Title
A novel non-linguistic audio-visual learning paradigm to test the cognitive correlates of learning rate

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Pasqualotto, Angela
Cochrane, Aaron
Bavelier, Daphne
et al.

Publication Date
2021

Peer reviewed

A novel non-linguistic audio-visual learning paradigm to test the cognitive correlates of learning rate

Angela Pasqualotto[1*] (angela.pasqualotto@unige.ch)

Aaron Cochrane[1*] (aaron.cochrane@unige.ch)

Daphne Bavelier[1] (daphne.bavelier@unige.ch)

Irene Altarelli[2] (irene.altarelli@unige.ch)

[1] Faculty of Psychology and Educational Sciences, University of Geneva, 1205 Geneva, Switzerland;

[2] Université de Paris, LaPsyDÉ, CNRS, 75005 Paris, France;

[*] Contributed equally

Abstract

Audio-visual (AV) associative learning is central to many aspects of cognitive development and is key in reading acquisition. Most studies thus far have examined AV associative learning involving linguistic stimuli. Yet it is of importance to examine cross-modal learning free of familiarity confounds. We therefore designed an AV learning paradigm relying on novel, non-linguistic auditory and visual stimuli, which were both unfamiliar to participants. On top of AV learning, we collected performance in reading-related abilities, as well as in more domain-general skills, in a population of healthy Italian-speaking adults (N=57). By fitting trial-by-trial performance in our novel learning task, we demonstrate the expected variability in speed of learning (learning rate) across participants. We then show that speed of learning in our novel learning task is positively associated with working memory and replicate this result in a set of French-speaking participants (N=32), showing that it holds in another language.

Keywords: audio-visual learning; associative learning; learning rate; cognitive correlates; working memory; individual differences

Introduction

Audio-visual associative learning is the process by which auditory and visual stimuli are linked together through exposure. This process is central to many aspects of cognitive development and is particularly crucial for the very initial steps of reading acquisition. Indeed, pairings between linguistic sounds (phonemes) and letters or groups of letters (graphemes) is essential for the successful decoding of print. Audio-visual (AV) learning abilities are present very early in the course of development, including in pre-verbal infants (Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Kersey & Emberson, 2017; Mersad, Kabdebon, & Dehaene-Lambertz, 2018). The process by which audio-visual associations are acquired has been studied extensively in children, especially in the linguistic domain. Infants' and children' AV learning has been documented in the context of both simultaneously heard and seen speech (Soto-Faraco, Calabresi, Navarra, Werker, & Lewkowicz, 2012) and arbitrary audio-visual pairings (Friedrich et al., 2017).

Studies of the cognitive correlates of AV learning in children indicate that learning linguistic audio-visual mappings is positively related to reading-related skills such as phonological awareness (de Jong, Seveke, & van Veen, 2000; Ehm et al., 2019; Georgiou, Liu, & Xu, 2017; Karipidis et al., 2017; Lervåg, Bråten, & Hulme, 2009; Litt, de Jong, van Bergen, & Nation, 2013; Windfuhr & Snowling, 2001) and rapid automatized naming (Georgiou et al., 2017; Lervåg et al., 2009). Very few studies in children have looked at the learning of non-linguistic pairings, such as environmental sounds and unknown symbols. Among those, Altarelli et al.'s study (2019) revealed that, in 5-6 years old, phonological awareness abilities again positively predicted AV learning, even though the to-be-learnt pairings did not entail any linguistic information.

As regards adult populations, while examining audio-visual associative learning in the non-linguistic domain is just as crucial for getting at the core of cross-modal learning free of familiarity confounds, very few studies have explored the time course and specificities of learning arbitrary AV pairings among unfamiliar, nonverbal auditory and visual stimuli. Barutchu et al. (2020) investigated both linguistic and non-linguistic audio-visual mapping in adults, allowing for relatively limited practice (128 trials per condition), highlighting learning in both cases, yet faster in the linguistic condition.

Moreover, and to the best of our knowledge, little is known as regards the cognitive correlates of audio-visual associative learning in adults. Schmalz and colleagues (2021) examined the relation between the acquisition of linguistic AV pairings and reading abilities in German-speaking adults, failing to uncover one. Xu et al. (2020) studied the link between linguistic AV associations learning and cognitive performance in a number of tasks, in Finnish-speaking participants, finding a correlation only between learning speed at the scale of training blocks and rapid automatized naming.

To best address the question of the cognitive abilities that may be related to AV association learning in adulthood, here we developed an audio-visual learning paradigm relying on novel, non-linguistic auditory and visual stimuli, which were both unfamiliar to participants.

Unlike Schmalz et al. (2021) and Xu et al. (2020), we used environmental sounds, avoiding any possible interference from phonological representations of the auditory stimuli as well as from the possibly unbalanced familiarity of the auditory versus visual stimuli. Progress in our AV learning task was analysed on a trial-by-trial basis, allowing us to fit a fully continuous-time model to the data, thus obtaining a fine-grained measure of learning rate for each participant.

Finally, on top of AV learning, we collected performance in a number of cognitive tasks tapping reading-related, domain-specific abilities as well as more domain-general skills. Our analyses are based on two samples: 57 Italian-speaking and 32 French-speaking healthy adults.

As regards our main hypotheses, we predict that the link between phonological awareness skills and non-linguistic AV learning, which was evidenced in young children (Altarelli et al., 2019), will not be found in adults – in line with the recent findings of Xu et al. (2020) in linguistic AV learning. In contrast, we expect greater reliance on domain-general executive functions such as working memory and cognitive flexibility, as the learning required in our task remains cognitively taxing.

Methods

Participants

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 the norms); (iii) intelligence within the normal range (cut-off score ≥ 7) as measured with the WAIS-IV subtest of Matrix Reasoning (Weschler, 2008), (iv) reported normal or corrected-to-normal vision and hearing. In total, data from seventy-two participants were included in the analysis (40 females, mean age 25 years old).

French-speaking sample. A total of forty-one French-speaking adult participants participated in the study (23 females, mean age 23 years old). Criteria for inclusion were the following: (i) no diagnosis of psychological/neurological disorders; (ii) reported normal or corrected-to-normal vision and hearing.

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.

Design and materials

Italian-speaking participants were tested individually in one, 1.5 hour-long session. In addition to the novel AV associative learning task, our battery of tasks comprised assessments of phonological awareness, rapid automatized naming, auditory and visual attention, short-term and working-memory skills. All tasks are presented below.

French-speaking participants were also tested individually. The battery of tasks was different compared to the one administered to the Italian-speaking participants, in that it was restricted to domain-general tasks (see below).

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 auditory stimuli (environmental sounds, 8 in total) and novel visual symbols (6 in total), as depicted in Figure 1. The task administered here consisted of three main parts. In the first part (duration: 3 minutes), participants familiarized themselves with the auditory stimuli, by passively listening to each of them, presented one at a time. All sounds were environmental sounds that are unfamiliar to the participants yet easily discriminable, as demonstrated in previous studies using similar stimuli (Seitz, Kim, van Wassenhove, & Shams, 2007). For the purpose of the audio-visual associative learning task, sounds to be associated with symbols were paired (total duration: 3850ms) and always presented as a sequence of two sounds, as illustrated in Fig. 1A.

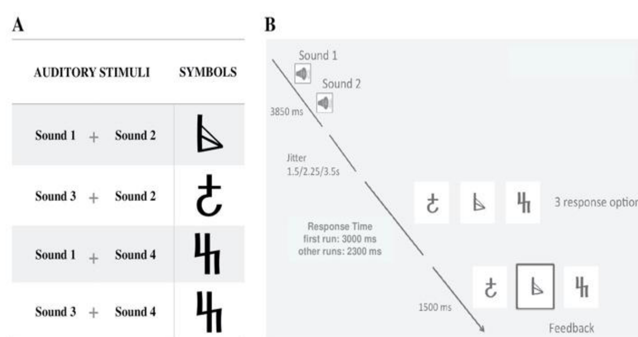

Figure 1: A. Audio-visual associative learning task: examples of four of the to-be-learnt audio-visual pairs. B. Each trial began with the presentation of a pair of auditory stimuli (3850

ms overall) followed by a variable ISI and then three response options. After the participant's response, a visual feedback was presented.

The second part of the task (duration: 5 minutes) consisted of a symbol familiarization phase in which the 6 symbols, adapted from the Bamum alphabet and unknown to participants, were presented in the context of a 1-back task. The choice of a 1-back task was made in order to ensure that participants would pay attention to the stimuli 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 second with a 1 second inter-stimulus interval. The task comprised 80 trials, with 10 1-back repetitions (2 of each symbol). The sequence of symbols was the same for all participants.

Finally, participants underwent the actual audio-visual association task for six blocks of five minutes each (see Fig. 1B). Participants were asked to learn the associations between the pairs of sounds and the symbols presented before, in a 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 is crucial (easy trials). Each block included hard and easy trials in a pseudo-randomized order. Following participants' response on each trial, feedback was provided, indicating the correct symbol.

Italian-speaking sample. *Reading-related measures.* Phonological awareness (PA) skills were evaluated through a phoneme deletion task and a phoneme blending task. Lexical access was tested through a rapid picture naming task (RAN of objects). In all of these tests, derived from the VALS – the Italian adaptation of the EVALAD battery (Pech-Georgel & George, 2011) – two scores are calculated: the accuracy and the speed with which the subject performs the test.
Domain-general skills. Fluid Intelligence was assessed by means of the Matrix Reasoning subtest of the WAIS-IV scale (Weschler, 2008). From the EVALAD battery (Pech-Georgel & George, 2011), we administered the forward and backward Digit span tasks, in order to assess participant's verbal short term and working memory skills, and a Listening span task to assess both working memory and resistance to interference.
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 target objects – among distractors – as they moved around the screen. After an interval of object motion, one of the objects was probed, and participants were required to decide whether this specific object was a target or a distractor at the beginning of the trial. Auditory selective attention was assessed by means of a novel task, based on the paradigm proposed by Hansen & Hillyard (1980). Two series of sounds were presented simultaneously to participants through headphones and differed 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.

French-speaking sample. *Domain-general skills.* Computerized versions of the Odd-one-out task (Hampshire, Highfield, Parkin, & Owen, 2012) and of the backward Digit span (Hampshire et al., 2012) were administered to evaluate fluid intelligence and working memory, respectively. Unlike the Italian sample, a forward Digit span task was not included here. The same MOT and auditory attention tasks were otherwise used as in the Italian sample.

Data analysis

The AV associative learning task was first screened for participants who did not demonstrate learning. 10 out of 72 Italian-speaking participants and 6 out of 41 French-speaking participants were excluded for not learning, as indicated by accuracy on adjacent tertiles of trials not increasing over time or not being above chance (33%) on the last tertile of the task.

The data was then fit with a mixed-effects Bayesian nonlinear regression using Stan via the **brms** package in R (Bürkner, 2017; for similar modeling approaches see Dale et al., 2021). In short, this nonlinear learning model approach estimated the trial-by-trial improvement in percent correct starting at 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 an exponential function of trial number (Dosher & Lu, 2007; Heathcote, Brown, & Mewhort, 2000; see Equation 1).

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

In this equation there are three free parameters, however, we fixed *start* to ⅓ 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* and *rate* [a time constant associated with half of change from start to asymptote] were simultaneously estimated within generalized linear mixed-effects models (see Equations 2 & 3, in **brms**/**lme4** "Wilkinson" model notation).

2. $$logit(asymptote) \sim trialDifficulty + (trialDifficulty \mid participant)$$
3. $$(rate) \sim trialDifficulty + (trialDifficulty \mid participant)$$

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 [*asymptote*s] bounded at [0,1] and time constants [*rate*s] bounded to positive reals. All priors were default where possible, with the priors for the logit-*asymptote* log-*rate* parameters having wide priors of *normal(0,3)* and *normal(4,3)* respectively. A Bernoulli response distribution was used due to the binary accuracy being modeled, and all models were run for 15,000 iterations, discarding the first 9,000 iterations as 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 .001 lapse rate). In this alternative model the only learning parameter was therefore the time constant of change, or *rate* (parameterized as described above). We compared the estimated-asymptote models to the fixed-asymptote models using bridge sampling estimates of Bayes Factors, and we found overwhelming support for models with fixed asymptotes. In the Italian-speaking sample, across 5 bridge sampling estimates, the smallest Bayes Factor was 172 in favor of the fixed-asymptote model. Using the same procedure, in the French-speaking sample the smallest Bayes Factor was 203 in favor of the fixed-asymptote model. As such, we used the fixed-asymptote model in tests of individual differences in learning (i.e., only using between-participant variation in learning rate).

Estimated model fixed effects regarding the difficulty manipulation were first assessed. Then, by-participant (i.e., random-effects) parameters were extracted from the model, and these parameters were subsequently used in tests of relations to other measures (e.g., attention or memory).

The Italian-speaking sample's 2 phonological awareness measures were independently converted into Inverse Efficiency Scores (i.e., RT/accuracy). Within each sample, all cognitive variables were then screened for multivariate outliers using the robust Mahalanobis distance method. Robust covariance estimation utilized a minimum of 90% of the sample, with outliers being identified using a chi-square cutoff with α=.05. This led to 5 Italian-speaking participants and 3 French-speaking participants being excluded due to being multivariate outliers (final sample size 57 Italian-speaking, 32 French-speaking). Univariate Yeo-Johnson power transformations were then applied, with Yeo-Johnson λ optimized to minimize the univariate skew of each variable. Variables were next z-scored. Last, a composite measure was calculated for phonological awareness; this composite measure used the dominant component from a PCA of the phonological awareness Inverse Efficiency Scores (i.e., the composite was the underlying dimension explaining the most variance in both measures; hereafter referred to as PA IES).

Our initial tests of individual differences in audio-visual associative learning used bivariate product-moment correlations. We next used bootstrapped (2000 resamples with replacement) robust linear models to test the reliability of regression estimates predicting rate of learning, while controlling for other effects. We additionally computed the out-of-sample proportion of variance explained by fitting 2000 robust linear models to random 80% subsamples of the data and assessing the reduction in OLS error to the corresponding held-out 20%. We report the median out-of-sample delta R-squared.

Results

Mixed-effects nonlinear model results

The estimated time to half of learning was 74 trials in the Italian-speaking sample and 59 trials in the French-speaking sample (see Table 1 and Table 2). Both samples showed reliable modulations of learning rate in response to difficulty manipulations.

Table 1. Model of Italian-speaking sample's learning: Fixed effects

	Estimate	Lower 95% CI	Upper 95% CI
Rate Intercept	6.22	5.97	6.46
Difficulty coefficient	-0.49	-0.80	-0.18

Table 2. Model of French-speaking sample's learning: Fixed effects

	Estimate	Lower 95% CI	Upper 95% CI
Rate Intercept	5.88	5.58	6.19
Difficulty coefficient	-0.49	-0.79	-0.21

Italian-speaking sample: Correlates of learning

Of the eight possible correlates of learning rate, only forward Digit span showed a reliable association ($r(55) = -0.39$, $CI_{95} = [-0.59,-0.15]$; see Fig. 2). Nonetheless, due to the relatively strong correlations with auditory attention ($r(55) = -0.23$, $CI_{95} = [-0.46,-0.03]$) and backward Digit span ($r(55) = -0.23$, $CI_{95} = [-0.46,-0.03]$), we selected all three measures for further analysis with stepwise robust regressions in the next section.

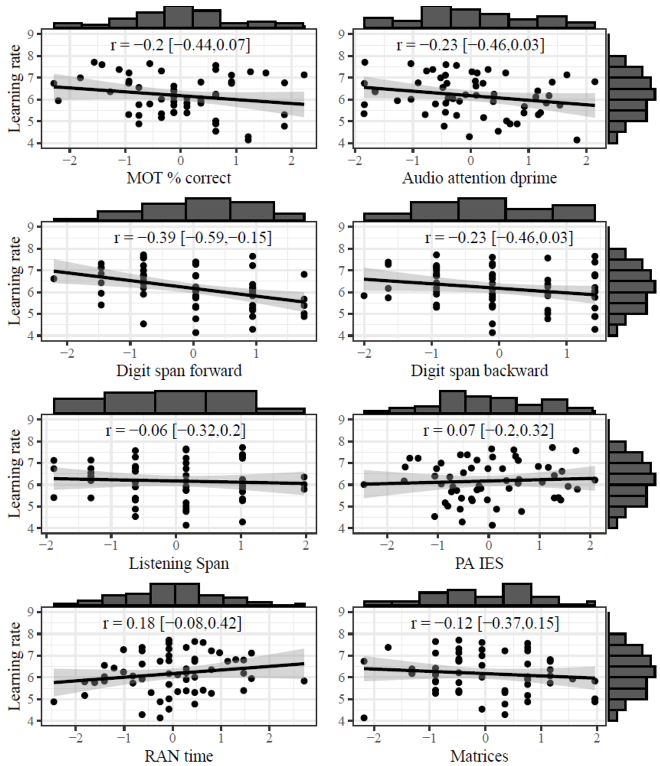

Figure 2: Correlates of AV learning in the Italian-speaking sample.

Italian-speaking sample: Regressions predicting learning

In step 1 of our stepwise regressions, we predicted learning rate using the covariate demographic variables of sex (b = 0.31, CI = [0.0051,0.56], ΔR^2_{oos} = 0.083) and age (b = 0.12, CI = [-0.17,0.35], ΔR^2_{oos} = -0.001). In step 2 we added PA IES (b = 0.023, CI = [-0.18,0.27], ΔR^2_{oos} = -0.014), with no reliable effect on learning rate. We also tested the addition of RT and accuracy variables to the model separately, with similar null effects. In step 3a we added forward Digit span (b = -0.36, CI = [-0.58,-0.13], ΔR^2_{oos} = 0.132), whereas in step 3b we instead added backward Digit span (b = -0.22, CI = [-0.5,0.063], ΔR^2_{oos} = 0.049) and in step 3c we added auditory attention (b = -0.11, CI = [-0.42,0.2], ΔR^2_{oos} = -0.006). Using the CI criterion, only the model in which forward Digit span was added showed a reliable effect, with participants scoring higher on forward Digit span also learning in less time. Notably, when using the out-of-sample variance explained criterion (i.e., >0), backward Digit span did predict around 5% of variance in the held-out samples.

To further test the robustness of our predictors we utilized penalized regression (LASSO) using the R package **glmnet**. We determined the optimal shrinkage parameter [lambda] through 200 runs of cross-validated fits to multiple lambdas (i.e., **cv.glmnet**; each cross-validation used 5 folds). Using the median lambda to minimize out-of-sample error (lambda = 0.204) we next fit a LASSO regression model using all cognitive variables as predictors, with the result that only the forward Digit span coefficient (b = -0.182) remained (i.e., coefficients from all other cognitive variables were penalized to 0). Clearly this corroborates the bivariate outcomes reported in the correlation matrices, in which higher forward Digit span scores are associated with shorter times taken to learn.

French-speaking sample: Correlations

In an independent sample with a different linguistic background, we replicated the correlation between a verbal working memory span measure and learning rate (see Fig. 3). Higher working memory span scores were associated with learning in fewer trials. However, this correlation involved backward Digit span, which in our Italian-speaking sample had demonstrated weaker associations with learning than forward Digit span.

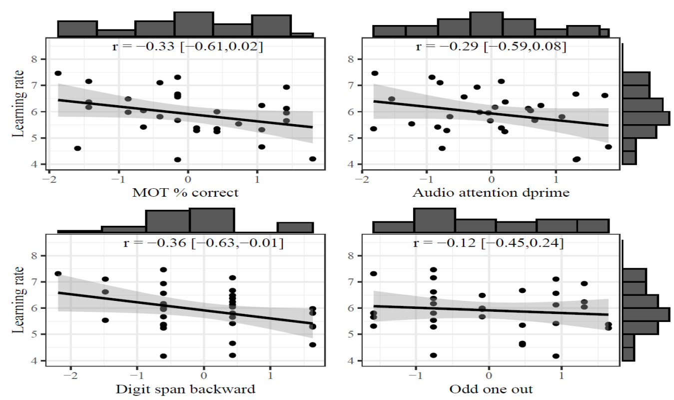

Figure 3: Correlates of AV learning in the French-speaking sample.

Discussion

Whereas AV associative learning is a crucial process in development and across the lifespan, e.g., for the acquisition of reading, little is known about the cognitive abilities it rests upon. The present study aimed to explore healthy adults' ability to learn arbitrary associations between novel visual stimuli and non-linguistic auditory stimuli, as well as the relationship between such learning and performance in reading-related and more domain-general skills. Data were collected in both an Italian- and a French-speaking sample for independent confirmation of results.

By fitting trial-by-trial performance of participants in this novel AV learning task, we demonstrated the expected variability in speed of learning (learning rate) across participants and the sensitivity of our rate measure to learning difficulty, as expected. In a second step we showed that the speed of learning in the AV learning task was positively associated with working memory capacity in both our samples. The weaker working memory effect in the Italian (i.e., backward Digit span) was the one that we could confirm in the French sample, in which forward Digit span data were lacking.

Prior studies – mostly in children – have investigated the cognitive correlates of learning audio-visual mappings, revealing correlations with phonological awareness abilities, rapid automatized naming (Ehm et al., 2019; Georgiou et al., 2017; de Jong et al., 2000; Lervåg et al., 2009; Karipidis et al., 2017) and verbal working memory (Ehm et al., 2019; Lervåg et al., 2009). Crucially, the vast majority of these studies used linguistic stimuli in the auditory domain, like native-language phonemes, to be paired with unfamiliar symbols. 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). In addition, the exact same paradigm and stimuli cannot be applied to participants speaking different languages. The current AV learning experiment introduces non-linguistic, environmental sounds in order to assess the very process of building cross-modal AV associations free of familiarity confounds and in a way that is comparable across participants speaking different languages.

Of importance, while many previous studies examined learning in a test phase that followed passive exposure to the audio-visual pairs (Callan, Callan, & Masaki, 2005; Madec et al., 2016), 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. Nonlinear mixed-effects models provided for the simultaneous estimation of all participants' full trajectories of learning. Beyond the currently reported results this may allow, for example, additional trial-by-trial comparisons of each participant's difference in performance between easier and harder trials (which is beyond our scope here). Such models also allow for fully Bayesian model comparisons. Here we demonstrated that between-participant and difficulty-modulated variations in learning were due to differential rates of learning, as opposed to differences in asymptotic performance or stimulus "learnability".

Very similar non-linguistic audio-visual learning processes were previously investigated in preschoolers (Altarelli et al., 2019), revealing a surprising relation between progress in the task and phonological awareness skills. We hypothesized and confirmed that in adults such a relationship is not present anymore. In young, pre-reading children, it is possible that variation in phonological awareness abilities is tightly connected with their ability to extract and integrate redundant auditory and visual information from talkers surrounding them (Melby-Lervåg, Lyster, & Hulme, 2012), thus linking greater audio-visual association abilities to greater phonological skills. In adults, who are expert readers of their language and for whom phonological awareness skills have been trained by the very process of learning to read, such coupling might be weaker.

In contrast, we found a positive relation between verbal working memory capacity and speed of AV learning, in both our samples, despite these having different linguistic backgrounds (Italian r= -.39; French r= -.36). These results are consistent with recent behavioral studies in children (Ehm et al., 2019; Lervåg et al., 2009) and neuroimaging findings in adults, suggesting the involvement of working memory throughout AV associative learning tasks (Tanabe et al., 2005). Future behavioral and neuroimaging studies will be needed to clarify the precise working memory sub-skills related to this form of AV associative learning.

The current study thus extends previous findings regarding the involvement of domain-general (e.g., executive functions) factors in the acquisition of speech sounds-to-symbols correspondences to non-linguistic mappings. It presents a novel AV learning task that allows modeling of each participant's full trajectory of learning. Finally, it suggests verbal working memory capacity as playing an important role in the acquisition of audio-visual mappings, even when non-verbal materials are involved, at least in adults.

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 to I.A. and the NCCR Evolving Language as well as the Office of Naval Research N00014-20-1-2074 to D.B.

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

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

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software; Vol 1, Issue 1 (2017)*. Retrieved from https://www.jstatsoft.org/v080/i01

Callan, A. M., Callan, D. E., & Masaki, S. (2005). When meaningless symbols become letters: Neural activity change in learning new phonograms. *NeuroImage*, *28*(3), 553–562. https://doi.org/10.1016/j.neuroimage.2005.06.031

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

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/https://doi.org/10.1006/jecp.1999.2549

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

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

Pech-Georgel C., & George F. (2011). EVALAD. Marseille. Soleil.

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

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

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

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/https://doi.org/10.1002/hbm.23437

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

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology*. American Psychological Association. https://doi.org/10.1037/a0014132

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

Litt, R. A., de Jong, P. F., van Bergen, E., & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associate learning (PAL): What drives the PAL–reading relationship? *Journal of Experimental Child Psychology*, *115*(1), 137–149. https://doi.org/https://doi.org/10.1016/j.jecp.2012.11.0

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/https://doi.org/10.1016/j.neuroimage.2016.02.010

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Mersad, K., Kabdebon, C., & Dehaene-Lambertz, G. (2018). A precursor of reading ability? 3-month-old infants easily learn to pair a phoneme with a visual shape. *BioRxiv*, 466318. https://doi.org/10.1101/466318

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/https://doi.org/10.1163/156856888X00122

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

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

Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J. F., & Lewkowicz, D. J. (2012). The development of audiovisual speech perception. In *Multisensory Development*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0009

Tanabe, H. C., Honda, M., & Sadato, N. (2005). Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *25*(27), 6409–6418. https://doi.org/10.1523/JNEUROSCI.0636-05.2005

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale--Fourth Edition (WAIS-IV)* [Database record]. APA PsycTests. https://doi.org/10.1037/t15169-000

Windfuhr, K. L., & Snowling, M. J. (2001). The Relationship between Paired Associate Learning and Phonological Skills in Normally Developing Readers. *Journal of Experimental Child Psychology*, *80*(2), 160–173. https://doi.org/10.1006/jecp.2000.2625

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