A novel task and methods to evaluate inter-individual variation in audio-visual associative learning

Angela Pasqualotto, Aaron Cochrane, Daphne Bavelier, Irene Altarelli

Cognition 242 (2024) 105658

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

A novel task and methods to evaluate inter-individual variation in audio-visual associative learning

Angela Pasqualotto [a,b,1], Aaron Cochrane [a,b,1], Daphne Bavelier [a,b,*], Irene Altarelli [c]

[a] Faculty of Psychology and Education Sciences (FPSE), University of Geneva, Geneva, Switzerland
[b] Campus Biotech, Geneva, Switzerland
[c] Université Paris Cité, LaPsyDÉ, CNRS, Paris, France

ARTICLE INFO

Keywords:
Audio-visual learning
Associative learning
Learning rate
Cognitive correlates
Working memory
Individual differences

ABSTRACT

Learning audio-visual associations is foundational to a number of real-world skills, such as reading acquisition or social communication. Characterizing individual differences in such learning has therefore been of interest to researchers in the field. Here, we present a novel audio-visual associative learning task designed to efficiently capture inter-individual differences in learning, with the added feature of using non-linguistic stimuli, so as to unconfound language and reading proficiency of the learner from their more domain-general learning capability. By fitting trial-by-trial performance in our novel learning task using simple-to-use statistical tools, we demonstrate the expected inter-individual variability in learning rate as well as high precision in its estimation. We further demonstrate that such measured learning rate is linked to working memory performance in Italian-speaking ($N = 58$) and French-speaking ($N = 51$) adults. Finally, we investigate the extent to which learning rate in our task, which measures cross-modal audio-visual associations while mitigating familiarity confounds, predicts reading ability across participants with different linguistic backgrounds.

The present work thus introduces a novel non-linguistic audio-visual associative learning task that can be used across languages. In doing so, it brings a new tool to researchers in the various domains that rely on multisensory integration from reading to social cognition or socio-emotional learning.

1. Introduction

Audio-visual associations are central to many aspects of behavior from mapping sounds to print (e.g., when learning to read) to mapping sounds to emotional faces (e.g., in social cognition; FeldmanHall & Dunsmoor, 2019). Many of our behaviors rely on the process by which auditory stimuli are linked together through exposure with visual ones. In addition, learned associations of auditory and visual stimuli allow for a reduction of uncertainty compared to processing either modality in isolation. Audio-visual associative learning abilities are present very early on in the course of development, including in pre-verbal infants (Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Kersey & Emberson, 2017; Mersad, Kabdebon, & Dehaene-Lambertz, 2021) and can lead not only to explicit, high-level rules of associations between stimuli, but can also influence the low-level, pre-attentive perception of ambiguous stimuli (Kafaligonul & Oluk, 2015; Piazza, Denison, & Silver, 2018; Schmack, Weilnhammer, Heinzle, Stephan, & Sterzer, 2016), even generating predictions at the sensory level (Kersey & Emberson, 2017).

Audio-visual associative learning — like other types of learning — results in changes in behavioral performance thanks to repeated experiences (Harlow, 1949). Indeed, in learning contexts ranging from mental multiplication (Thorndike, 1908) to motor learning to word-list acquisition (Schmidt & Bjork, 1992), and in humans as well as other animals (Gallistel, Fairhurst, & Balsam, 2004), detailed analyses of the time courses of learning can provide valuable insights into learning constraints and affordances. Yet, inferences about associative learning as it unfolds have typically been constrained by the use of behavioral methodologies and associated analytical approaches that are limited in the identification of the full within-subject learning trajectory. Here our focus is on the identification of such learning trajectories, at the individual level, using only a short sequence of trials (under half an hour) to better characterize the processes of learning itself, as opposed to just one

* Corresponding author at: Faculty of Psychology and Education Sciences (FPSE), University of Geneva, and Campus Biotech, Geneva, Switzerland.
E-mail address: daphne.bavelier@unige.ch (D. Bavelier).
[1] These authors contributed equally.

https://doi.org/10.1016/j.cognition.2023.105658
Received 10 August 2022; Received in revised form 24 October 2023; Accepted 31 October 2023
Available online 11 November 2023

of its trajectory points or its putatively stable ending outcome when using longer learning durations. We thus developed a novel audio-visual associative learning task tapping the acquisition of associations among unfamiliar sounds and symbols and paired it with innovative, simple-to-use analytical tools. Of interest is the understanding of how differences in associative learning speed between individuals or populations may relate to other outcomes, such as executive functions or attentional control. Indeed, recently several authors have proposed that such central cognitive functions may allow for more efficient learning (Bavelier & Green, 2019; Miller & Unsworth, 2020; Radulescu, Niv, & Ballard, 2019; Zhang et al., 2021), although this proposal has not yet been fully tested in the context of associative learning.

Our audio-visual associative task and the measures it allows collecting bear several theoretical and empirical implications. On the theoretical side, it is to be noted that distinct sets of computations are involved in different forms of learning, thus the exact task choice matters as to precisely which of those computations are called for (e.g., as argued by Siegelman, Bogaerts, Christiansen, & Frost, 2017 for statistical learning). In its early steps, reading acquisition requires learning the correct associations between letters (or groups of letters) and sounds of language, thus relying on audio-visual associative learning. At least two sets of computations can be highlighted here, namely the encoding of modality-specific inputs and the cross-modal building of associations between these inputs. The choice of learning task we make in the present work therefore reflects a theoretically constrained focus on computations that are relevant for reading acquisition. The choice of unfamiliar stimuli in both auditory and visual modalities is also theoretically motivated, as it further guarantees that these computations are not confounded by the outputs of previous processing – e.g., the encoding of well-known linguistic stimuli.

On the empirical side, the proposed task has the advantage of tracking learning in real time, rather than averaging across trials or estimating learning after it has occurred, which both run the risk of overlooking crucial inter-individual variability throughout the learning process. By examining the evolution of performance on a trial-by-trial basis, we gain a deeper understanding of the underlying processes contributing to performance and how they unfold over time. Indeed, inter-individual differences in learning provide a richer output than simply end performance, such as accuracy reached by the end of the task. Our approach draws upon a rich tradition in behavioral science that puts learning at the center of cognitive processes (e.g., Baddeley & Longman, 1978; Gallistel et al., 2004; Harlow, 1949; Thorndike, 1908), as well as recent interest in the field of reading acquisition about the importance of studying learning progression to better characterize the cognitive mechanisms supporting reading skills (e.g. see the individual trajectories of grapheme-phoneme associations acquisition in Dehaene-Lambertz, Monzalvo, & Dehaene, 2018 as well as Siegelman, 2020 for a discussion of the dynamics and inter-individual variability in statistical learning and their relation to linguistic processes such as reading).

From a methodological point of view, effective characterization of inter-individual differences in learning is not a simple task. One major hurdle of typical methodologies is that they require relatively large amounts of data per learner to reliably estimate trajectories of training-induced change (Schmack et al., 2016: >500 trials; Xu, Kolozsvari, Oostenveld, & Hämäläinen, 2020: 75 mins; Younger & Booth, 2018: 3 × 50 mins). For instance, Xu et al. (2020) asked participants to learn the associations between novel foreign letters and familiar speech sounds on two consecutive days (first day ~50 min; second day ~25 min). Another hurdle concerns proper titration of the learning task difficulty, especially in the face of a limited number of trials, which if not properly implemented may result in many participants being either at floor or at ceiling (for a similar discussion see Siegelman, Bogaerts, & Frost, 2017). Here we present a combination of experimental and analytical methods for quantifying learning trajectories within a behavioral task of fewer than 200 trials, lasting about 30 min, leveraging task constraints to increase the precision of estimates of learning trajectories even when the amount of time on task is limited. We then use simulations to demonstrate the efficiency of the method with even fewer trials.

Another challenge that has prevented inferences about the detailed time course of learning is that many studies exploring associative learning, its cognitive correlates, and/or underlying mechanisms, have used designs where the encoding of associations and their retrieval are separated in time. In such designs, often only retrieval blocks are tested. Participants are typically exposed to associations between cues and outcomes in one or multiple subphases of the experiment, while being tested on their knowledge of them in one or multiple other subphases of the experiment. Periods of passive learning and testing are most often interleaved (see, for instance, Gonzalo, Shallice, & Dolan, 2000; Xu et al., 2020; Younger & Booth, 2018). In such cases, the collected data provides only a discontinuous assessment of participants' learning trajectories, leading to limitations we address in the current work.

Given the goal of modeling the unfolding of learning in a short period of time, it is necessary to analyze the data in such a way that leverages the information provided by each trial's data, rather than treating blocks of trials as independently and identically distributed samples of performance. In our recent work, we have shown that by modeling the trajectory of learning using performance as a continuous function of time, along with implementing computational constraints provided by the experimental methods, separate components of learning trajectories (e.g., rate of change; asymptotic performance) can be reliably estimated even from fairly short behavioral tasks (Cochrane & Green, 2021a, 2021b; Cochrane & Green, 2023; see also Zhang, Zhao, Dosher, & Lu, 2019). In a less constrained approach, such as calculated performance estimates over small blocks of trials, estimated performance may appear to get better or worse simply due to noise rather than true changes in knowledge or skill. To the extent that learning can be modeled as a monotonic improvement in knowledge or skills, this provides robustness against shorter-term fluctuations due to noise or other transitory confounding factors (Kattner, Cochrane, & Green, 2017).

Additional constraints can allow for even short behavioral tasks to provide accurate estimates of performance. First, consider a situation wherein the participant completing a task was completely unfamiliar with the correct categorizations prior to beginning the task. It must necessarily be true, then, that their performance was at chance on the first trial (i.e., when they had no information about correct performance). In this case, it makes no sense for a researcher to treat a set of trials (e.g., a block of the first 32 trials) as interchangeable samples of performance (as with the block-analysis methods mentioned in the following paragraph), but instead as a trajectory of change can be modeled that starts at chance performance. Such an assumption is especially powerful when, initially - at the start of the task - all participants' performance estimates can be constrained to the same point, thereby providing a point of ground-truth equivalence that can facilitate the interpretation of later-task divergences between participants. The empirical estimation of performance is likewise assisted by a ground-truth anchoring point as well, which makes model fitting and comparison more tractable. Further, consider a simple question: Is a person performing a task with an accuracy above chance? If that participant responded incorrectly on the first trial, but correctly on every subsequent trial, in a two-alternative forced-choice categorization task it would take 8 trials total for a one-tailed binomial test to conclude that the participants' performance was significantly above chance. In contrast, in a three-alternative forced choice task it would only take 5 trials total for such a conclusion to be made (i.e., 62.5% of the number of trials). Increasing the number of alternatives leads each correct trial to be less likely to be due to chance (Shelton & Scarrow, 1984; Vancleef et al., 2018).

Estimated trajectories of learning may be constrained in additional ways, for example, by using informative priors in Bayesian estimation. Certain parameters may even be fixed to plausible values (e.g., if learners are likely to completely master a task, asymptotic performance could be fixed to 100% accuracy). These latter constraints provide

further ease of estimating other values of interest and reduce the likelihood of trade-offs in correlated parameters. They cannot be known a priori, however, and thus differ greatly from constraints such as fixing starting accuracy at chance. Rigorous comparisons between models with or without such constraints should instead be used to justify the imposition or relaxation of assumptions regarding learning trajectories.

Unfortunately, and to the best of our knowledge, no example exists of the application of such methods in the field of associative learning, or at least audio-visual forms of such learning. Indeed, many studies examined progress in the task by averaging accuracy over subsections of the data (e.g., mean accuracy and RTs in the 1st, 2nd, 3rd and 4th quarters of the experiment, as in Hämäläinen, Parviainen, Hsu, & Salmelin, 2019; mean accuracy and median RTs in each learning block, as in Madec et al., 2016). In the same vein, Younger and Booth (2018) used latent growth curve modeling to estimate the overall change across measurement points, that is, testing blocks over three days; nonetheless, within-block across-trial aggregation of performance still reduced the possibility of identifying learning on shorter scales of time.

Apart from examining learning achievement in discrete learning episodes through time, one can also estimate the time needed to reach a certain level of performance. Karipidis et al. (2017) used this method, namely examining the number of training blocks needed to reach a criterion accuracy, in order to estimate the learning rate. Such methods are not uncommon, yet they remain reliant on within-block aggregation of performance while also providing only a single point estimate of the time taken to learn. That is, the estimation of the time taken to learn does not take into account full trajectories of performance change, thereby limiting the available inferences from such data (e.g., using model comparisons of learning). In sum, in previously published studies, performance is computed by averaging over large chunks of data, assuming constant performance within blocks or testing phases. As such, rather indirect estimations of learning speed are gathered. Such assumptions of performance stationarity within blocks and changes between blocks imply a corresponding assumption of process-level (i.e., knowledge or ability) stationarities and disjunctions, which is both contrary to theoretical views of learning but also may bias inferences regarding the learning that occurred (Kattner, Cochrane, Cox, Gorman, & Green, 2017; Kattner, Cochrane, & Green, 2017). Interestingly, even in associative learning studies that have collected trial-by-trial data, inferences regarding learning have been mostly limited by aggregating across trials (Barutchu, Fifer, Shivdasani, Crewther, & Paolini, 2020).

In order to address the issues described above, and to best serve our goal of modeling progress in audio-visual associative learning continuously, here we developed a novel paradigm where participants were asked to learn arbitrary associations between unfamiliar sounds and symbols, and output a response on every trial. We then applied a fully continuous-time (i.e., trial-by-trial accuracy) model to the data.

First, in order to establish the constraint that participants began learning at a chance level, we used environmental sounds unknown to participants as well as unfamiliar symbols (derived from an archaic non-European alphabet). Using stimuli unknown to participants in both auditory and visual modalities ensured avoiding any familiarity unbalance between modalities, contrary to many previous studies (e.g., Schmalz, Schulte-Körne, De Simone, & Moll, 2021; Xu et al., 2020; Younger & Booth, 2018).

Progress in our audio-visual associative learning task was modeled by fitting a fully continuous-time (i.e., trial-by-trial) model to the data. We followed a methodological approach which prioritizes direct characterizations of performance trajectories in terms of individual learners' parameters of interest, such as learning rate and asymptotic level reached (Crossman, 1959; Dosher & Lu, 2007; Kattner, Cochrane, Cox, et al., 2017; Newell, Liu, & Mayer-Kress, 2001; Newell & Rosenbloom, 1981). Treating accuracy as continuously varying over time allowed us to identify inter-individual differences from dissociable components of learning (e.g., rate of learning, acceleration, or asymptotic performance) and to provide methodologically-driven constraints to our data while estimating theoretically-meaningful parameters of learning. We applied such methods to two distinct samples, 58 Italian-speaking and 63 French-speaking healthy adults, for internal replication of results in two different languages. Importantly, these methods use a statistical package that requires minimal technical understanding of the underlying modeling approach; a single function in **R** fits the full nonlinear mixed-effects model, from which group-level effects, model predictions, individual-level effects, and model diagnostics can all be extracted (Cochrane, 2020). Recovery analyses indicated that inter-individual differences were able to be estimated with high precision.

Additionally, to confirm that our novel audio-visual learning task was capturing meaningful inter-individual variation in learners' abilities, we collected performance in a series of cognitive tasks in the same participants. Specifically, the selected measures encompassed both domain-specific abilities related to language (i.e., speed and accuracy of reading) and more domain-general skills (i.e., fluid intelligence, working memory and attentional control). Despite the importance of determining the processes implicated in the acquisition of audio-visual mappings (e.g., for reading), very few studies investigated their cognitive correlates, particularly when the to-be-learned pairings contained no linguistic information (but see, Altarelli, Dehaene-Lambertz, & Bavelier, 2019 in children). Finally, through the assessment of reading skills in the same participants, we delved into the putative relation between reading skills, nonlinguistic associative learning, and domain-general skills. By investigating these interconnected factors, we aim to contribute to the growing body of literature exploring the complex cognitive processes underlying reading.

In the present study, we addressed several prior approaches' weaknesses by implementing a fully non-linguistic audio-visual associative learning environment. In order to extract learners' trajectories of associations' acquisition on a trial-to-trial timescale, we modeled continuously-changing indices of performance.

2. Methods

2.1. Participants

2.1.1. Italian-speaking sample

Seventy-five native Italian-speaking adult participants took part in this behavioral study. Criteria for inclusion were the following: (i) no diagnosis of psychological/neurological disorders; (ii) no reading delay in word, non-word and text reading tasks (−1.5 sd from published norms); (iii) reported normal or corrected-to-normal vision and hearing, (iv) intelligence within the normal range (cut-off score ≥ 7) as measured with the WAIS-IV subtest of Matrix Reasoning (Wechsler, 2008). This latter cut-off score led to the removal of three participants; thus, data from seventy-two (72) participants remained (40 females, mean age: 25 years old).

2.1.2. French-speaking sample

Sixty-four French-speaking adult participants participated in the study. Criteria for inclusion were the following: (i) no diagnosis of psychological/neurological disorders; (ii) no reading delay in word, non-word and text reading tasks (−1.5 sd from published norms); (iii) reported normal or corrected-to-normal vision and hearing. One female participant was excluded from the analysis due to undergoing an ADHD diagnosis process. In total, data from sixty-three (63) participants remained (53 females, mean age: 25 years old).

In Supplementary Section 3, we have included an additional sample consisting of fourty-one French-speaking participants. For this group, data was collected solely for domain-general tasks and our novel audio-visual learning task.

In both samples, written informed consent was obtained prior to participation. The study was approved by the research ethics committee of the University of Trento (IT) and that of the University of Geneva (CH), respectively.

2.2. Design and materials

Both Italian and French-speaking participants were tested individually in one, 1.5 h-long session. In addition to the novel audio-visual associative learning task, our comprehensive battery of tasks encompassed assessments of reading skills, domain-specific abilities such as phonological awareness and rapid automatized naming, as well as domain-general skills including auditory and visual attention, and short-term and working memory skills. It is important to note that the battery of tasks slightly differed between Italian-speaking and French-speaking participants, with measures of phonological awareness and rapid automatized naming only collected in the Italian-speaking group. The specific tasks administered to the participants are detailed below.

All computerized tasks were administered on a 13″ screen and headphones were used when auditory stimuli were involved. The Multiple Object Tracking task was programmed using Matlab v. R2017, Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The audio-visual associative task was programmed using PsychoPy v1.82 (Peirce, 2007). The Auditory Attention task was programmed using Javascript. Finally, the Odd-one-out task was provided by Creyos (see, Hampshire, Highfield, Parkin, & Owen, 2012).

2.2.1. Audio-visual associative learning task

A novel audio-visual associative learning task was created for this study. The overall goal of the participant was to learn associations between pairs of unknown auditory stimuli (environmental sounds, 8 in total) and novel visual symbols (6 in total), as depicted in Fig. 1. Because participants had no pre-existing knowledge of the arbitrary associations in our task, they all presented the same starting point, that is, performing initially at chance, enforcing a common starting point across all subjects for our modeling. A similar version of the task, yet simplified to be suitable for preschool children, can be found in Altarelli et al. (2019).

The task administered here consisted of three main parts, in which the first two constitute stages – of very short duration – of familiarization with auditory and visual stimuli, respectively. In the first part (duration: 2.5 min; 16 trials), participants familiarized themselves with the auditory stimuli by passively listening to each of them be presented one at a time. All sounds were environmental sounds that were unfamiliar to the participants yet easily discriminable, as demonstrated in previous studies using similar stimuli (Seitz, Kim, van Wassenhove, & Shams, 2007). This choice ensured avoiding confounds related to varying levels of pre-existing sound familiarity among participants. Indeed, when linguistic stimuli are used, participants' oral language abilities (e.g., phonological skills) may influence their progress in the audio-visual associative learning task, interfering with the measure of associative learning per se. For the purpose of the audio-visual associative learning task, sounds were always presented as a sequence of two sounds (total duration: 3850 ms), as illustrated in Fig. 1A. The use of paired sounds allowed us to efficiently titrate the difficulty of the task without significantly inflating its overall duration.

The second part of the task (duration: 4.5 min) consisted of a symbol familiarization phase in which the 6 symbols, adapted from the Bamum alphabet and unknown to the participants, were presented in the context of a 1-back task to ensure attention throughout. Participants were required to respond to a stimulus only if it matched the stimulus that immediately preceded it. Each symbol was presented for 1 s, with a 1 s inter-stimulus interval. The task comprised 80 trials, with 12 1-back repetitions (2 of each symbol). The sequence of symbols was the same for all participants.

Finally, participants underwent the audio-visual associative learning task for six blocks of under five minutes each. Participants were told they would be taught an unfamiliar language by listening to the sounds uttered by an alien and choosing which is the correct symbol for each pair of sounds, in parallel to what learning to read requires. In order to be learned, half of the audio-visual pairings required the participants to pay attention to both sounds (difficult trials), whereas in the other half of the audio-visual pairings, only the last sound was crucial (easy trials; for examples of the audio-visual pairs, see Fig. 1A). The task consisted of 6 blocks of 32 trials each with difficult and easy trials in pseudo-randomized order. In each trial (see Fig. 1B), one of the 8 pairs of auditory stimuli was presented, followed briefly by a blank screen (jitter duration: 1500, 2250, or 3500 ms). Then three response options were presented until the participant responded and for a maximum of 2300 ms (first block: 3000 ms). The location of the symbols changed from one trial to the next, to avoid spatial learning. Following participants' response on each trial, feedback was provided, indicating the correct symbol (duration: 1500 ms). All participants completed the task in <32 min, including the breaks they could take between blocks.

At the end of the audio-visual associative learning task, participants underwent a short additional task (8 trials) which assessed the extent to which only the second sound was discriminative of the response symbol. In each of the new trials, the same final sound that was previously used

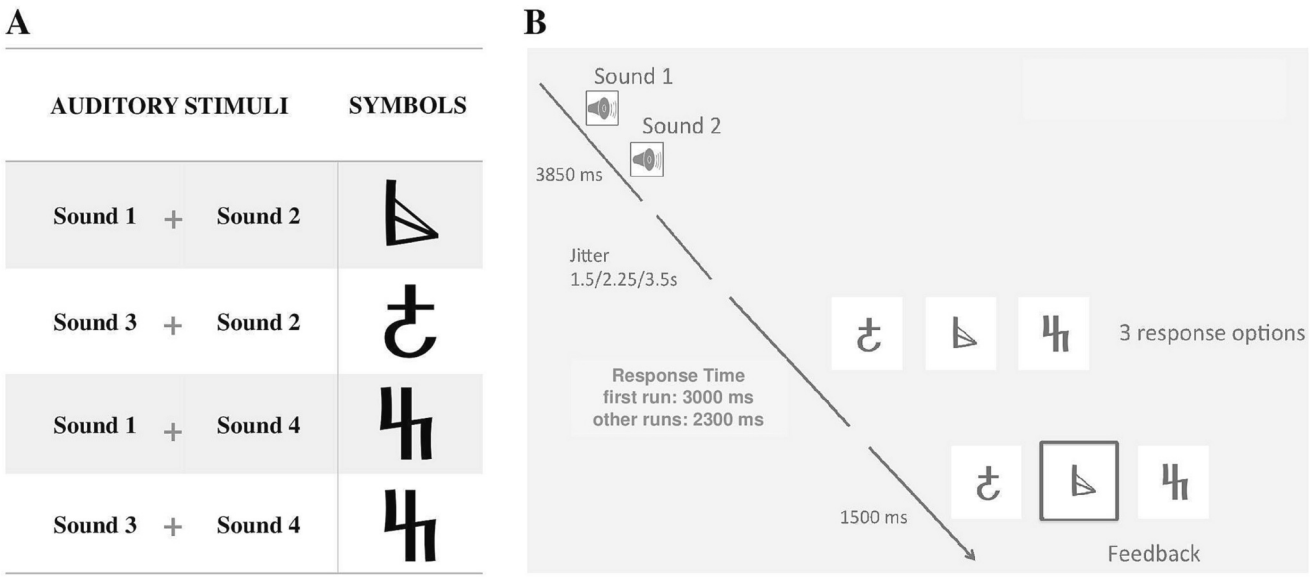

Fig. 1. A. Audio-visual associative learning task: examples of four of the to-be-learnt audio-visual pairs. **B.** Each trial began with the presentation of a pair of auditory stimuli (3850 ms overall) followed by a variable ISI and then three response options. After the participant's response, visual feedback was presented.

in the easy trials was inserted to assess the generalization of the easier association rule.

2.2.2. Italian-speaking sample

2.2.2.1. Reading skills. The reading task employed utilized lists of words and non-words from the VALS, which is the Italian adaptation of the EVALAD battery (Pech-Georgel & George, 2011). Participants were instructed to read aloud the presented lists of words and pseudo-words (as defined by legal sequences of maximum four syllables that do not correspond to actual words in Italian). For each participant, z-scores were calculated for both reading speed (measured in seconds) and reading errors (each incorrect word or non-word counted as one error).

2.2.2.2. Reading-related measures. Phonological awareness skills were evaluated through a phoneme deletion task and a phoneme blending task. Lexical access was tested through a rapid picture naming task (Rapid Automatized Naming; RAN of objects). In all of these tests, derived from the VALS battery (Pech-Georgel & George, 2011) - two scores were calculated: the accuracy and the speed with which the subject performs the test.

2.2.2.3. Domain-general skills. Fluid Intelligence was assessed by means of the Matrix Reasoning subtest of the WAIS-IV scale (Wechsler, 2008). From the EVALAD battery (Pech-Georgel & George, 2011), we administered a Listening Span task to assess both working memory and resistance to interference. In this complex span task, participants listened to sentences, verified their semantic plausibility, and then repeated the last word of each of the sentences in the set. The length of each set of sentences increased throughout the task. For each of the Span tasks, an accuracy score was calculated. In addition, the forward and backward Digit Span tasks were also administered, in order to assess participants' verbal short-term and working memory skills using again the the EVALAD battery.

An adaptation of the Multiple Object Tracking task (MOT – Pylyshyn & Storm, 1998) was used to evaluate attentional control in a dynamic setting. Participants were asked to track objects as they moved around the screen. Objects were initially still, with some presented as targets (sad blue face) and others as distractors (happy yellow face). After this initial period allowing participants to identify the to-be-tracked targets (200 ms), all objects turned back to happy yellow faces. Participants were required to keep track of the initially blue ones for a duration of 400 ms. After that tracking period, the objects stopped moving, one of them was flagged and participants had to decide whether this specific object had initially been a target (blue) or a distractor (yellow). Accuracy score reflects participants' attentional control skills, including the capacity to update object information in working memory.

Auditory selective attention was assessed by means of a novel task, based on the paradigm proposed by Hansen and Hillyard (1980). Two series of sounds were presented simultaneously to participants through headphones, differing in terms of duration and frequency. The participant had to identify and respond only to the low-pitch and long sounds (i.e., 300 Hz, 102 ms), by pressing a button after each occurrence. Performance in this task was evaluated by calculating a value of d' for each participant (using Bendixen & Andersen, 2013, for computing d' in high event rate paradigms).

2.2.3. French-speaking sample

2.2.3.1. Reading skills. Reading skills were assessed with word and pseudo-word decoding tasks from the VALS battery (Pech-Georgel & George, 2011). As for the Italian sample, z-scores were calculated for both reading speed (measured in seconds) and reading errors (each incorrect word or non-word counted as one error).

2.2.3.2. Domain-general skills. The same domain general constructs as in the Italian sample were evaluated but at times using different assessment. A computerized version of the Odd-one-out task (Hampshire et al., 2012) was used to assess fluid intelligence. During each trial, participants had to identify which of the presented figures (varying in color, shape, and number of sub-figures) should be excluded. Participants had to solve as many trials as possible within 3 min and earned one point for each correct answer (score: maximum number of problems solved).

The digit span subtest from the WAIS-IV (Wechsler, 2011) was administered to assess forward and in backwards digit span. The same MOT and auditory attention tasks were otherwise used as in the Italian sample.

2.3. Data analysis

2.3.1. Audio-visual learning task

The audio-visual associative learning task was first screened for participants who did not demonstrate learning. Participants were excluded according to two different exclusion criteria: either percent correct was not significantly above chance on the last $^1/_3$ of trials (64 trials; chance performance of 33%; assessed using a one-tailed binomial test) or their accuracy decreased with time (i.e., mean accuracy from the first 64 trials to the second, or from the second 64 trials to the third). With partially overlapping sets of participants, this led to 10 out of 72 Italian-speaking participants being excluded, and 6 out of 63 French-speaking participants being excluded.

The data was then fit with mixed-effects Bayesian nonlinear regression using the **TEfits** package in R (Cochrane, 2020; see Supplement for model code) which itself utilizes Stan via the **brms** package (Bürkner, 2017; for similar modeling approaches see Dale, Cochrane, & Green, 2021). In short, this nonlinear learning model approach estimated the trial-by-trial improvement in percent correct from the chance performance (i.e., 33% accuracy) on the first trial through some above-chance performance on the last trials. Learning took the form of improvement in accuracy as a nonlinear function of trial numbers (Cochrane & Green, 2021a; Dosher & Lu, 2007; Heathcote, Brown, & Mewhort, 2000). We fit three such models, with different nonlinear functions of change giving rise to learning, and used Bayesian model comparison to adjudicate between the possible underlying trajectories of change. These included a 3-parameter power function (Eq. 1), a 3-parameter exponential function (Eq. 2), and a 4-parameter Weibull function (an augmentation of the 3-parameter exponential function; see Eq. 3).

$$accuracy = asymptote + (start - asymptote) \times trialNumber^{rate} \quad (1)$$

$$accuracy = asymptote + (start - asymptote) \times 2^{(1-trialNumber)/rate} \quad (2)$$

$$accuracy = asymptote + (start - asymptote) \times 2^{((1-trialNumber)/rate)^{shape}} \quad (3)$$

In Eqs. 1 and 2 there are three free parameters, however, in all three functions' models we fixed *start* to $^1/_3$ or the expected chance performance when first encountering the task. Then, within the nonlinear mixed-effects model, the trajectories of accuracy change defined by *asymptote, rate* [a time constant associated with a fixed percent of change from start to asymptote], and *shape* [acceleration] were simultaneously estimated within generalized linear mixed-effects models (see Eqs. 4, 5 & 6, in **brms/lme4** "Wilkinsen" model notation)

$$logit(asymptote) \sim Intercept + trialDifficulty + language + (trialDifficulty \mid participant) \quad (4)$$

$$log2(rate) \sim Intercept + trialDifficulty + language + (trialDifficulty \mid participant) \quad (5)$$

$$log(shape) \sim Intercept + (1 \mid participant) \quad (6)$$

Estimation of parameters on logit or log scales provided the ability for parameter estimates to, in principle, vary along all real numbers, while constraining the trajectories of learning to have accuracies [*asymptotes*] bounded at [0,1] while time constants [*rates*] and *shapes* were bounded to positive reals. All priors were defaulted within **TEfits**. A Bernoulli response distribution was used due to the by-trial binary accuracy being modeled, and all models were run for 20,000 iterations, discarding the first 5000 iterations as a warm-up.

We also fit models with an additional constraint, namely, that all participants' performance on all trial types would asymptote at 99.9% correct (i.e., perfect stimulus-response learning, with a 0.001 lapse rate; Wichmann & Hill, 2001). In these alternative models, the only learning parameter was therefore the time constant of change, or *rate* (parameterized as described above), and in the Weibull model there was the additional by-participant *shape* parameter.

A last model was fit for descriptive purposes, in order to assess some of the assumptions in the learning models described above. In particular, instead of a parameterized trajectory of change over time, this model converted the linear vector of trial numbers into a matrix of overlapping normalized basis functions with centers placed every 20 trials (see Appendix for details). These basis functions, entered as linear predictors in a mixed-effects model, approximate arbitrary nonlinear changes in performance. Allowing flexibility in the shape of the estimated trajectories of performance, we could cross-check our assumptions of monotonic increases implemented in the previous models with parameterized functions of change.

Models were compared using Bayes Factors estimated using bridge sampling. Due to the possible instability of bridge sampling estimates, we ran the bridge sampling algorithm 15 times for each relevant model comparison and used the median Bayes Factor for each. We conducted model comparisons in order of increasing complexity, first comparing the two simpler models from different families, the 3-parameter power and exponential models (with only a single free parameter, rate, for these models; for discussion of mechanistic differences between families of learning, see Dosher & Lu, 2007; Heathcote et al., 2000; A. Newell & Rosenbloom, 1981). We next compared the "winner" of this model comparison to a version of the analogous model with a freely-estimated asymptote. Last, we compared the "winner" of the second model comparison to the more highly-parameterized learning trajectory, that which used the Weibull function.

Given our initial adjudication between functional forms of learning, we used the best-fitting model in subsequent tests of individual differences. Estimated model fixed effects were first assessed. Then, point estimates of by-participant (i.e., random-effects) parameters were extracted from the model, and these parameters were subsequently used in tests of relation to other measures (e.g., attention or memory; see Cochrane & Green, 2021a, 2021b; Dale et al., 2021).

Last we used a simulation-based approach to determine the sensitivity of our analyses to individual-level variations and to identify a lower bound on the experimental method's number of trials. We simulated true learning trajectories for new participants given the means and standard deviations of participant-level parameter point estimates, with each simulated dataset to test recovery varying in number of trials (from 15 to 200 in 5-trial increments) and number of participants (from 10 to 100 in 5-participant increments). After randomly choosing a combination of trial numbers and participant numbers from this grid of possible values, and sampling participant-level parameters from normal distributions matching the empirical participant-level distributions, latent participant-level interleaved easy-trial and difficult-trial curves were generated. Given these participant-level learning curves (which were themselves accuracy percentage values), accuracies were generated for each trial by sampling from a Bernoulli distribution with an expected value of the learning curve's value at that trial. We then fit an identical model to the main results, except for the removal of the linguistic background covariate, to each of the simulated samples of participants. Nearly all models converged (with all r-hat below 1.05), and only converged models ($n = 1024$) were included in the final analyses.

2.3.2. All other skills

Within each sample, multivariate outliers were screened entering all available cognitive and reading variables and using the robust Mahalanobis distance method of Leys and colleagues (Leys, Klein, Dominicy, & Ley, 2018). Robust covariance estimation utilized a minimum of 90% of the sample, with outliers being identified using a chi-square cutoff with $\alpha = 0.01$. This led to 4 Italian-speaking participants and 6 French-speaking participants being excluded due to being multivariate outliers (final sample size 58 Italian-speaking, 51 French-speaking). Note that for the Italian sample, the phonological awareness measures collected through RT-based tasks were first independently converted into Inverse Efficiency Scores (i.e., RT/accuracy).

Then, Yeo-Johnson power transformations were applied, with Yeo-Johnson λ optimized to minimize the univariate skew of each variable. Variables were next z-scored. Last, separate composite measures were calculated for working memory and attentional control using the dominant component from a PCA of respectively the two working memory scores, and of the two attentional control measures. The composite represented the underlying dimension that accounted for the highest amount of variance across the two measures used within each respective domain.

Our initial tests of individual differences in audio-visual associative learning used bivariate product-moment correlations with 3000 bootstrap resamples to determine confidence intervals. Next, given the links we observed between learning and reading, we used mediation models to assess the extent to which learning abilities' predictiveness of reading ability was attributable to working memory as expected, and possibly attentional control. Mediations were fit controlling for age, sex, and language group. Bias-corrected and accelerated confidence intervals were estimated using a nonparametric bootstrap with 2000 iterations.

3. Results

3.1. Model comparisons of functional form of learning

We first tested the relative evidence for the two models with fixed asymptotes, namely, the exponential-function model and the power-function model (see, e.g., Dosher & Lu, 2007; Heathcote et al., 2000). Each of these models had starting values fixed to 33.3% and asymptote values fixed to 99.9%. Over 15 estimations of the Bayes Factors, estimated using bridge sampling (Gronau, Singmann, & Wagenmakers, 2020), there was decisive evidence in favor of the exponential model over the power model (in fact, the evidence ratio was so large that it was unable to be estimated precisely, resulting in an infinite Bayes Factor).

We next tested whether the better of the previous two models, the model of exponential change from a chance-level start (i.e., 33.3%) to a fixed asymptote, was improved by instead allowing the asymptote parameter to be freely estimated. Bayes Factors indicated slightly more evidence for the fixed-asymptote model over a model with freely-estimated asymptote parameters (over 15 bridge sampling runs, median $\log_3$ Bayes Factor was 0.16). This indicated that the variation in learning was more likely primarily due to variations in learning rate, and not in asymptotic performance.

Given the above evidence for the fixed-asymptote exponential model of learning, we next tested one additional model within an exponential family of functional changes. An additional "shape" parameter, augmenting the exponential function, leads to a Weibull function of change; because the exponential function is nested within the Weibull function,

the exponential function is a special case of the Weibull function (i.e., with a shape parameter fixed a priori). There is evidence in previous work that relaxing this assumption and allowing the Weibull function's shape parameter to be estimated provides better fit to learning data (Cochrane & Green, 2021b; Gallistel et al., 2004; Leibowitz, Baum, Enden, & Karniel, 2010). Indeed, when allowing by-participant estimates of Weibull shape to augment the previous exponential (with a fixed starting and asymptotic accuracy), we observed decisive evidence for the Weibull model over the more restricted exponential model (over 15 runs, median $BF_{log3} = 36.90$). As such, we used the fixed-asymptote Weibull model in all further analyses. After selecting the learning model using the above comparisons, we extracted by-participant point estimates of parameters.

The fixed-asymptote Weibull model also fit better than the descriptive model with basis functions allowing for flexible estimation of changes over time (median $BF_{log3} = 331.10$). This was as we expected, since the descriptive model had many more parameters than the other models and was therefore likely to be overfit (e.g., due to estimating a starting performance level rather than setting it to the a priori level of 0.333). Still, this result provides support for various assumptions implemented in the primary models, such as monotonic increases in performance (see also Fig. S1).

3.2. Mixed-effects nonlinear model results

The estimated time to half of learning was 82.6 trials in the Italian-speaking sample and 87.9 trials in the French-speaking sample (see Table 1 and Fig. 2). Due to the fixed starting and asymptotic accuracies, these numbers of trials indicated the amount of time necessary to reach an accuracy of $^2/_3$. As expected, reliable modulations of learning rate in response to difficulty manipulations were present ($b = -0.37$, $CI_{95} = [-0.50,-0.23]$). The Weibull shape parameter was parameterized such that a value of zero would be equivalent to the simpler 3-parameter exponential function, and that larger values would indicate a sigmoid shape (i.e., slow start and acceleration) to learning. The fixed effect was reliably higher than zero, further providing evidence for an acceleration of learning (i.e., sigmoid shape; $b = 0.26$, $CI_{95} = [0.16,0.36]$). Upon visual inspections of the model fits, predicted accuracy closely followed participants' accuracy over time (see Fig. 3).

3.3. Italian-speaking sample: Correlates of learning

Of the possible correlates of learning rate, both the working memory composite ($r = -0.33$, $CI_{95} = [-0.52,-0.09]$) and reading accuracy ($r = -0.41$, $CI_{95} = [-0.61,-0.19]$) showed a reliable bivariate association (see Fig. 4 and Fig. S3).

3.4. French-speaking sample: Correlations

In the French-speaking sample learning rate correlated again, albeit not reaching statistical reliability, with the working memory composite ($r = -0.18$, $CI_{95} = [-0.43,0.08]$) and reading accuracy ($r = -0.23$, $CI95 = [-0.48,0.05]$, as well as reliably correlating with the attention control composite ($r = -0.24$, $CI_{95} = [-0.44,-0.01]$ - see Fig. 5 and Figs. S4; see also Fig. S5 for the additional French-speaking sample).

Table 1
Model of learning: Fixed effects of the Weibull-change fixed-asymptote model.

	Estimate	Lower 95% CI	Upper 95% CI
Learning rate (Intercept)	6.05	6.09	6.36
Learning rate (Difficulty)	−0.37	−0.50	−0.23
Learning rate (Sample)	0.28	0.01	0.55
Learning shape (Intercept)	0.26	0.16	0.36

Note. See Eqs. 5 and 6 for the fixed-effects specifications.

3.5. Testing working memory and attention control as mediators of the relationship between learning rate and reading accuracy

To assess the extent to which working memory, and possibly attention control, could account for links between learning rate and reading accuracy, the Italian- and the French-speaking sample were combined, and two mediation models were tested (see Fig. 6). Each model controlled for participants' age, sex, and language. While working memory scores reliably mediated 20% of learning ability's predictiveness of reading accuracy (indirect effect $b = -0.081$, $CI_{95} = [-0.182,-0.021]$), attention control scores did not reliably mediate the link (indirect effect $b = -0.036$, $CI_{95} = [-0.125,0.02]$). In all cases, the direct effect remained reliable, with the time taken to learn continuing to predict reading accuracy.

3.6. Recovery analyses: Sensitivity of methods to individual differences

A core question addressed in these two studies involved the extent to which inter-individual differences in learning could be rapidly assessed. Our behavioral results demonstrated that individual differences in learning were reliably measured and related to cognitive measures. As an additional demonstration of our ability to capture learning, we used the empirical model to simulate 1024 new datasets and applied the same analytical approach to these new datasets. The ability to recover inter-individual differences in learning (i.e., corresponding to the point estimates on the y-axis of Fig. 4 and Fig. 5) was then assessed.

Recovery of inter-individual differences was quite good (see Fig. 7). Participant-level Spearman rank correlations between estimated and true generative parameters were fairly insensitive to participant number, and tended to be very high with trial numbers over about 112. This means that, in order to identify inter-individual variations in our sample with an assurance that the estimates would have at least a correlation of 0.9 with the generative learning rates, we would have only needed to measure four learning blocks rather than six. We also assessed estimation bias and confirmed that, with four or more blocks of trials, there was very little systematic difference between generative and estimated parameters (see Supplementary Fig. S2). In total, these results show that our experimental methods are efficient and robust even at sample sizes and trial numbers smaller than our empirical sample.

4. Discussion

Though audio-visual associative learning is a crucial process in development and across the lifespan, very few studies have focused on finely characterizing each participant's learning performance by modeling progress in audio-visual associative learning continuously. This study set out with the aim of demonstrating that — even in a short task with a limited number of trials — it is possible to obtain precise and efficient individual-level estimates of healthy adults' ability to learn arbitrary associations between novel, non-linguistic, visual stimuli and auditory stimuli. Crucially, continuous-time modeling was applied to our trial-by-trial data, in order to estimate the time taken to learn for each participant – a methodological choice, we argue, that could also be useful in many other learning contexts. In the current study, we gathered data from both an Italian- and a French-speaking sample, for confirmation of the results' language-independent nature.

While many of the previous studies examined learning in a test phase that followed passive exposure to the audio-visual pairs (e.g., Xu et al., 2020), we developed a task that tracks learning progression on a trial-to-trial basis, thus allowing us to fit a fully continuous-time model to the data. Research to date has typically averaged participants' performance over long periods of time (e.g., blocks) and used rather indirect measures of learning speed (e.g., training duration). Such procedures typically assume that learning is constant over large portions of time. Many of the previous approaches thus use questionable assumptions in the estimation of learning rates. With a view to reducing these errors at both the

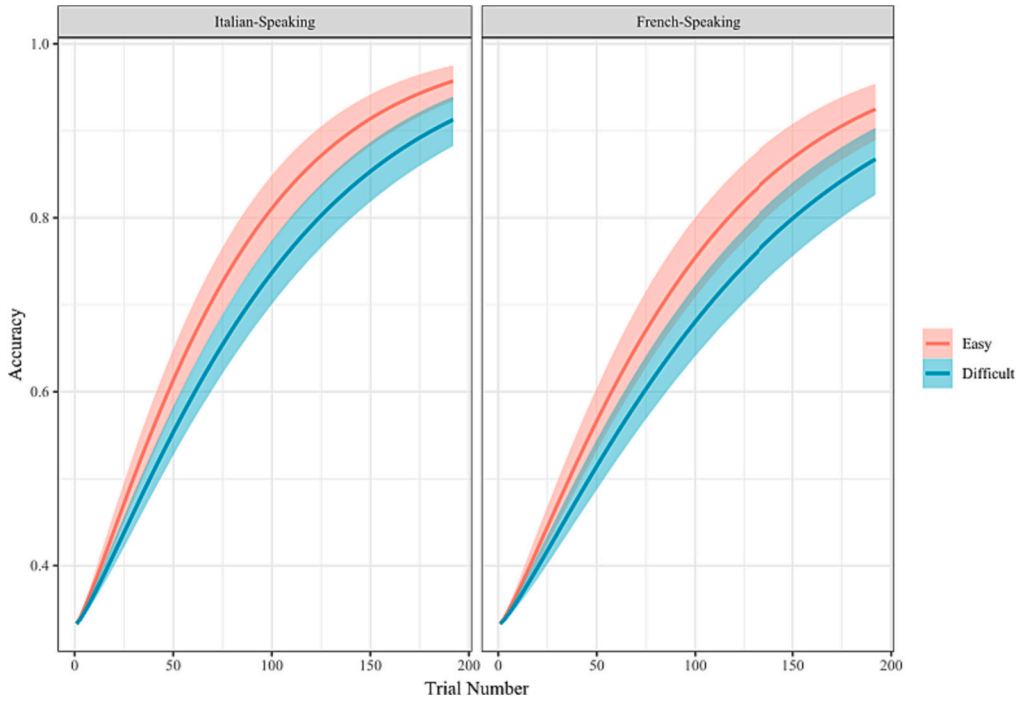

Fig. 2. Group-level effects from each sample. Both samples approached asymptotic performance by the end of the task on easy trials, with reliable difficulty-related effects being evident.

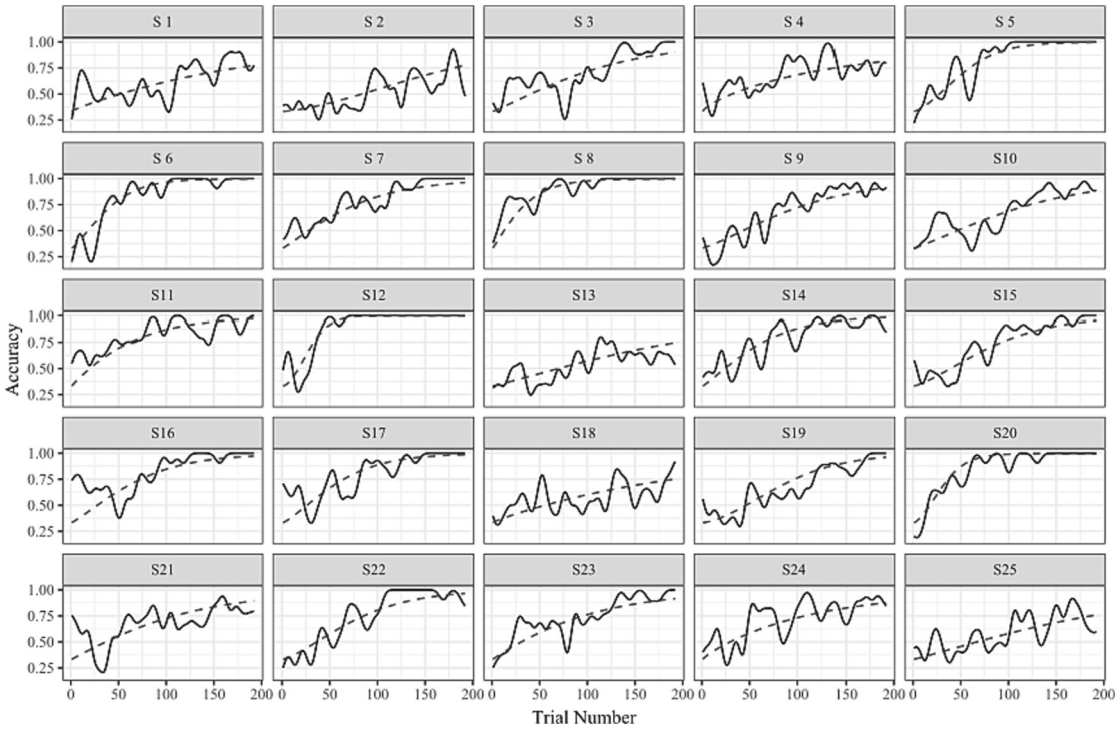

Fig. 3. Example participants' learning trajectories. 25 participants were randomly chosen to plot the raw accuracy (solid black line; smoothed with a Gaussian kernel) and the model fits (red dashed line). A sigmoid shape of learning is evident for some participants, such as S12 and S19. Note that the model fits were evaluated at an intermediate difficulty level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

single-subject level and in group-level analyses, our methodological approach prioritizes direct characterizations of performance trajectories in terms of parameters of interest, such as learning rate and asymptotic accuracy reached by the participant. Nonlinear mixed-effects models provided for the simultaneous estimation of all participants' full trajectories of learning. While these models provided a key performance measure of time taken to learn, which is very similar to various other studies (e.g., Karipidis et al., 2017), our models additionally allow for

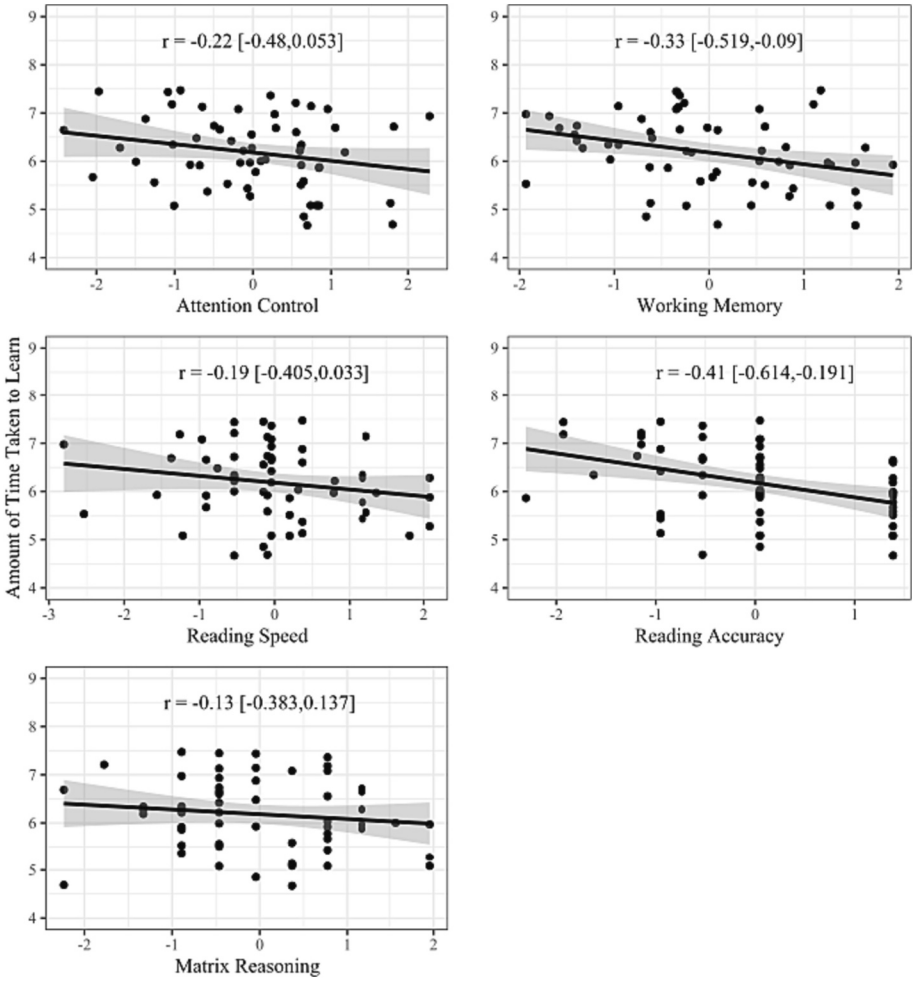

Fig. 4. Correlates of audio-visual associative learning in the Italian-speaking sample with the composite scores of attention control and of working memory, as well as with reading speed, reading accuracy and the measure of fluid intelligence.

trial-to-trial estimates of performance (allowing in principle, e.g., the post hoc comparison of groups or individuals at any point in time) as well as allowing for model comparisons adjudicate between model parameterizations or constraints. Such models also allowed for fully Bayesian model comparisons.

We review below a few key benefits of our theoretically-driven innovative task and analysis methods. First, most previous studies used linguistic stimuli in the auditory domain, like native-language phonemes, to be paired with unfamiliar symbols (as described for audio-visual associative learning in Altarelli et al., 2019). Yet in these cases, familiarity with the auditory stimuli cannot be controlled for, a factor of importance in determining the processes implicated in audio-visual learning (Li et al., 2016). Indeed, these types of learning are scaffolded by a person's previous learning (i.e., prior linguistic knowledge narrows the environmental dimension-exploration greatly, influencing the number of possible answers to a task). In addition, the exact same paradigm and stimuli cannot be applied to participants speaking different languages. The current audio-visual associative learning paradigm introduces non-linguistic, environmental sounds in order to assess the very process of building cross-modal audio-visual associations free of familiarity confounds and in a way that is comparable across participants speaking two different languages (i.e., Italian and French).

Second, one key methodological benefit of our task was that, given the 3AFC nature of the task as well as the stimuli themselves and their associations being entirely novel to participants, initial task accuracy was a priori known to be 33.3% (Kattner, Cochrane, Cox, et al., 2017). This greatly facilitated the specification of a theoretically-constrained model from which to draw inferences regarding learning. While other approaches to trial-by-trial learning could be applied to associative learning, such as reinforcement learning algorithms (Gershman, 2015; Steingroever, Wetzels, & Wagenmakers, 2014), standard implementations of such models would need to be modified in unclear ways in order to provide the constraints and model comparisons reported here (e.g., between accelerating learning functions or non-accelerating learning functions, or between fixed asymptotes and by-participant asymptotes). In addition, it is not clear the efficiency with which such modified models would be able to capture individual differences in interpretable parameters (such as *time taken to learn*) in a fairly short period of time. Simulations of parameter recoverability further showed that, in our case, reliable estimation of inter-individual differences is actually possible in much less time than in our empirical data.

Third, by treating accuracy as continuously varying over time (i.e., improving with learning), we characterized inter-individual variations from dissociable sources (e.g., rate of learning or asymptotic performance; learning on easier or harder trial types). This contrasts with learning measured at the level of blocks or testing sessions (e.g., Xu et al., 2020). Bayesian model comparison showed that inter-individual variations in performance asymptotes should not be included in our learning models, and similar model comparisons would be straightforward to implement in novel datasets. In this way, we confirmed the primacy of individual differences in learning rate, as opposed to variations in asymptote. By applying the aforementioned methods, we were thus able to demonstrate that between-participant and difficulty-modulated variations in learning were due to differential rates of

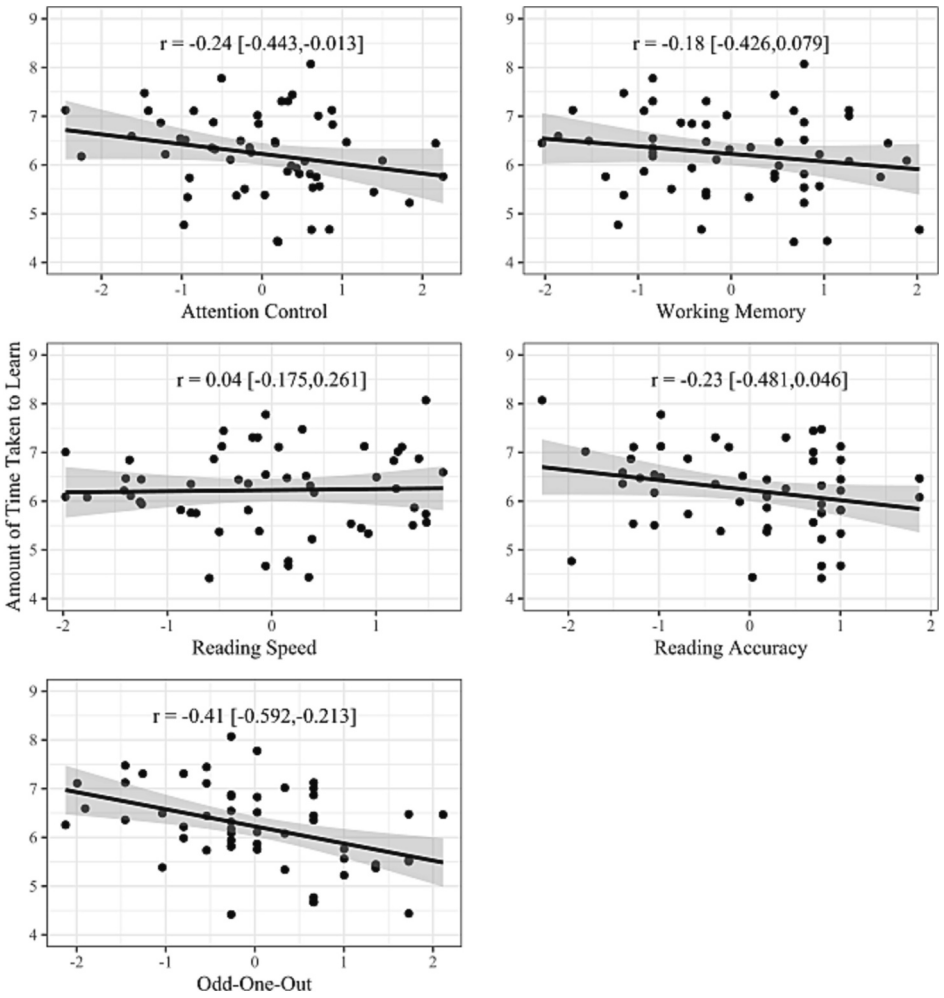

Fig. 5. Cognitive correlates of audio-visual associative learning in the French-speaking sample with the composite scores of attention control and of working memory, as well as with reading speed, reading accuracy and the measure of fluid intelligence.

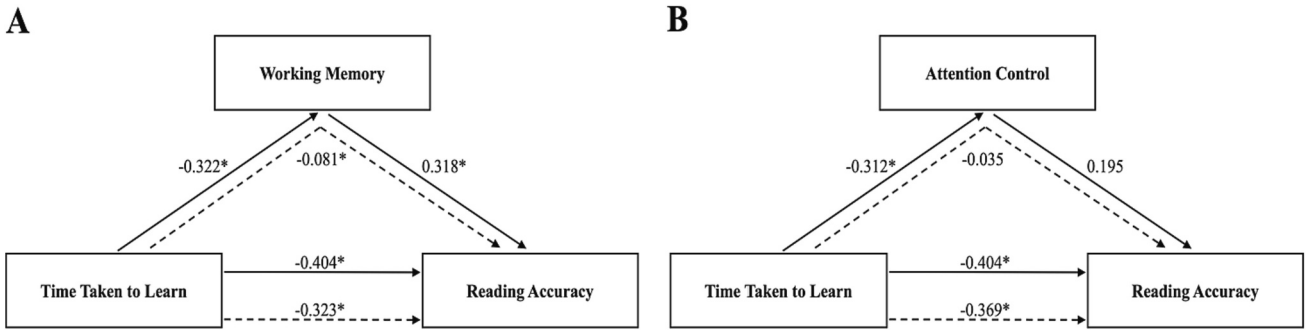

Fig. 6. Mediation model between time-taken-to-learn (i.e., learning rate) and reading accuracy with working memory and with attention control.

learning, as opposed to differences in asymptotic performance. This observation — and especially the absence of differences in asymptotic performance — suggests that the observed disparities between participants are not due to variations in stimulus "learnability" (i.e., because all stimuli had the same fixed asymptote). All models we tested enforced monotonic increases in performance over time (i.e., treating decrements as noise rather than signal), which in longer studies or with richer datasets could be relaxed in order to test models including additional time-sensitive phenomena (e.g., fluctuations in sustained attention, fatigue). While our comparison to the more-flexible descriptive model did not support the need for such flexibility in our data, increased flexibility may be particularly important in less high-functioning populations (e.g., young children, patients).

Fourth, we identified the functional form of change, which conformed much more closely to an exponential function than to a power function. Functional forms of change provide indications about the underlying processes of change: exponential functions tend to imply a simpler learning mechanism than power functions (Dosher & Lu, 2007; Newell & Rosenbloom, 1981). This also supports empirical inferences, for instance that in similar tasks, future work may apply models from the exponential family of functions rather than power functions (Cochrane & Green, 2021b).

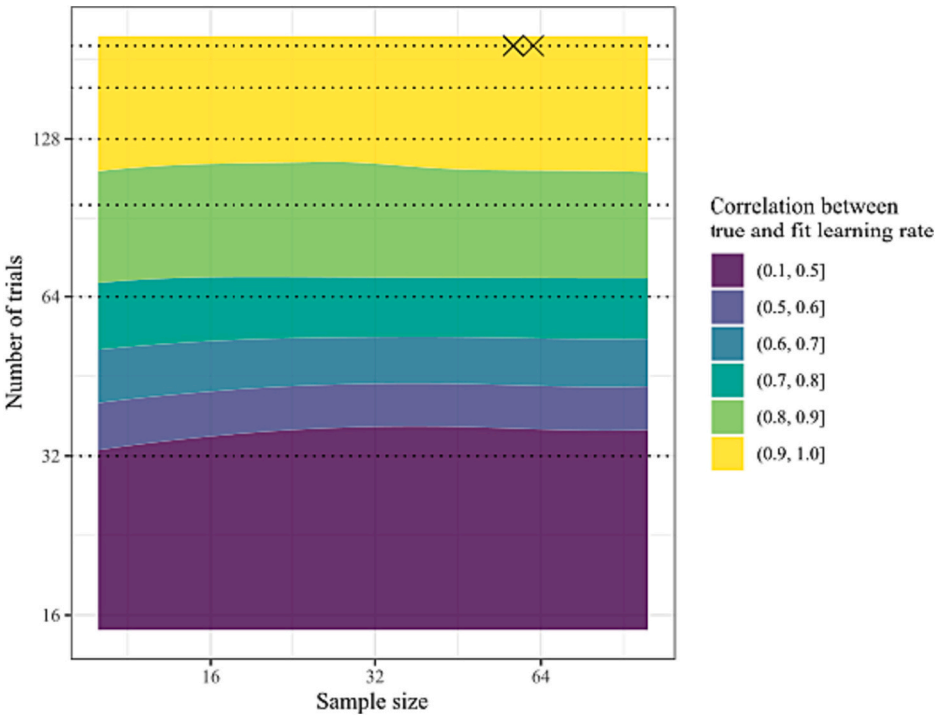

Fig. 7. Recovery of simulated inter-individual differences in the time taken to learn, given the behavioral and analytical methods used here. Spearman ρ was calculated between participant-level generative and estimated parameters for each simulation, then LOESS regression was used to fit and interpolate the surface of ρ by participant number and trial number. Recovery correlation of $\rho > 0.80$ would likely generally be acceptable, while $\rho > 0.90$ would be very good. Two ` × ` marks near the top of the plot show the trial numbers and participant numbers corresponding to the Italian-speaking and French-speaking samples, respectively. Dotted horizontal lines correspond to each block of trials. The axes are log-log scaled to emphasize any differences at relatively small trial numbers or participant numbers.

Fifth, model comparisons further showed that the progression of participants' learning demonstrated an accelerating or decelerating hazard function (i.e., the Weibull shape parameter). This indicated several things, most notably (1) that the majority of participants had accelerating learning trajectories, and more broadly, (2) that the possibility of sigmoid learning may need to be tested in a broader array of learning contexts. Each of these points reinforces previous arguments that learning may often be delayed or even step-function-like in some cases (Gallistel et al., 2004), while averaging across individuals' learning curves may provide spurious indications about the underlying function (Brown & Heathcote, 2003; Dosher & Lu, 2007). Thus, because individual-level trajectories were modeled, the comparisons in both points five and six were additionally useful because they provided an empirical foundation for our eventual individual-differences inferences. The process of model comparison allowed those inferences to be theoretically and empirically stronger.

Sixth, it should be noted that very few studies have assessed, as we did, the learning of audio-visual mappings in the context of non-linguistic stimuli. To our knowledge, only a few studies have explored the cross-modal learning of non-linguistic stimuli within the domain of statistical learning (e.g., Ball, Michels, Thiele, & Noesselt, 2018; Piazza et al., 2018; Shams, Seitz, & van Wassenhove, 2006). Shams et al., 2006, for instance, were among the first to probe statistical learning of arbitrary audio-visual non-linguistic pairings. Subsequent work demonstrated how statistical learning of such pairings comes to influence later visual perception, contributing to resolving sensory ambiguity (Piazza et al., 2018) and shaping temporal expectations (Ball et al., 2018). These works, however, have typically not characterized the dynamics of learning nor tried to link it to other cognitive or everyday skills like reading.

Seventh, by conjointly assessing both audio-visual associative learning and cognitive correlates of learning we were in a position to explore the relationship between non-linguistic audio-visual associative learning and cognitive performance in various tasks. The few studies that have done so have used linguistic audio-visual learning. For example, Xu et al. (2020) found that the only factor related to learning performance was rapid automatized naming, a language-related skill. Other studies, mostly on children, confirmed a correlation between audio-visual learning and phonological awareness abilities (de Jong, Seveke, & van Veen, 2000; Ehm et al., 2019; Karipidis et al., 2017; Lervåg, Bråten, & Hulme, 2009) and rapid automatized naming (Georgiou, Liu, & Xu, 2017; Lervåg et al., 2009). A link between linguistic audio-visual learning and verbal working memory has also been documented in a few behavioral studies in children (Ehm et al., 2019; Lervåg et al., 2009) as well as in neuroimaging studies in adults (Tanabe, Honda, & Sadato, 2005), suggesting the involvement of working-memory throughout audio-visual associative learning tasks, at least when the stimuli are linguistic in nature. The present results extend these findings by highlighting a positive relation between verbal working memory capacity and speed of non-linguistic audio-visual associative learning. Of note, this finding was observed in both our samples, despite these having different linguistic backgrounds (Italian $r = -0.33$; French $r = -0.18$). The current study thus extends previous findings regarding the involvement of domain-general (e.g., executive functions) factors on the acquisition of speech sounds-to-symbols correspondences to non-linguistic mappings. Future behavioral and neuroimaging studies will be needed to clarify the precise working memory sub-skills related to this form of audio-visual associative learning.

Finally, our study highlighted a link between learning rate in our task and reading accuracy. This finding extends an association repeatedly reported in the literature, which emphasizes the connection between reading and the acquisition of audio-visual pairings when the task comprises linguistic stimuli (Ehm et al., 2019; Lervåg et al., 2009). For example, several significant contributions within the broader field of statistical learning and reading also point to links between audiovisual learning with linguistic materials and visual word identification, a key

reading function (for recent reviews, see Frost, Armstrong, & Christiansen, 2019; Schmalz, Moll, Mulatti, & Schulte-Körne, 2019). Of note, by not involving linguistic stimuli but still showing a link to reading and doing so in populations with two different language backgrounds, the observed association between reading accuracy, working memory and mastery of audio-visual associations highlight a separate contribution of audio-visual association processes, independent of language capacities. In future investigations, it will be of interest to assess how these relationships vary across development (see for instance, Altarelli et al., 2019 for data in kindergartners) and whether they hold in populations with clinical populations such as children and adults with developmental dyslexia (Swanson, Zheng, & Jerman, 2009, and Reis, Araújo, Morais, et al., 2020; respectively), who often display reduced working memory skills.

Audio-visual associative learning, and its cognitive correlates, have been a useful test case applied in various populations of different linguistic backgrounds. The data presented extends this work by providing a novel task and method for measuring audio-visual learning using non-linguistic stimuli, thus allowing the deployment of the same task across populations with different language backgrounds. Our study suggests that there may be underlying learning mechanisms that transcend linguistic domains and contribute to both reading ability and the acquisition of audio-visual associations. In addition, many other areas of everyday functioning rely on similar learning processes. Social knowledge, certain occupations' expertise, reading abilities, and even music skills are likely to be supported in part by associating auditory and visual stimuli. Although our findings are limited to the domains of reading and cognition, other disciplines may benefit from using methods similar to those implemented here.

CRediT authorship contribution statement

Angela Pasqualotto: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Aaron Cochrane:** Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Daphne Bavelier:** Conceptualization, Methodology, Funding acquisition, Project administration, Supervision, Visualization, Writing – review & editing. **Irene Altarelli:** Conceptualization, Methodology, Funding acquisition, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Data availability

The data is available on OSF: https://osf.io/qknfy/

Acknowledgements

This work has been funded by the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant no. 661667, LearningDeterminants, as well as a Junior grant from Institut Universitaire de France to I.A. and the NCCR Evolving Language SNF 51NF40_180888, SNF 100014_178814, as well as the Office of Naval Research N00014-20-1-2074 to D.B.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105658.

References

Altarelli, I., Dehaene-Lambertz, G., & Bavelier, D. (2019). Individual differences in the acquisition of non-linguistic audio-visual associations in 5 year olds. *Developmental Science, 23*(4), 1–13. https://doi.org/10.1111/desc.12913

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21*(8), 627–635.

Ball, F., Michels, L. E., Thiele, C., & Noesselt, T. (2018). The role of multisensory interplay in enabling temporal expectations. *Cognition, 170*, 130–146. https://doi.org/10.1016/j.cognition.2017.09.015

Barutchu, A., Fifer, J. M., Shivdasani, M. N., Crewther, S. G., & Paolini, A. G. (2020). The interplay between multisensory associative learning and IQ in children. *Child Development, 91*(2), 620–637. https://doi.org/10.1111/cdev.13210

Bavelier, D., & Green, C. S. (2019). Enhancing attentional control: Lessons from action video games. *Neuron, 104*(1), 147–163. https://doi.org/10.1016/j.neuron.2019.09.031

Bendixen, A., & Andersen, S. K. (2013). Measuring target detection performance in paradigms with high event rates. *Clinical Neurophysiology, 124*(5), 928–940. https://doi.org/10.1016/j.clinph.2012.11.012

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brown, S., & Heathcote, A. (2003). Averaging learning curves across and within participants. *Behavior Research Methods, Instruments, & Computers, 35*(1), 11–21.

Bürkner, P. C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1). https://doi.org/10.18637/jss.v080.i01

Cochrane, A. (2020). TEfits: Nonlinear regression for time-evolving indices. *Journal of Open Source Software, 5*(52), 2535.

Cochrane, A., & Green, C. S. (2021a). Trajectories of performance change indicate multiple dissociable links between working memory and fluid intelligence. *Npj Science of Learning, 6*(1), 33. https://doi.org/10.1038/s41539-021-00111-w

Cochrane, A., & Green, C. S. (2021b). Assessing the functions underlying learning using by-trial and by-participant models: Evidence from two visual perceptual learning paradigms. *Journal of Vision, 21*(13), 5. https://doi.org/10.1167/jov.21.13.5

Cochrane, A., & Green, C. S. (2023). Working memory is supported by learning to represent items as actions. *Attention, Perception, & Psychophysics, 1-12*. https://doi.org/10.3758/s13414-023-02654-z

Crossman, E. R. (1959). A theory of the acquisition of speed-skill*. *Ergonomics, 2*(2), 153–166.

Dale, G., Cochrane, A., & Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-021-02268-3

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biology, 16*(3), Article e2004103.

Dosher, B. A., & Lu, Z.-L. (2007). The functional form of performance improvements in perceptual learning: Learning rates and transfer. *Psychological Science, 18*(6), 531–539. https://doi.org/10.1111/j.1467-9280.2007.01934.x

Ehm, J.-H., Lonnemann, J., Brandenburg, J., Huschka, S. S., Hasselhorn, M., & Lervåg, A. (2019). Exploring factors underlying children's acquisition and retrieval of sound-symbol association skills. *Journal of Experimental Child Psychology, 177*, 86–99. https://doi.org/10.1016/j.jecp.2018.07.006

FeldmanHall, O., & Dunsmoor, J. E. (2019). Viewing adaptive social choice through the Lens of associative learning. *Perspectives on Psychological Science, 14*(2), 175–196. https://doi.org/10.1177/1745691618792261

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The sleeping infant brain anticipates development. *Current Biology, 27*(15), 2374–2380.e3. https://doi.org/10.1016/j.cub.2017.06.070

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America, 101*(36), 13124. https://doi.org/10.1073/pnas.0404965101

Georgiou, G., Liu, C., & Xu, S. (2017). Examining the direct and indirect effects of visual–verbal paired associate learning on Chinese word reading. *Journal of Experimental Child Psychology, 160*, 81–91. https://doi.org/10.1016/j.jecp.2017.03.011

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS Computational Biology, 11*(11), Article e1004567.

Gonzalo, D., Shallice, T., & Dolan, R. (2000). Time-dependent changes in learning audiovisual associations: A single-trial fMRI study. *NeuroImage, 11*(3), 243–255. https://doi.org/10.1006/nimg.2000.0540

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). Bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software, 1*(10), 2020. https://www.jstatsoft.org/v092/i10.

Hämäläinen, J. A., Parviainen, T., Hsu, Y.-F., & Salmelin, R. (2019). Dynamics of brain activation during learning of syllable-symbol paired associations. *Neuropsychologia, 129*, 93–103. https://doi.org/10.1016/j.neuropsychologia.2019.03.016

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron, 76*(6), 1225–1237. https://doi.org/10.1016/j.neuron.2012.06.022

Hansen, J. C., & Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49* (3–4), 277–290. https://doi.org/10.1016/0013-4694(80)90222-9
Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56*(1).
Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185–207. https://doi.org/10.3758/BF03212979
de Jong, P. F., Seveke, M.-J., & van Veen, M. (2000). Phonological sensitivity and the Acquisition of new Words in children. *Journal of Experimental Child Psychology, 76* (4), 275–301. https://doi.org/10.1006/jecp.1999.2549
Kafaligonul, H., & Oluk, C. (2015). Audiovisual associations alter the perception of low-level visual motion. *Frontiers in Integrative Neuroscience, 9*, 26. https://doi.org/10.3389/fnint.2015.00026
Karipidis, I., Pleisch, G., Röthlisberger, M., Hofstetter, C., Dornbierer, D., Stämpfli, P., & Brem, S. (2017). Neural initialization of audiovisual integration in prereaders at varying risk for developmental dyslexia. *Human Brain Mapping, 38*(2), 1038–1055. https://doi.org/10.1002/hbm.23437
Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017). Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Current Biology, 27*(6), 840–846.
Kattner, F., Cochrane, A., & Green, C. S. (2017). Trial-dependent psychometric functions accounting for perceptual learning in 2-AFC discrimination tasks. *Journal of Vision, 17*(11). https://doi.org/10.1167/17.11.3
Kersey, A. J., & Emberson, L. L. (2017). Tracing trajectories of audio-visual learning in the infant brain. *Developmental Science, 20*(6). https://doi.org/10.1111/desc.12480
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.
Leibowitz, N., Baum, B., Enden, G., & Karniel, A. (2010). The exponential learning equation as a function of successful trials results in sigmoid performance. *Journal of Mathematical Psychology, 54*(3), 338–340. https://doi.org/10.1016/j.jmp.2010.01.006
Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45*(3), 764–781. https://doi.org/10.1037/a0014132
Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology, 74*(September 2017), 150–156. https://doi.org/10.1016/j.jesp.2017.09.011
Li, Y., Wang, F., Huang, B., Yang, W., Yu, T., & Talsma, D. (2016). The modulatory effect of semantic familiarity on the audiovisual integration of face-name pairs. *Human Brain Mapping, 37*(12), 4333–4348. https://doi.org/10.1002/hbm.23312
Madec, S., Le Goff, K., Anton, J. L., Longcamp, M., Velay, J. L., Nazarian, B., … Rey, A. (2016). Brain correlates of phonological recoding of visual symbols. *NeuroImage, 132*, 359–372. https://doi.org/10.1016/j.neuroimage.2016.02.010
Mersad, K., Kabdebon, C., & Dehaene-Lambertz, G. (2021). Explicit access to phonetic representations in 3-month-old infants. *Cognition, 213*, 104613.
Miller, A. L., & Unsworth, N. (2020). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(12), 2277.
Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition, 1*.
Newell, K. M., Liu, Y.-T., & Mayer-Kress, G. (2001). Time scales in motor learning and development. *Psychological Review, 108*(1), 57.
Pech-Georgel, C., & George, F. (2011). *EVALAD: Evaluation du langage écrit et des compétences transversales – Adolescents de 1ère et de terminale ou adultes*. De Boeck Supérieur.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897X00366
Piazza, E. A., Denison, R. N., & Silver, M. A. (2018). Recent cross-modal statistical learning influences visual perceptual selection. *Journal of Vision, 18*(3), 1. https://doi.org/10.1167/18.3.1
Pylyshyn, Z. W., & Storm, R. W. (1998). Tracking multiple independent targets: Evidence for a parallel tracking mechanism*. *Spatial Vision, 3*(3), 179–197. https://doi.org/10.1163/156856888X00122
Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic reinforcement learning: The role of structure and attention. *Trends in Cognitive Sciences, 23*(4), 278–292.
Reis, A., Araújo, S., Morais, I. S., et al. (2020). Reading and reading-related skills in adults with dyslexia from different orthographic systems: A review and meta-analysis. *Annals of Dyslexia, 70*, 339–368. https://doi.org/10.1007/s11881-020-00205-x
Schmack, K., Weilnhammer, V., Heinzle, J., Stephan, K. E., & Sterzer, P. (2016). Learning what to see in a changing world. *Frontiers in Human Neuroscience, 10*, 263. https://doi.org/10.3389/fnhum.2016.00263
Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading, 23*(1), 64–76.
Schmalz, X., Schulte-Körne, G., De Simone, E., & Moll, K. (2021). What do artificial orthography learning tasks actually measure? Correlations within and across tasks. *Journal of Cognition, 4*(1), 7. https://doi.org/10.5334/joc.144
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–218.
Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception, 36* (10), 1445–1453. https://doi.org/10.1068/p5843
Shams, L., Seitz, A., & van Wassenhove, V. (2006). Audio-visual statistical learning. *Journal of Vision, 6*(6), 152. https://doi.org/10.1167/6.6.152
Shelton, B., & Scarrow, I. (1984). Two-alternative versus three-alternative procedures for threshold estimation. *Perception & Psychophysics, 35*(4), 385–392.
Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass, 14*(3), Article e12365.
Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 372*(1711), 20160059. https://doi.org/10.1098/rstb.2016.0059
Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research, 49*, 418–432. https://doi.org/10.3758/s13428-016-0719-z
Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2014). Absolute performance of reinforcement-learning models for the Iowa gambling task. *Decision, 1*(3), 161.
Swanson, H. L., Zheng, X., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities, 42*(3), 260–287.
Tanabe, H. C., Honda, M., & Sadato, N. (2005). Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *Journal of Neuroscience, 25*(27), 6409–6418. https://doi.org/10.1523/JNEUROSCI.0636-05.2005
Thorndike, E. L. (1908). The effect of practice in the case of a purely intellectual function. *The American Journal of Psychology, 19*(3), 374–384. https://doi.org/10.2307/1413197
Vancleef, K., Read, J. C., Herbert, W., Goodship, N., Woodhouse, M., & Serrano-Pedraza, I. (2018). Two choices good, four choices better: For measuring stereoacuity in children, a four-alternative forced-choice paradigm is more efficient than two. *PLoS One, 13*(7), Article e0201366.
Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). TX: Pearson.
Wechsler, D. (2011). *WAIS-IV- Wechsler adult intelligence scale- fourth edition. Manuel d'administration*. Paris: ECPA.
Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293–1313.
Xu, W., Kolozsvari, O. B., Oostenveld, R., & Hämäläinen, J. A. (2020). Rapid changes in brain activity during learning of grapheme-phoneme associations in adults. *NeuroImage, 220*(March). https://doi.org/10.1016/j.neuroimage.2020.117058
Younger, J. W., & Booth, J. R. (2018). Parietotemporal stimulation affects Acquisition of Novel Grapheme-Phoneme Mappings in adult readers. *Frontiers in Human Neuroscience, 12*. https://doi.org/10.3389/fnhum.2018.00109
Zhang, P., Zhao, Y., Dosher, B. A., & Lu, Z.-L. (2019). Assessing the detailed time course of perceptual sensitivity change in perceptual learning. *Journal of Vision, 19*(5), 9. https://doi.org/10.1167/19.5.9
Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuehl, 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