4(1). <https://doi.org/10.1037/tmb0000102>
- Belchior, P., Marsiske, M., Sisco, S. M., Yam, A., Bavelier, D., Ball, K., & Mann, W. C. (2013). Video game training to improve selective visual attention in older adults. *Computers in Human Behavior*, 29(4), 1318–1324. <https://doi.org/10.1016/j.chb.2013.01.034>
- Berger, E. M., Fehr, E., Hermes, H., Schunk, D., & Winkel, K. (2020). The impact of working memory training on children's cognitive and noncognitive skills. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3622985>
- Bigorra, A., Garolera, M., Guijarro, S., & Hervás, A. (2016). Long-term far-transfer effects of working memory training in children with ADHD: A randomized controlled trial. *European Child & Adolescent Psychiatry*, 25(8), 853–867. <https://doi.org/10.1007/s00787-015-0804-3>
- Blackner, K. J., & Curby, K. M. (2013). Enhanced visual short-term memory in action video game players. *Attention, Perception, & Psychophysics*, 75(6), 1128–1136. <https://doi.org/10.3758/s13414-013-0487-0>
- Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, 8(4), 445–454. <https://doi.org/10.1177/1745691613491271>
- Bostock, S., Crosswell, A. D., Prather, A. A., & Steptoe, A. (2019). Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. *Journal of Occupational Health Psychology*, 24(1), 127–138. <https://doi.org/10.1037/ocp0000118>
- Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00063>
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, Article e9414. <https://doi.org/10.7717/peerj.9414>
- Bruhn, M., & McKenzie, D. (2023). *Pursuit of balance: Randomization in practice in development field experiments* (Vol. 1) (4).
- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 5(8), 980–989. <https://doi.org/10.1038/s41562-021-01143-3>
- Chen, F.-T., Etner, J. L., Chan, K.-H., Chiu, P.-K., Hung, T.-M., & Chang, Y.-K. (2020). Effects of exercise training interventions on executive function in older adults: A systematic review and meta-analysis. *Sports Medicine*, 50(8), 1451–1467. <https://doi.org/10.1007/s40279-020-01292-x>
- Choi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, 40(6), 531–542. <https://doi.org/10.1016/j.intell.2012.07.004>
- Chuang, L.-Y., Hung, H.-Y., Huang, C.-J., Chang, Y.-K., & Hung, T.-M. (2015). A 3-month intervention of dance dance revolution improves interference control in elderly females: A preliminary investigation. *Experimental Brain Research*, 233(4), 1181–1188. <https://doi.org/10.1007/s00221-015-4196-x>
- Collins, C. L., Pina, A., Carrillo, A., Ghil, E., Smith-Peirce, R. N., Gomez, M., ... Seitz, A. R. (2022). Video-based remote Administration of Cognitive Assessments and Interventions: A comparison with in-lab administration. *Journal of Cognitive Enhancement*, 6(3), 316–326. <https://doi.org/10.1007/s41465-022-00240-z>
- De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, 8(85), 5351. <https://doi.org/10.21105/joss.05351>
- Denkinger, S., Spano, L., Bingel, U., Witt, C. M., Bavelier, D., & Green, C. S. (2021). Assessing the impact of expectations in cognitive training and beyond. *Journal of Cognitive Enhancement*, 5(4), 502–518. <https://doi.org/10.1007/s41465-021-00206-7>
- Deveau, J., Jaeggi, S. M., Zordan, V., Phung, C., & Seitz, A. R. (2015). How to build better memory training games. *Frontiers in Systems Neuroscience*, 8. <https://doi.org/10.3389/fnys.2014.00243>
- Dotson, V. M., & Duarte, A. (2020). The importance of diversity in cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464(1), 181–191. <https://doi.org/10.1111/nyas.14268>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fedewa, A., & Ahn, S. (2011). The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: A meta-analysis. *Research Quarterly for Exercise and Sport*, 82(3). <https://doi.org/10.5641/027013611X13275191444107>
- Feldt, L. S. (1958). A comparison of the precision of three experimental designs employing a concomitant variable. *Psychometrika*, 23(4), 335–353. <https://doi.org/10.1007/BF02289783>
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, 18(10), 850–855. <https://doi.org/10.1111/j.1467-9280.2007.01990.x>
- Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences*, 113(27), 7470–7474. <https://doi.org/10.1073/pnas.1601243113>
- Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current Biology*, 23(6), 462–466. <https://doi.org/10.1016/j.cub.2013.01.044>
- Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., & Granger, C. B. (2015). *Fundamentals of clinical trials*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-18539-2>
- Gavelin, H. M., Dong, C., Minkov, R., Bahar-Fuchs, A., Ellis, K. A., Lautenschlager, N. T., ... Lampit, A. (2021). Combined physical and cognitive training for older adults with and without cognitive impairment: A systematic review and network meta-analysis of randomized controlled trials. *Ageing Research Reviews*, 66, Article 101232. <https://doi.org/10.1016/j.arr.2020.101232>
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534–537. <https://doi.org/10.1038/nature01647>

- Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1465–1478. <https://doi.org/10.1037/0096-1523.32.6.1465>
- Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, 18(1), 88–94. <https://doi.org/10.1111/j.1467-9280.2007.01853.x>
- Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Anson, U., Ball, K. K., ... Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, 3(1), 2–29. <https://doi.org/10.1007/s41465-018-0115-y>
- Green, C. S., Strobach, T., & Schubert, T. (2013). On methodological standards in training and transfer experiments. *Psychological Research*, 78(6), 756–772. <https://doi.org/10.1007/s00426-013-0535-3>
- Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., & Bavelier, D. (2012). The effect of action video game experience on task-switching. *Computers in Human Behavior*, 28(3), 984–994. <https://doi.org/10.1016/j.chb.2011.12.020>
- Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, 21(2), 181–192. <https://doi.org/10.1006/ceps.1996.0015>
- Haimov, I., & Shatil, E. (2013). Cognitive training improves sleep quality and cognitive function among older adults with insomnia. *PLoS One*, 8(4), Article e61390. <https://doi.org/10.1371/journal.pone.0061390>
- Hale, G. E., Colquhoun, L., Lancaster, D., Lewis, M., & Tyson, P. J. (2021). Review: Physical activity interventions for the mental health and well-being of adolescents — A systematic review. *Child and Adolescent Mental Health*, 26(4), 357–368. <https://doi.org/10.1111/camh.12485>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods*, 48(4), 1718–1724. <https://doi.org/10.3758/s13428-015-0678-9>
- Hillman, C. H., Castell, D. M., & Buck, S. M. (2005). Aerobic fitness and neurocognitive function in healthy preadolescent children. *Medicine & Science in Sports & Exercise*, 37(11), 1967–1974. <https://doi.org/10.1249/01.mss.0000176680.79702.ce>
- Honick, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, 17, 63–84. <https://doi.org/10.1016/j.edurev.2015.11.002>
- Hou, J., Jiang, T., Fu, J., Su, B., Wu, H., Sun, R., & Zhang, T. (2020). The long-term efficacy of working memory training in healthy older adults: A systematic review and meta-analysis of 22 randomized controlled trials. *The Journals of Gerontology: Series B*, 75(8), e174–e188. <https://doi.org/10.1093/geronb/gbaa077>
- Houle, S. (2015). An introduction to the fundamentals of randomized controlled trials in pharmacy research. *The Canadian Journal of Hospital Pharmacy*, 68(1). <https://doi.org/10.4212/cjhp.v68i1.1422>
- Hutchinson, C. V., Barrett, D. J. K., Nitka, A., & Raynes, K. (2016). Action video game training reduces the Simon effect. *Psychonomic Bulletin & Review*, 23(2), 587–592. <https://doi.org/10.3758/s13423-015-0912-6>
- Inquisit 5. (2016) [Computer software] <https://www.millisecond.com>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829–6833. <https://doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394–412. <https://doi.org/10.1080/09658211003702171>
- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42(3), 464–480. <https://doi.org/10.3758/s13421-013-0364-z>
- Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence*, 12.
- Jaušovec, N., & Jaušovec, K. (2012). Working memory training: Improving intelligence—Changing brain activity. *Brain and Cognition*, 79(2), 96–106. <https://doi.org/10.1016/j.bandc.2012.02.007>
- Joessel, F. (2022). *Development of a video game to investigate the AVG features promoting attentional control* [Université de Genève]. doi: 10.13097/ARCHIVE-OUVERTE/UNI GE:164866.
- Källai, I., Barke, A., & Voss, U. (2004). The effects of experimenter characteristics on pain reports in women and men. *Pain*, 112(1), 142–147. <https://doi.org/10.1016/j.pain.2004.08.008>
- Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: A systematic review and meta-analysis. *Ageing Research Reviews*, 15, 28–43. <https://doi.org/10.1016/j.arr.2014.02.004>
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 10.
- Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24(6), 781–791. <https://doi.org/10.1076/jcen.24.6.781.8395>
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567. <https://doi.org/10.1146/annurev.psych.50.1.537>
- Levoy, K., Rivera, E., McHugh, M., Hanlon, A., Hirschman, K. B., & Naylor, M. D. (2022). Caregiver engagement enhances outcomes among randomized control trials of transitional care interventions: A systematic review and meta-analysis. *Medical Care*, 60(7), 519–529. <https://doi.org/10.1097/MLR.0000000000001728>
- Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, 12(5), 549–551. <https://doi.org/10.1038/nn.2296>
- Li, R. W., Ngo, C., Nguyen, J., & Levi, D. M. (2011). Video-game play induces plasticity in the visual system of adults with amblyopia. *PLoS Biology*, 9(8), Article e1001135. <https://doi.org/10.1371/journal.pbio.1001135>
- Libertus, M. E., Liu, A., Pikul, O., Jacques, T., Cardoso-Leite, P., Halberda, J., & Bavelier, D. (2017). The impact of action video game training on mathematical abilities in adults. *AERA Open*, 3(4), Article 233285841774085. <https://doi.org/10.1177/2332858417740857>
- Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, 20(1), 135–141. <https://doi.org/10.3758/s13423-012-0335-6>
- Loosli, S. V., Buschkuhl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*, 18(1), 62–78. <https://doi.org/10.1080/09297049.2011.575772>
- Ludyga, S., Gerber, M., Pühse, U., Looser, V. N., & Kamijo, K. (2020). Systematic review and meta-analysis investigating moderators of long-term effects of exercise on cognition in healthy individuals. *Nature Human Behaviour*, 4(6), 603–612. <https://doi.org/10.1038/s41562-020-0851-8>
- Maillat, P., Perrot, A., & Hartley, A. (2012). Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. *Psychology and Aging*, 27(3), 589–600. <https://doi.org/10.1037/a0026268>
- Marquez, D. X., Aguiñaga, S., Vázquez, P. M., Conroy, D. E., Erickson, K. I., Hillman, C., ... Powell, K. E. (2020). A systematic review of physical activity and quality of life and well-being. *Translational Behavioral Medicine*, 10(5), 1098–1109. <https://doi.org/10.1093/tbm/ibz198>
- Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology*, 44(4), 645–657. <https://doi.org/10.1348/014466604X17948>
- Mawjee, K., Woltering, S., Lai, N., Gotlieb, H., Kronitz, R., & Tannock, R. (2017). Working memory training in ADHD: Controlling for engagement, motivation, and expectancy of improvement (pilot study). *Journal of Attention Disorders*, 21(11), 956–968. <https://doi.org/10.1177/1087054714557356>
- Mayo, E. (1993). *The human problems of an industrial civilization* (2nd ed.) (2nd ed., Vol. 3). MacMillan.
- McCallum, E. B., & Peterson, Z. D. (2015). Effects of experimenter contact, setting, inquiry mode, and race on women's self-report of sexual attitudes and behaviors: An experimental study. *Archives of Sexual Behavior*, 44(8), 2287–2297. <https://doi.org/10.1007/s10508-015-0590-5>
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: Evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>
- Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, 110(1), 40–48. <https://doi.org/10.1037/0021-843X.110.1.40>
- Mishra, J., Sagar, R., Joseph, A. A., Gazzaley, A., & Merzenich, M. M. (2016). Training sensory signal-to-noise resolution in children with ADHD in a global mental health setting. *Translational Psychiatry*, 6(4), e781. <https://doi.org/10.1038/tp.2016.45>
- Modic-Stanke, K., & Ivanec, D. (2016). Pain threshold—Measure of pain sensitivity or social behavior? *Psihologija*, 49(1), 37–50. <https://doi.org/10.2298/PSI1601037M>
- Moreau, D., Kirk, I. J., & Waldie, K. E. (2017). High-intensity training enhances executive function in children in a randomized, placebo-controlled trial. *eLife*, 6, Article e25062. <https://doi.org/10.7554/eLife.25062>
- Moreau, D., Morrison, A. B., & Conway, A. R. A. (2015). An ecological approach to cognitive enhancement: Complex motor training. *Acta Psychologica*, 157, 44–55. <https://doi.org/10.1016/j.actpsy.2015.02.007>
- Moreno-Peral, P., Pino-Postigo, A., Conejo-Cerón, S., Bellón, D., Rodríguez-Martín, B., Martínez-Vizcaíno, V., & Bellón, J. A. (2022). Effectiveness of physical activity in primary prevention of anxiety: Systematic review and meta-analysis of randomized controlled trials. *International Journal of Environmental Research and Public Health*, 19(3), Article 1813. <https://doi.org/10.3390/ijerph19031813>
- Morris, D., Fraser, S., & Wormald, R. (2007). Masking is better than blinding. *BMJ*, 334(7597), 799. <https://doi.org/10.1136/bmj.39175.503299.94>
- Netz, Y., Wu, M.-J., Becker, B. J., & Tenenbaum, G. (2005). Physical activity and psychological well-being in advanced age: A meta-analysis of intervention studies. *Psychology and Aging*, 20(2), 272–284. <https://doi.org/10.1037/0882-7974.20.2.272>
- Nguyen, L., Murphy, K., & Andrews, G. (2019). Immediate and long-term efficacy of executive functions cognitive training in older adults: A systematic review and meta-analysis. *Psychological Bulletin*, 145(7), 698–733. <https://doi.org/10.1037/bul0000196>
- Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, 135(2), 151–165.
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 152–158. <https://doi.org/10.1145/191666.191729>
- Novak, E., & Tassell, J. (2015). Using video game play to improve education-majors' mathematical performance: An experimental study. *Computers in Human Behavior*, 53, 124–130. <https://doi.org/10.1016/j.chb.2015.07.001>
- Nyquist, J. B., Lappin, J. S., Zhang, R., & Tadin, D. (2016). Perceptual training yields rapid improvements in visually impaired youth. *Scientific Reports*, 6(1), Article 37431. <https://doi.org/10.1038/srep37431>

- Oei, A. C., & Patterson, M. D. (2013). Enhancing cognition with video games: A multiple game training study. *PLoS One*, 8(3), Article e58546. <https://doi.org/10.1371/journal.pone.0058546>
- Osman, A. M., Jaffe, P. I., Ng, N. F., Kerlan, K. R., & Schafer, R. J. (2023). Transfer of learning: Analysis of dose-response functions from a large-scale, online, cognitive training dataset. *PLoS One*, 18(5), Article e0281095. <https://doi.org/10.1371/journal.pone.0281095>
- Owens, M., Koster, E. H. W., & Derakshan, N. (2013). Improving attention control in dysphoria through cognitive training: Transfer effects on working memory capacity and filtering efficiency: Working memory training in dysphoria. *Psychophysiology*, 50(3), 297–307. <https://doi.org/10.1111/psyp.12010>
- Parong, J., Seitz, A. R., Jaeggi, S. M., & Green, C. S. (2022). Expectation effects in working memory training. *PNAS*, 119(37), 10.
- Parong, J., Vodyanik, M., Green, C. S., Jaeggi, S. M., & Seitz, A. R. (2023). Experiment effects. In A. L. Nichols, & J. Edlund (Eds.), *The Cambridge handbook of research methods and statistics for the social and behavioral sciences* (1st ed., pp. 224–243). Cambridge University Press. <https://doi.org/10.1017/9781009010054.012>
- Pasqualotto, A., Altarelli, I., De Angeli, A., Menestrina, Z., Bavelier, D., & Venuti, P. (2022). Enhancing reading skills through a video game mixing action mechanics and cognitive training. *Nature Human Behaviour*, 6(4), 545–554. <https://doi.org/10.1038/s41562-021-01254-x>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Plass, J., Mayer, R. E., & Homer, B. D. (Eds.). (2019). *Handbook of game-based learning*. The MIT Press.
- Pocock, S. J., & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31(1), 103. <https://doi.org/10.2307/2529712>
- Prakash, R. S., Voss, M. W., Erickson, K. I., & Kramer, A. F. (2015). Physical activity and cognitive vitality. *Annual Review of Psychology*, 66(1), 769–797. <https://doi.org/10.1146/annurev-psych-010814-015249>
- Prins, P. J. M., DAVIS, S., Ponsioen, A., Ten Brink, E., & Van Der Oord, S. (2011). Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? *Cyberpsychology, Behavior, and Social Networking*, 14(3), 115–122. <https://doi.org/10.1089/cyber.2009.0206>
- Prolific. (2023, October). Why Prolific? <https://www.prolific.com/prolific-vs-mturk>
- Rand, D. G. (2012). The promise of mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172–179. <https://doi.org/10.1016/j.jtbi.2011.03.004>
- Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, 26(6), 462–483. <https://doi.org/10.1016/j.tics.2022.03.007>
- Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H., King, J. W., ... Willis, S. L. (2014). Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society*, 62(1), 16–24. <https://doi.org/10.1111/jgs.12607>
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359–379. <https://doi.org/10.1037/a0029082>
- Riedl, D., & Schüller, G. (2017). The influence of doctor-patient communication on health outcomes: A systematic review. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 63(2), 131–150. <https://doi.org/10.13109/zptm.2017.63.2.131>
- Sala, G., Aksayli, N. D., Tatlidil, K. S., Gondo, Y., & Gobet, F. (2019). Working memory training does not enhance older adults' cognitive skills: A comprehensive meta-analysis. *Intelligence*, 77, Article 101386. <https://doi.org/10.1016/j.intell.2019.101386>
- Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, 144(2), 111–139. <https://doi.org/10.1037/bul0000139>
- Sassenberg, K., & Dittrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, 2(2), 107–114. <https://doi.org/10.1177/2515245919838781>
- Schmidt, M., Jäger, K., Egger, F., Roebers, C. M., & Conzelmann, A. (2015). Cognitively engaging chronic physical activity, but not aerobic exercise, affects executive functions in primary school children: A group-randomized controlled trial. *Journal of Sport and Exercise Psychology*, 37(6), 575–591. <https://doi.org/10.1123/jsep.2015-0069>
- Schmiedek, F. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*. <https://doi.org/10.3389/fnagi.2010.00027>
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166. <https://doi.org/10.1080/00461520.2015.1036274>
- Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLoS One*, 6(9), Article e24372. <https://doi.org/10.1371/journal.pone.0024372>
- Sella, F., Raz, G., & Cohen Kadosh, R. (2021). When randomization is not good enough: Matching groups in intervention studies. *Psychonomic Bulletin & Review*, 9. <https://doi.org/10.3758/s13423-021-01970-5>
- Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: A meta-analysis. *Pediatric Exercise Science*, 15(3), 243–256. <https://doi.org/10.1123/pes.15.3.243>
- Sink, K. M., Espeland, M. A., Castro, C. M., Church, T., Cohen, R., Dodson, J. A., ... Williamson, J. D. (2015). Effect of a 24-month physical activity intervention vs health education on cognitive outcomes in sedentary older adults: The LIFE randomized trial. *JAMA*, 314(8), 781. <https://doi.org/10.1001/jama.2015.9617>
- Stepankova, H., Lukavsky, J., Buschkuhl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults. *Developmental Psychology*, 50(4), 1049–1059. <https://doi.org/10.1037/a0034913>
- Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychologica*, 140(1), 13–24. <https://doi.org/10.1016/j.actpsy.2012.02.001>
- Stroth, S., Reinhardt, R. K., Thöne, J., Hille, K., Schneider, M., Härtel, S., ... Spitzer, M. (2010). Impact of aerobic exercise training on cognitive functions and affect associated to the COMT polymorphism in young adults. *Neurobiology of Learning and Memory*, 94(3), 364–372. <https://doi.org/10.1016/j.nlm.2010.08.003>
- Taves, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clinical Pharmacology & Therapeutics*, 15(5), 443–453. <https://doi.org/10.1002/cpt.1974155443>
- Tetlow, A. M., & Edwards, J. D. (2017). Systematic literature review and Meta-analysis of commercially available computerized cognitive training among older adults. *Journal of Cognitive Enhancement*, 1(4), 559–575. <https://doi.org/10.1007/s41465-017-0051-2>
- Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS One*, 8(5), Article e63614. <https://doi.org/10.1371/journal.pone.0063614>
- Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality and Quantity*, 40(3), 435–456. <https://doi.org/10.1007/s11355-005-8081-8>
- Vance, D., Dawson, J., Wadley, V., Edwards, J., Roenker, D., Rizzo, M., & Ball, K. (2007). The accelerate study: The longitudinal effect of speed of processing training on cognitive performance of older adults. *Rehabilitation Psychology*, 52(1), 89–96. <https://doi.org/10.1037/0090-5550.52.1.89>
- Vedamurthy, I., Nahum, M., Huang, S. J., Zheng, F., Bayliss, J., Bavelier, D., & Levi, D. M. (2015). A dichoptic custom-made action video game as a treatment for adult amblyopia. *Vision Research*, 114, 173–187. <https://doi.org/10.1016/j.visres.2015.04.008>
- Verhaeghen, P. (2017). *Presence: How mindfulness and meditation shape your brain, mind, and life* (1 edition). Oxford University Press.
- Verhaeghen, P. (2021). Mindfulness as attention training: Meta-analyses on the links between attention performance and mindfulness interventions, long-term meditation practice, and trait mindfulness. *Mindfulness*, 12(3), 564–581. <https://doi.org/10.1007/s12671-020-01532-1>
- Von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, 41(4), 611–624. <https://doi.org/10.3758/s13421-012-0280-7>
- Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, 16(1), 1–12. <https://doi.org/10.1016/j.lindif.2005.06.004>
- Wang, C., Jaeggi, S. M., Yang, L., Zhang, T., He, X., Buschkuhl, M., & Zhang, Q. (2019). Narrowing the achievement gap in low-achieving children by targeted executive function training. *Journal of Applied Developmental Psychology*, 63, 87–95. <https://doi.org/10.1016/j.appdev.2019.06.002>
- Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology*, 30(2), 190–212. <https://doi.org/10.1037/neu0000227>
- Wu, S., & Spence, I. (2013). Playing shooter and driving videogames improves top-down guidance in visual search. *Attention, Perception, & Psychophysics*, 75(4), 673–686. <https://doi.org/10.3758/s13414-013-0440-2>
- Young, J., Angevaren, M., Rusted, J., & Tabet, N. (2015). Aerobic exercise to improve cognitive function in older people without known cognitive impairment. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD005381.pub4>
- Yung, A., Cardoso-Leite, P., Dale, G., Bavelier, D., & Green, C. S. (2015). Methods to test visual attention online. *Journal of Visualized Experiments*, 96, Article 52470. <https://doi.org/10.3791/52470>
- Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuhl, M., ... Bavelier, D. (2021). Action video game play facilitates “learning to learn”. *Communications Biology*, 4(1), 1154. <https://doi.org/10.1038/s42003-021-02652-7>