Spatial attention underpins social word learning in the right fronto-parietal network

Laura Verga [a,b], Sonja A. Kotz [a,b,*]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, Dept. of Neuropsychology, Leipzig, Germany
[b] Maastricht University, Faculty of Psychology and Neuroscience, Dept. of Neuropsychology and Psychopharmacology, Maastricht, the Netherlands

ARTICLE INFO

ABSTRACT

In a multi- and inter-cultural world, we daily encounter new words. Adult learners often rely on a situational context to learn and understand a new word's meaning. Here, we explored whether interactive learning facilitates word learning by directing the learner's attention to a correct new word referent when a situational context is non-informative. We predicted larger involvement of inferior parietal, frontal, and visual cortices involved in visuo-spatial attention during interactive learning. We scanned participants while they played a visual word learning game with and without a social partner. As hypothesized, interactive learning enhanced activity in the right Supramarginal Gyrus when the situational context provided little information. Activity in the right Inferior Frontal Gyrus during interactive learning correlated with post-scanning behavioral test scores, while these scores correlated with activity in the Fusiform Gyrus in the non-interactive group. These results indicate that attention is involved in interactive learning when the situational context is minimal and suggest that individual learning processes may be largely different from interactive ones. As such, they challenge the ecological validity of what we know about individual learning and advocate the exploration of interactive learning in naturalistic settings.

1. Introduction

Language learning, either in classrooms or by immersion, is typically a social activity and so is language use. Yet, language learning is often studied at the individual level, and only one previous study explored the neural underpinnings of social word learning in adults (Jeong et al., 2010). The authors reported activation of the right Supramarginal Gyrus (rSMG) during the retrieval of words learned while watching video clips of non-interactive social learning situations (Jeong et al., 2010). This finding aligns with results in social neuroscience studies where social stimuli activate areas inside and around the right Temporo-Parietal Junction (TPJ), a brain region at the boundary of the temporo-parietal cortex (Jeong et al., 2010; Carter and Huettel, 2013).

While social neuroscience studies consistently report TPJ activation, this region also engages in visuo-spatial attention (Decety and Lamm, 2007) where the re-directing of attention toward a new target plays a central role (Corbetta et al., 2000). This, in turn, facilitates the activation of task-relevant visual cortices via top-down connections from the inferior parietal cortex (Supramarginal and Angular Gyrus) and superior visual cortices (including the Lingual Gyrus) to the striate cortex

(Horwitz et al., 1998; Chambers et al., 2004; Verghese et al., 2014) and ultimately results in facilitated stimulus processing at attended locations (Bressler et al., 2008; Thiel et al., 2004). Based on this evidence, we hypothesized that the role of TPJ in social interaction relates to a modulation of visuo-spatial attention. Consistent with this hypothesis, first language (L1) learning studies proposed that sharing visual attention with a caregiver is essential in successful verbal learning (Kuhl et al., 2003; Waxman and Gelman, 2009), as a social partner directs the learner's attention toward the correct referent among possible targets (Tomasello, 2000; Dominey and Dodane, 2004; Verga and Kotz, 2013). However, adult learners typically use different strategies than children during word learning; for example, most often a situational context (i.e., the context a novel item is presented in) is sufficient to extract the correct meaning of a new word (Mestres-Misse et al., 2007, 2008). Yet, a situational context is not always informative. Does interactive learning help in this scenario? (Verga and Kotz, 2013, 2017)

To test this hypothesis, we used fMRI in a previously validated contextual learning game (Verga et al., 2015; Verga and Kotz, 2017) to investigate adult word learning in a simulated interactive context. Participants performed a visual interactive learning game, in which new

* Corresponding author. Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neuropsychology, Stephanstrasse 1a, Leipzig, 04103, Germany.
E-mail addresses: kotz@cbs.mpg.de, sonja.kotz@maastrichtuniversity.nl (S.A. Kotz).

words were repeated in different sentence contexts. Participants were originally divided into two groups to perform a behavioral training session: The "interactive" group played with a partner; they were also told that the same interactive set-up was going to be employed during the fMRI learning phase, but that the social partner would control the game from the console room because of technical constraints (cover story). The "non-interactive" group performed the task alone on a computer. Unbeknownst to participants, during the fMRI learning session the game was controlled by a computer program in both groups. To strengthen the cover story, the timing of the stimulus presentation was jittered to mimic the behavior of a human partner as observed in previous studies (Verga et al., 2015; Verga and Kotz, 2017; see the Methods section for details).

We hypothesized that the presence of a social partner may help the learner by directing her/his attention toward the correct target referent when contextual information is too minimal to learn and understand a new word's meaning (Tomasello, 2000; Dominey and Dodane, 2004; Verga and Kotz, 2013). Consequently, we expected activation in brain areas involved in orienting attention, such as the inferior posterior parietal cortex (Angular and Supramarginal Gyrus) and visual associative areas (e.g., Lingual Gyrus) whenever the learner "uses" a social partner to learn a new word's referent. Furthermore, if visuo-spatial attention represents an alternative learning strategy employed by learners (as compared to contextual learning), not only should different brain regions be recruited during learning, but also the outcome of learning (i.e., the accuracy to generalize learned words to novel contexts) should correlate with the activation of different brain areas.

2. Methods

2.1. Participants

45 participants took part in the experiment. 4 participants were discarded due to excessive movement during scanning. Hence, 41 participants (20F, mean age 26.98 ± 3.35 years) were included in the final analyses. All were native speakers of German, right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971), reported normal or corrected-to-normal vision, and none of them had any history of hearing or neurological disorders. All participants gave written informed consent, were paid for their participation, and were debriefed at the end of the experiment about its real purpose (for further details see Task and Procedure). The experiment was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Leipzig.

2.2. Material

The stimulus material and task employed in the current experiment have been described in detail elsewhere (Kotz et al., 2010; Verga et al., 2015; Verga and Kotz, 2017). In brief, the task required participants to find an object correctly completing a plausible German sentence consisting of a subject and a transitive verb (Sentence Context, SC) among several possible referents. All elements of the sentence as well as the possible referents were represented by pictures inserted in a visual checkerboard. For example, the SC "The boy eats" ("Der Junge isst") could be correctly completed by a picture representing an egg ("das Ei"), but not by pictures representing a map, a baby, a letter, or the actions to peel, to water, to pick (Table 1 and Fig. 1). SCs could either change from sentence to sentence, thus being differently informative each time (different Sentence Context, dSC) or stay consistent, thus being maximally informative (same Sentence Context, sSC). Importantly, while the amount of information provided in sSC and dSC sentences differed, the sentence structure and the control of linguistic parameters did not. We opted for this variation in sentence context to mimic immersive learning where learners are exposed to different objects in different sentence contexts (e.g., a piece of bread can be eaten, sliced, etc.) while others may appear in similar contexts (e.g., a knife should always be positioned on

Table 1

Examples of sentences presented during the experiment. Abbreviations: sSC = same Sentence Context, dSC = different Sentence Context.

Repetition	sSC		dSC	
	Sentence Context	Target Object	Sentence Context	Target Object
1	Die Fuchs beißt … *The fox bites …*	… das Huhn … *the* *chicken*	Der Junge isst … *The boy eats …*	… das Ei … *the egg*
2	Die Fuchs beißt …		Der Kellner schält … *The waiter peels …*	
3	Die Fuchs beißt …		Die Frau kocht … *The woman cooks* …	

the right side of a plate during each meal).

The pictures (330×245 pixels, 72 dpi) included 49 black drawings of objects, humans, animals, or actions (Bates et al., 2003; Székely et al., 2003, 2004, 2005; http://crl.ucsd.edu/experiments/ipnp/), inserted in a visual checkerboard. We selected pictures with high naming frequency (CELEX database) representing humans, animals, food, clothes, artifacts, and transitive actions. Pictures representing natural phenomena, multiple objects, objects with a compound name, general categories, and impersonal or intransitive verbs were not used. Objects appearing in the action pictures were modified when possible to avoid priming or interference effects (i.e., the human leg in the image representing the verb "to bite" was modified to avoid participants being primed or confused by it−see Fig. 1). Modifications were warped with the deformation option implemented in the graphic suite GIMP 2.6.8 (www.gimp.org; deformation radius = 15 mm, deformation amount = 100 mm). The pictures selected at this stage were combined to create all possible plausible German sentences, excluding those with the same subject and object (for example, "Der Hund verfolgt den Hund", the dog follows the dog) or an inanimate subject (for example, "Der Wind biegt den Baum", the wind bends the tree). The pictures representing nouns and used as subjects or objects of a sentence were balanced for naming frequency, gender, and animacy. The "objects" were divided in two balanced groups: Objects for sentences whose context was the same with each repetition (same Sentence Context, sSC) and objects for sentences whose context was different for each repetition (Different Sentence Context, dSC). The sentences were validated in two norming studies. In the first study, 20 native German speakers (9F, mean age 24.50 ± 2.74 years) were presented with 564 plausible transitive sentences, intermixed with 188 implausible fillers. They were asked to judge, on a scale from 1 to 5, sentence plausibility and how well the object completed a sentence. In a second study, participants were presented with three pictures (subject, verb, object) forming each sentence. 20 native German speakers (11F, mean age 24.10 ± 2.94 years) were presented with 376 sentences judged as plausible in the first validation study. Participants were required to write each sentence as they intended it and to evaluate its plausibility on a 1–5 scale. The agreement between the meaning of the picture and the participant's answer was calculated and images with plausibility scores below 3 or with an agreement rate below 50% were excluded.

The components of each sentence (i.e., pictures representing subject, verb, and object) were inserted in cells that touched each other at least corner to corner. This arrangement enforced that only one sentence led to the correct target object. Care was taken to ensure that each of the elements in the sentence appeared in each position of the grid a comparable number of times; similarly, all possible combinations of positions on the grid were employed a comparable number of times. The remaining six "distractor" pictures on each checkerboard were chosen from an initial image pool and were balanced between pictures representing nouns (either animals, humans, or objects) and actions. These distractors were selected to ensure that none of them could be considered as an additional plausible object for a sentence context; given the constraint that the
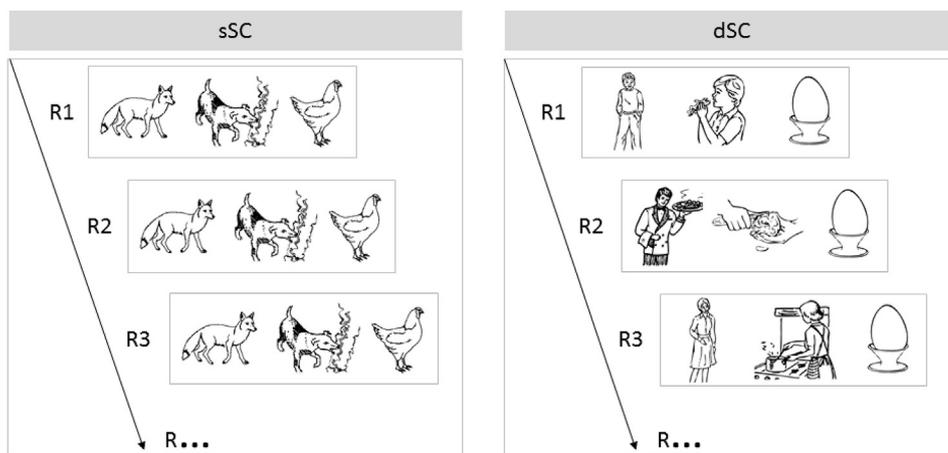
Fig. 1. Examples of pictorial sentences presented during the learning part of the experiment. In the same Sentence Context (sSC) condition, the target object ("the chicken") was presented each time when the same sentence was repeated ("the fox bites"). In the different Sentence Context (dSC) condition, the target object followed a different sentence context for each repetition (from top to bottom: "the boy eats", "the waiter peels", "the woman cooks"). 10 target objects were presented in each condition. Abbreviations: sSC = same Sentence Context; dSC = different Sentence Context; R = repetition.

elements of the sentence needed to be positioned close to each other, only one object on each checkerboard could represent the correct target object. The checkerboards were further balanced in terms of the mean naming frequency of the items depicted by the 9 pictures both within and between checkerboards.

All images were presented on a light gray background to avoid excessive luminance contrast inside the scanner. In total 180 checkerboards (20 target objects x 9 repetitions) were presented, each containing nine images (3×3).

Each target object (N = 20, each repeated 9 times) was associated with an Italian pseudo-word (Kotz et al., 2010). All pseudo-words were disyllabic strings (4–6 characters) balanced across conditions for syllabic complexity, initial, and final letter (either "a" or "o" to avoid confusion with the typical plural form of Italian words–usually "i" or "e", since all of the pictures depicted nouns in the singular form). Five German native speakers evaluated the pseudo-words to ensure that they were not

cognates of existing German words and that they "sounded Italian". In addition, the plausibility of the pseudo-words was evaluated by 15 Italian native speakers. All the pseudo-words were rated above 3 on a Likert scale from 1 to 5.

2.3. Experimental design

We employed a $3 \times 2 \times 2$ mixed design, with Run (Run1, Run2, Run3), Social Interaction (simulated, absent), and Sentence Context (same, different).

Social interaction was evaluated between-subjects: Participants of comparable age [U = 136.5, z = −1.93, p > .05] were pseudo-randomly assigned to a Simulated Social Interaction group (S+, N = 21, 10F, mean age 25.86 ± 2.87 years) or Non-Social Interaction group (S-, N = 20, 10M, 9F, mean age 27.95 ± 3.68 years).

SC was evaluated within-subjects: For the same participant, half of the objects (N = 10) occurred in the sSC condition (e.g., the image
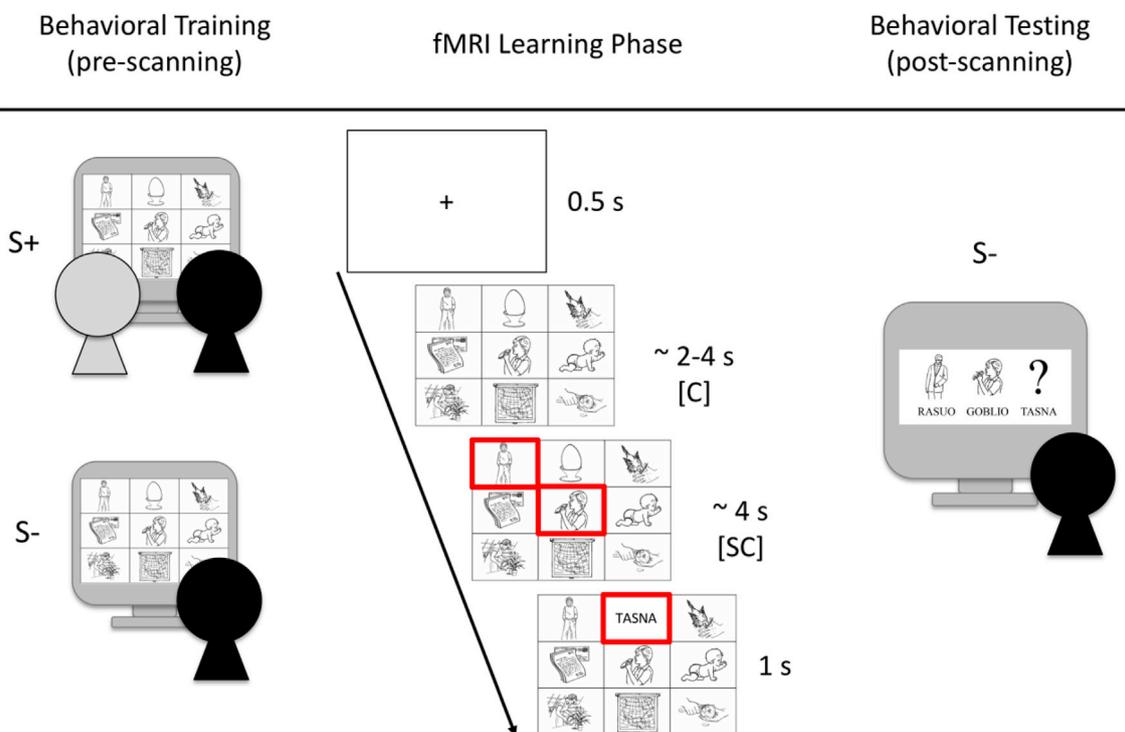

Fig. 2. Schematic representation of the experiment and trial characteristics. Upon their arrival at the scanner, participants were assigned to either one of two groups: S+ participants performed a behavioral training (10 trials; pre-scanning) with the experimenter as social partner; S- participants performed the same task alone. Participants then progressed into the fMRI learning phase. Lastly, participants performed a Behavioral Testing (post-scanning) outside the scanner.

representing "the chicken" always completed the SC "the fox bites"). The other half (N = 10) occurred in the dSC condition (e.g., the image for "the egg" followed at each occurrence a different SC, such as "the boy eats", "the waiter peels", "the woman cooks", etc.–see Table 1 and Fig. 1). The alternation between sSC and dSC checkerboards was randomized.

2.4. Task and Procedure

2.4.1. Behavioral training (pre-scanning) and task description

Participants received detailed instructions and were pseudo-randomly assigned to either the S+ or the S- group to initially perform a 10 practice trials training outside the scanner (Fig. 2). The participants' task was to find the correct ending (i.e., object) for a given SC amongst the checkerboard images. Each trial began with the presentation of a fixation cross (500 ms), followed by a checkerboard.

In the S- condition, after a jittered delay (range: 2.63–4.65 s, $M_{delay} = 3.59$ s ± 0.56 s) a red frame appeared around the image corresponding to the SC subject, followed after 500 ms by a second red frame surrounding the SC verb. Both frames remained on the screen until the object of the sentence was selected by the participant.

In the S+ condition, the selection of subject and verb was done manually by a social partner–the experimenter–sitting side by side with the participant in front of the computer. The rationale for having the experimenter as social partner instead of a confederate has been described elsewhere (Verga et al., 2015; Verga and Kotz, 2017).

Participants of both groups had to choose the correct target object from the remaining images. No time limit was imposed in this phase. In case of a correct response, a pseudo-word appeared immediately after the picture selection, in black capital letters (font Arial, size 40 points) over a light grey background in the selected object cell for 1000 ms. In case on incorrect response no pseudo-word was displayed, and the following trial began immediately.

2.4.2. fMRI learning phase

The learning phase took place in the fMRI scanner. S+ participants were told as a cover story that the experimenter's computer was located in the scanner's console room due to technical constraints. To further strengthen the cover story in between functional runs the experimenter asked S+ participants whether her pace while playing the game together was adequate, or if it had to be adjusted. However, for both groups the game was controlled by the same computer program following the procedure described for the S- training. In this phase participants had a maximum response time of 4 s; if they did not answer in this timeframe the next trial began.

Total scanning time was about 48 min, divided into 3 functional runs each containing 60 stimuli (3 repetitions for each of the 20 target objects) alternating with 20 null events (~length 12s). Visual stimuli were presented via back-projection (projector SANYO PLC-XP50L) onto a screen positioned at the rear-end of the bore and were visible to the participant via a mirror mounted on the head-coil.

Participants selected the correct object on the checkerboard with an fmri-compatible in-house touch pad positioned under the right hand and confirmed their choice by pressing a response-button with their left thumb. Since we were interested in cognitive processes occurring before the motor response (i.e., exploration of the visual stimulus and target identification), response hand was not counterbalanced.

2.4.3. Behavioral testing (post-scanning)

After learning, participants were behaviorally tested outside the scanner in a separate room. The test consisted in selecting the correct pseudo-word matching a SC that had not been presented during the learning phase in the scanner (Fig. 2).

Lastly, participants filled in a paper-and-pencil questionnaire with 12 questions to verify that S+ participants believed the cover story (e.g., "Was your partner in the game too slow?", "Were you sure your partner in the game was a real person?"). S- participants filled in a similar 12-questions questionnaire in which "the partner" was substituted with "the computer" (see Inline Supplementary Material).

All stimuli were presented using a desktop computer running Presentation 16.0 (Neurobehavioral Systems, Albany, USA). Two standard wheel mice (Logitech Premium Optical Wheel Mouse) were connected to the same Windows laptop and used as response devices for the S+ training. In all other test conditions only one standard wheel mouse (Logitech Premium Optical Wheel Mouse) was employed. At the end of the experiment, participants were debriefed about the real set-up and aim of the study. A new participant was admitted to the training room only after the previous one left, to avoid him/her hearing the debriefing. All S+ participants confirmed that they believed the cover story.

2.5. fMRI data acquisition

Brain images were acquired using a 3T whole-body Siemens TRIO scanner (Siemens Healthcare, Erlangen, Germany) at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. Functional whole-brain images were collected using a T2*-weighted gradient echo, echo-planar (EPI) pulse sequence, using the Blood Oxygenation Level Dependent (BOLD) contrast. Each functional volume consisted of 30 axial slices parallel to the AC-PC line with a slice thickness of 3 mm and a 33% inter-slice gap (TR = 2000 ms, TE = 30 ms, flip angle 90°, FOV 192.2 cm, matrix size 64 × 64, 116 kHz acquisition bandwidth). The first 2 volumes in each run were discarded to account for saturation effects.

A field map (short TE = 4.92 ms, long TE = 7.38 ms, echo spacing = 0.69 ms) was acquired for each participant at the beginning of the session to provide an estimate of local field inhomogeneity and later used in the spatial pre-processing of functional images.

Previously acquired high-resolution anatomical images were used for co-registration of the functional data. These images were T1 weighted 3D MP-RAGE structural scans with a sagittal orientation and a spatial resolution of 1 × 1x1.5 mm (TR = 1300 ms, TE = 3.93 ms, FOV 256 × 240 mm, 67 kHz acquisition bandwidth).

2.6. Behavioral data analysis

Statistical analyses of behavioral data were performed using MATLAB R2013a (The Mathworks Inc., Natick, USA) and IBM SPSS Statistics 18 (IBM Corporation, New York, USA).

Behavioral data were corrected for outliers (Response Times [RTs] ± 2 Standard Deviations [SDs]; mean of rejected trials across participants = 6%). Dependent variables were accuracy (proportion of correct responses), RTs of correct responses, and their SDs. In addition, a temporal coordination index was computed as correlation coefficients between participants' and experimenter/computer's RTs (Verga et al., 2015; Verga and Kotz, 2017). Unless otherwise specified, separate ANCOVAs were conducted on these variables with SC as a within-subject factor, and SDs of the SC presentation times as covariate. For the testing phase, separate ANCOVAs were conducted on accuracy scores and RTs with the mean number of pseudo-words repetitions for each object during the learning phase as a covariate.

When the assumption of sphericity was not met, a Greenhouse-Geisser correction was applied to the degrees of freedom. Two-tailed t-tests and simple-effect analyses were employed to compare individual experimental conditions, with an alpha level of p < .05 to ascertain significance. Bonferroni correction was applied when required to account for multiple comparisons.

2.7. fMRI data analyses

Data pre-processing and statistical analysis were conducted using SPM8 (Statistical Parametric Mapping, Wellcome Trust Centre for NeuroImaging, London, UK) running on MATLAB 7.11 (The Mathworks Inc., Natick, USA). For each participant, slices were corrected for differences

in acquisition time, realigned to the first image of the first session, and unwarped to correct for field inhomogeneity. Images were then normalized to the Montreal Neurological Institute (MNI) standard space and smoothed with an 8 mm full-width at half-maximum Gaussian isotropic kernel.

Data were statistically analyzed with a standard general linear model procedure in two stages. At the first level, regressors for the experimental conditions were convolved with a canonical hemodynamic response function. The design matrix included single regressors for the experimental conditions (sSC, dSC) in each specific moment of the learning task (i.e., onset of the checkerboards, SC identification). Error trials (including wrong responses as well as responses exceeding the maximum response time) were modeled as an additional regressor of no interest and functional run (1, 2 and 3) were included as constant regressors. Temporal series for each participant were filtered with a 100 Hz high-pass filter to remove slow signal drifts, and first-level Student's t-tests for each experimental condition were calculated (checkerboard [sSC], checkerboard [dSC], sentence context [sSC], sentence context [dSC]). Contrasts directly comparing the single conditions were also calculated at the first level (for example, sSC > dSC). At the second level, t-contrasts defined at the first level were fed into one-sample (within-subject comparisons) and paired-sample (between-subject comparisons) t-tests.

Correlations between brain activity and behavioral data were estimated with a multiple regression analysis at the second level. As between-group comparisons increase the variance in the sample and consequently reduce statistical power, we employed a cluster extent threshold criterion to correct for multiple comparisons (Forman et al., 1995). Monte Carlo simulations were conducted using the program AlphaSim implemented in AFNI software; the threshold for single voxels was set at p < .001 and a Monte Carlo simulation with 2000 iterations was run, resulting in an extent threshold of 30 voxels to achieve a corrected threshold of p < .05.

Images are displayed superimposed onto the skull-stripped mean anatomical image of the 41 subjects warped to the MNI space.

3. Results

3.1. Behavioral results

3.1.1. fMRI learning phase
During the learning phase, no differences emerged between S+ and S- on any of the relevant dependent measures (reaction times, standard deviations, correlation coefficients; all ps > .05).

3.1.2. Behavioral testing (post-scanning)
Participants' accuracy during the testing phase was 70% on average, with comparable performances in the two groups [$F(1,38) = 2.777$, $p = .104$, $\eta_p^2 = 0.068$]. However, participants in the S+ group were significantly slower [$M = 4.737$, $SD = 1.214$] than participants in the S- group [$M = 3.960$, $SD = 0.839$] [$F(1,38) = 6.767$, $p = .013$, $\eta_p^2 = 0.151$] (Fig. 3).

3.2. fMRI results

3.2.1. Learning related activity changes
In the following, we report results specifically related to our hypothesis and research questions. A complete overview of all the results, including learning and contextual effects, can be found in the Supplementary Materials.

Of particular interest with respect to how social interaction modulates learning are activity changes across time (i.e., across the three functional runs) during the exploration of the sentence context [S], identified by a Group (S+ vs. S-) x Run interaction testing for increases (run 1 > 2> 3) or decreases (run 3 > 2>1) of brain activity. We found a significant activation decrease in the right Supramarginal Gyrus (MNI: 57–34 37; z peak level = 3.82; k = 39; p < .05 AlphaSim corrected) (Fig. 4) in the S- group
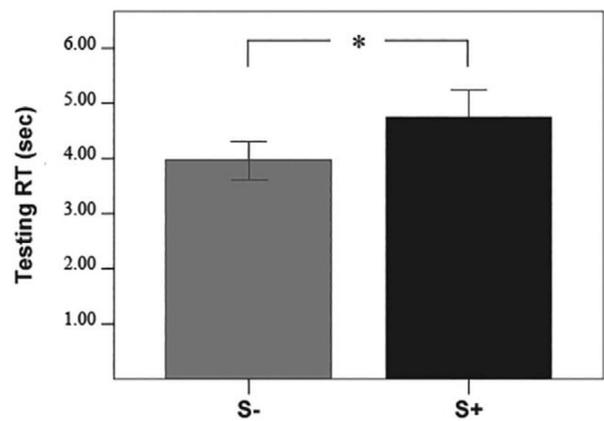

Fig. 3. Response times of correct responses during the behavioral testing (post-scanning). Vertical bars represent 95% CI. Abbreviations: S- = Non-Social Interaction; S+ = Simulated Social Interaction.
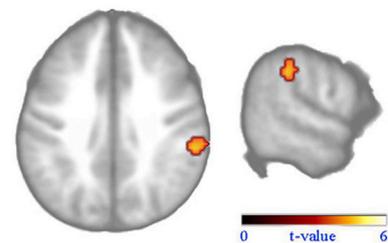

Fig. 4. Brain activity in the right Supramarginal gyrus for the contrast S+ > S- comparing the end of the learning phase to the beginning during the sentence context exploration [S].

as compared to the S+ during the SC exploration.

3.2.2. Correlation between brain activity and learning outcome
Accuracy scores obtained in the testing phase significantly correlated with activity in different brain regions during the exploration of dSC checkerboards [C] for the two groups of participants: The contrast S+>S- elicited an activation in the right Inferior Frontal Gyrus, pars orbitalis (MNI: 48 35–11; z at peak level = 3.91, k = 51, p < .05 AlphaSim corrected); the opposite contrast S->S+ correlated with activity in the right Lingual Gyrus (MNI: 12–70 -11; z at peak level = 4.17; k = 74, p < .05 AlphaSim corrected) (Fig. 5). No correlations were significant when looking at the same Sentence Context condition.

#### 3.2.3. Correlation between brain activity and temporal coordination during learning
In the S+ group more than in the S- group (contrast: S+>S-), activity of the right Caudate Nucleus (MNI: 12–1 16; z at peak level = 4.44; k = 126; p < .05 AlphaSim corrected) and Lingual Gyrus (MNI: 18–49 4; z = 3.90; k = 36, a at peak level = 3.90; p < .05 AlphaSim corrected) during the observation of the sSC checkerboards significantly correlated with the temporal coordination index (Fig. 6; see Methods 2.6 for details). For the opposite contrast (S- > S+) and for all contrasts in the dSC condition, no correlations were found to be significant.

4. Discussion

The current study employed a simulated interactive approach (Schilbach et al., 2013; Schilbach, 2014) to investigate the neural substrates of adult word learning. Results indicate that i) interactive word
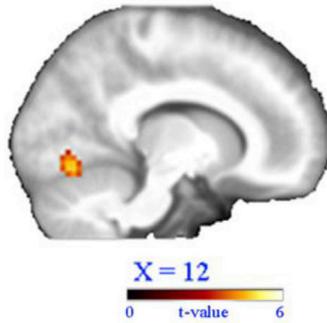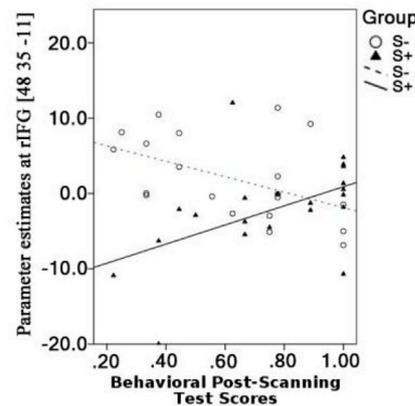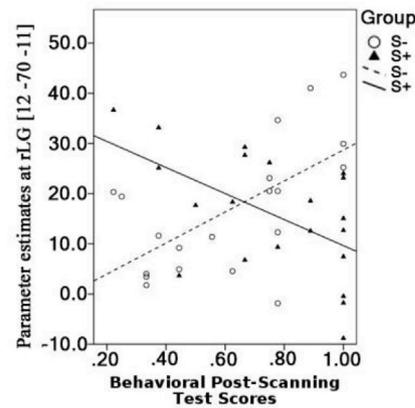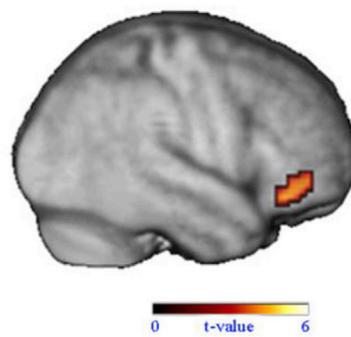
Fig. 5. Correlation between behavioral testing (post-scanning) accuracy and functional results of the contrast [C]dSC.

learning recruits the right Supramarginal Gyrus (rSMG); ii) in non-informative contexts, word retrieval correlates with activation of the right Inferior Frontal Gyrus (rIFG) in interactive learning and of the right Lingual Gyrus (rLG) in individual learning; d) in informative contexts, activity in right caudate and rLG during interactive learning is linked to temporal coordination.

4.1. The right Supramarginal Gyrus in interactive learning

The right SMG is involved in the retrieval of L2 words encoded in a social setting (Jeong et al., 2010) but it is also frequently reported in social cognition studies (Carter and Huettel, 2013). Furthermore, the same region is often activated in visuo-spatial attention tasks (Decety and Lamm, 2007) when attention is re-directed toward new targets (Corbetta et al., 2000) to facilitate the processing of a stimulus at an attended location (Thiel et al., 2004; Bressler et al., 2008). An explanation for this functional overlap is that during learning an interactive partner acts as an attentional beacon directing the learner's attention to a new word's meaning. This hypothesis has been extensively validated in children learning a first language (e.g., Dominey and Dodane, 2004; Tomasello, 2000).

It has been suggested that similar cognitive processes support word acquisition in L2 and L1 in social contexts (Rodriguez-Fornells et al., 2009; Jeong et al., 2010). Our results substantiate this suggestion: rSMG activity differentiated social and non-social learners in the exploration of a sentence context (i.e., the combination of subject and verb) leading to a set of target referents. Sentence contexts had exactly the same temporal and spatial characteristics in both groups; yet, the observed differences in activation patterns suggest that interactive rather than individual learning may be enough to "set" a specific learning strategy: If a partner is

believed to be present, learners rely on her/him to learn the target referent when the context provides minimal input; if the sentence context is maximal, they resort to an individual strategy (e.g., extract relevant information from context). The reliance on an interaction partner during learning makes intuitive sense, as it mirrors one of the most efficient learning strategies—learning from con-specifics (Whiten 2017)—from both a comparative as well as from an ontogenetic point of view. Indeed, learning from a knowledgeable partner is the most widespread mean of acquiring new knowledge in humans as well in several other species (Frith and Frith, 2012) and often relates to the ability of shift attention to follow a partner's gaze (Tomasello and Carpenter, 2007).

4.2. Retrieval of interactively vs. non-interactively learned words

If new words are differently encoded when learned interactively, different substrates are likely involved in their retrieval. Indeed, in the S+ group testing performance correlated with rIFG activity during learning. Activation of this region has been related to executive control such as the inhibition of competing memories during memory retrieval (Aron et al., 2004, 2014; Hampshire et al., 2010). For example, previous studies on social learning found IFG activation to reflect a mismatch between the learning context—social—and the testing context—non-social (Jeong et al., 2010). While this mismatch was also present in the current set-up, participants did not know in advance about the non-social testing phase; hence, they could not predict during learning that a mismatch between the learning and testing phase (that is learning interactively but being tested individually) would occur. A more likely explanation for this activation is the role of rIFG in attentional control, and, in particular, in the detection of salient stimuli (Hampshire et al., 2010) and inhibition of irrelevant ones (Sacchet et al., 2015): When
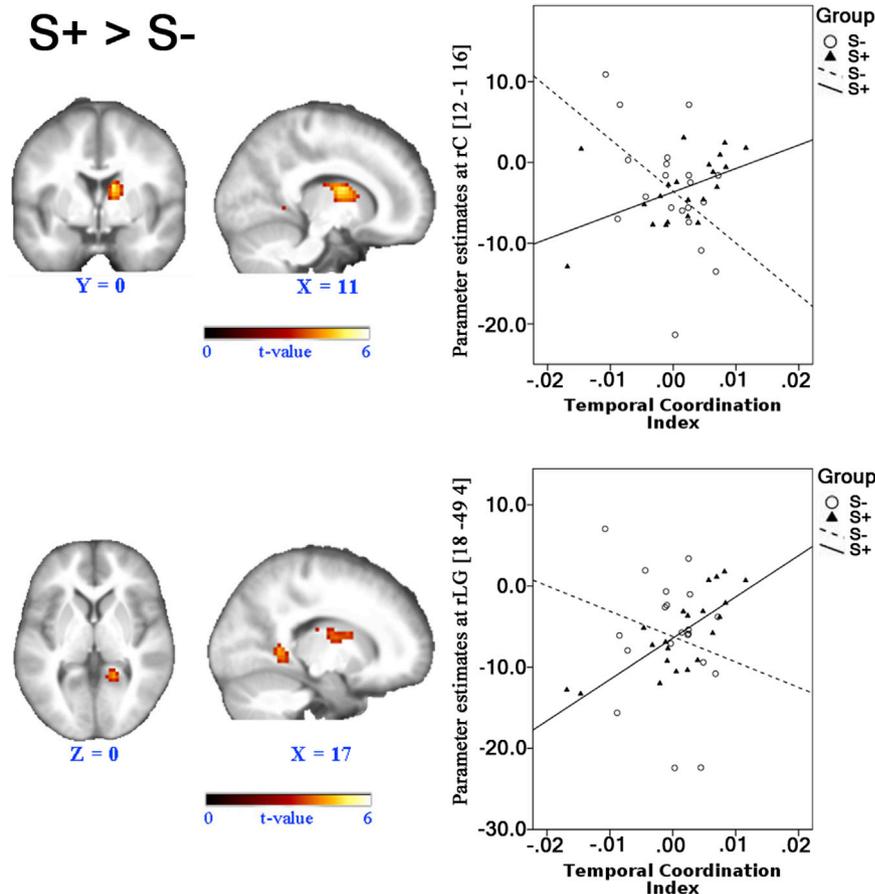
Fig. 6. Correlation between temporal coordination during learning and functional results of the contrast [C]sSC including the right Caudate and the right Lingual Gyrus.

faced with a new checkerboard in the non-informative condition (dSC), participants had to identify possible targets and filter out irrelevant ones; hence, the better they were at this task, the better the learning outcome. It is then noteworthy that rIFG activity correlated with learning outcome in the interactive group but not in the non-interactive group, supporting our initial hypothesis that social interaction works as an attentional beacon to facilitate the learner's attention to new word meaning (Verga and Kotz, 2017). Instead, during the observation of non-informative checkerboards, the behavioral test performance of the S- group correlated with activity in the right Lingual Gyrus, a visual region associated with the analysis of non-linguistic (Macaluso et al., 2000) and linguistic (Mechelli et al., 2000) visual stimuli. rLG activation increase has been reported in children with Autism Spectrum Disorder (ASD), a condition frequently associated with socio-attentional deficits (e.g., Fitzpatrick, 2018; Mundy and Jarrold, 2010; Schilbach et al., 2013): In a semantic category decision task, ASD children showed increased rLG activation as compared to normally developing children. This finding has been ascribed to a different strategy used by ASD children in lexico-semantic tasks, based on increased visualization of target items in lieu of more complex lexico-semantic processing of the target word (Gaffrey et al., 2007). Similarly, in the current study participants in the S- group, who did not have an interactive partner to rely on, employed a visualization strategy when analyzing the checkerboard, as opposed to a more partner-dependent strategy employed by the S+ group.

Taken together the current results suggest that interactive and individual learning may engage different learning strategies in search for a word referent in a visually non-informative learning context (e.g. a

visually displayed varying sentence context on the checkerboard). Slower reaction times observed during the behavioral post-scanning testing phase for the interactive learning group may thus reflect a difference in decoding strategy: If words encoded in interactive learning require visuo-spatial attention, visuo-spatial attentive mnemonics may be necessary to recall the stored information. In contrast, non-interactive learners, who used a simpler encoding based on visualization, would also have direct access to the learned word meaning. Both learning strategies seem to be equally effective in our task, as demonstrated by similar accuracy scores recorded in the testing phase.

4.3. Informative contexts and interactive learning

Is a social partner equally useful when the learner can easily identify a new word's referent in a situational context? We hypothesized that in this scenario adult learners would not rely on their partner to obtain information. Here, this condition was represented by repeating sentence frames on a checkerboard, in which participants simply had to retrieve the path to the target from memory. Nevertheless, the S+ group showed greater activation in the rLG—possibly reflecting the visual search process (Macaluso et al., 2000)—and right Caudate Nucleus, a subcortical region involved in number of cognitive, motor, and spatial functions, including goal directed actions (Hollerman et al., 2000; Jahanshahi et al., 2015) and the retrieval of previously learned routes (Brown and Stern, 2014).

Activity in the latter regions correlated with an index of temporal coordination between the learner and the "interactive partner": The

higher LG and caudate activity, the better the temporal coordination between partners during learning. This suggests that the well-known tendency for people in social situations to get "in synch" with each other (e.g., Richardson et al., 2007; Sebanz et al., 2006; Verga and Kotz, 2017) could sub-serve the exploration of a visual environment. This explanation would favor the hypothesis of a visuo-spatial attentional account for interactive learning in analogy to L1 even when the context is by itself informative; indeed, our task can be interpreted as a simplified visual natural environment, in which several possible referents for a new word are present. This is the same situation in which a child is learning new words; in both cases, the timing of the interaction between a learner and a partner is essential to focus on the same object at the same time—which would facilitate learning. It is also possible that the facilitation of navigating a visual display (the checkerboard), which is always the same—indicated by greater activity of LG and caudate—would reduce the brain's computation load, thus freeing cognitive resources that could then be re-directed to synchronizing with a partner (Koban et al., 2017). This account, however, would not explain why the same activation pattern was not observed in individual learners or why participants should want to synchronize with their partner in the first place.

4.4. Additional results

The learning of new words revealed differences in how the checkerboards were processed over time. Early in the learning process brain areas engaged in visual exploration and cognitive control (anterior and middle portions of the Cingulate Gyrus, the inferior Parietal Cortex and the Insula; Conci and Müller, 2012) as well as in language comprehension were activated (Temporal and Angular Gyrus; Binder et al., 2009). Towards the end increased activation was focused in primary and secondary visual areas (Inline Supplementary Material). These differential results suggest that early learning requires greater cognitive effort to navigate the checkerboards and identify the meaning of the picture sequences, while later learning more reflects the visual recognition of picture sequences. These results are further strengthened by the fact that a similar learning dynamic is evident for the Sentence Context, which shifts from "classic" left-lateralized language learning regions early on (i.e., Inferior Frontal, Angular, and Middle Temporal Gyrus – e.g., Mestres-Misse et al., 2008) to a predominantly "visual" pattern of activation encompassing the Calcarine Gyrus bilaterally toward the end of the experiment (Inline Supplementary Material).

As expected, participants processed same and different sentence context conditions quite differently: Checkerboards depicting consistent contexts stimulated participants to visually inspect and direct their attention towards the possible targets as soon as the checkerboards appeared. This is supported by the activation of a right-lateralized network encompassing Inferior Frontal Gyrus, Angular Gyrus, and Precuneus, typically involved in visuo-spatial attention. However, checkerboards depicting different sentence contexts required participants to recognize pictures constituting a hidden sentence, led to bilateral activation of the Fusiform Gyrus (Inline Supplementary Material). Interestingly, activations in the same region emerged in an interaction between group and sentence context type, suggesting that activation of the right Fusiform Gyrus may be up-regulated by perceived social interaction (Inline Supplementary Material). This interpretation is further supported by the results derived from the exploration of Sentence Context: consistent (same) sentence contexts elicited widespread bilateral activations in Superior Temporal Gyrus, Supramarginal Gyrus, and Cerebellum, in the right Cingulate Cortex, Putamen, and left Angular Gyrus, Amygdala, Insula. These combined areas are often reported in tasks dealing with spatial attention and planning of action execution, including covert speech production (Nardo et al., 2014; Thoma and Henson, 2011). Taken together, these findings complement the ones of the checkerboard observation: When a sentence context is known to a participant, s/he can focus on what to do next, namely locate and select the target object and possibly rehearse the corresponding pseudo-word. However, when the

sentence context changes with each repetition of a novel word, activation increases in areas involved in the retrieval and integration of syntactic and semantic information (Inferior Frontal Gyrus, bilaterally), together with areas involved in spatial analysis (Calcarine Gyrus, Occipital Gyrus) and areas involved in the identification and naming of visual stimuli (Lingual and Fusiform Gyrus; Inline Supplementary Material). This pattern of activation is consistent with the idea that participants work towards the construction of the sentence meaning.

In summary, these results suggest that when a context is constantly repeated, the decoding of a sentence is not necessary, while if a context is changing, the sentence context has to be decoded each time it occurs. Importantly, an increasing role of regions involved in visual analysis and recognition emerges as a result of the interaction between learning and type of context, in the checkerboard (Inline Supplementary Material) and sentence context observations (Inline Supplementary Material). This reflects a shift as learning progresses in a consistent environment, from more anterior regions involved in cognitive control to posterior regions supporting the analysis of the visual environment.

4.5. Limitations and future directions

Lastly, we put forward some caveats of the current study. Contrary to our previous studies (Verga et al., 2015; Verga and Kotz, 2017), we did not observe behavioral learning differences between S+ and S- participants, while such differences emerged during the testing phases which was conducted in close proximity with the experimenter. Hence, whether face-to-face interaction with a social partner is necessary to modify the learner's behavior is a question for further investigations. In addition, choosing same (semantically and visually) and different sentence contexts was motivated to mimic a natural visual learning situation. Thus both linguistic and visual variability may have influenced the results. A possible way to disentangle these effects would be to differentiate between conditions in which contextual and visual information are both consistently repeated (as in the current same Sentence Context condition), both varied (different Sentence Context), only varied visually, or only linguistically varied. Such a design would certainly help in clarifying whether social learning is driven by linguistic or visual information.

Another important venue for future research concerns the role of social interaction as rewarding or motivating. This proposition stems from the outcome of the post-testing paper and pencil questionnaire (Inline Supplementary Material). The primary scope of this questionnaire was to provide an objective measure to verify that S+ participants believed the cover story. Answers to indirect questions asking for details about the social partner's behavior (e.g., "How much do you think the presence of another person facilitated the task?"; "Did the partner/ computer adapted to your speed?") were particularly interesting: Participants' answers to these questions suggest that a human partner is perceived and adapted to. Rather unexpectedly, participants in the social group reported that they enjoyed the experiment significantly more than participants in the non-social group (Question1: "Did you enjoy the experiment?), that they were significantly more satisfied with the behavior of their game partner (Question2: "How satisfied were you with the behavior of your partner?"), and that they thought the presence of a human partner facilitated their task (Question 3: How much do you think the presence of another person facilitated the task?"). Consequently, a question for future studies could be whether social interaction (real or perceived) might push learning by engaging the reward system (Pfeiffer et al., 2014; Alkire et al., 2018).

5. Conclusion

The current study explored the neural correlates of contextual word learning in interaction. Results suggest that social learners employ different strategies when learning, largely dependent on visuo-spatial attention in minimally informative contexts. This evidence provides additional support for a proposal that individual cognitive processes may

be largely different from interactive ones (Schilbach et al., 2013). In addition, they support the notion that language learning in adults may rely on similar mechanisms as first language learning as long as its premises (i.e., being interactive) are kept constant in a naturalistic setting. Based on these outcomes, we suggest that a future direction to explore language learning in adults should focus on exploring its interactive basis in more naturalistic and ecologically valid settings (Verga and Kotz, 2018).

Acknowledgments

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2019.03.071.

References

Alkire, D., Levitas, D., Warnell, K.R., Redcay, E., 2018. Social interaction recruits mentalizing and reward systems in middle childhood. Hum. Brain Mapp. 39, 3928–3942. https://doi.org/10.1002/hbm.24221.

Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. Trends Cognit. Sci. 8, 170–177.

Aron, A.R., Robbins, T.W., Poldrack, R.A., 2014. Inhibition and the right inferior frontal cortex: one decade on. Trends Cognit. Sci. 18, 177–185.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C.C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., 2003. Timed picture naming in seven languages. Psychon. Bull. Rev. 10, 344–380.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebr. Cortex 19, 2767–2796.

Bressler, S.L., Tang, W., Sylvester, C.M., Shulman, G.L., Corbetta, M., 2008. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. J. Neurosci. 28, 10056–10061.

Brown, T.I., Stern, C.E., 2014. Contributions of medial temporal lobe and striatal memory systems to learning and retrieving overlapping spatial memories. Cerebr. Cortex 24, 1906–1922.

Carter, R.M., Huettel, S.A., 2013. A nexus model of the temporal–parietal junction. Trends Cognit. Sci. 17, 328–336.

Chambers, C.D., Payne, J.M., Stokes, M.G., Mattingley, J.B., 2004. Fast and slow parietal pathways mediate spatial attention. Nat. Neurosci. 7, 217–218.

Conci, M., Müller, H.J., 2012. Contextual learning of multiple target locations in visual search. Vis. Cognit. 20, 746–770. https://doi.org/10.1080/13506285.2012.694376.

Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat. Neurosci. 3, 292–297.

Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist 13, 580–593.

Dominey, P.F., Dodane, C., 2004. Indeterminacy in language acquisition: the role of child directed speech and joint attention. J. Neurolinguistics 17, 121–145.

Fitzpatrick, P., 2018. The future of autism research: dynamic and process-oriented approaches. J. Am. Acad. Child Adolesc. Psychiatry 57, 16–17.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.

Frith, C.D., Frith, U., 2012. Mechanisms of social cognition. Annu. Rev. Psychol. 63, 287–313.

Gaffrey, M.S., Kleinhans, N.M., Haist, F., Akshoomoff, N., Campbell, A., Courchesne, E., Müller, R.A., 2007. A typical participation of visual cortex during word processing in autism: an fMRI study of semantic decision. Neuropsychologia 45, 1672–1684.

Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., Owen, A.M., 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. Neuroimage 50, 1313–1319.

Hollerman, J.R., Tremblay, L., Schultz, W., 2000. Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. In: Progress in Brain Research. Elsevier, pp. 193–215.

Horwitz, B., Rumsey, J.M., Donohue, B.C., 1998. Functional connectivity of the angular gyrus in normal reading and dyslexia. Proc. Natl. Acad. Sci. Unit. States Am. 95, 8939–8944.

Jahanshahi, M., Obeso, I., Rothwell, J.C., Obeso, J.A., 2015. A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. Nat. Rev. Neurosci. 16, 719–732.

Jeong, H., Sugiura, M., Sassa, Y., Wakusawa, K., Horie, K., Sato, S., Kawashima, R., 2010. Learning second language vocabulary: neural dissociation of situation-based learning and text-based learning. Neuroimage 50, 802–809.

Koban, L., Ramamoorthy, A., Konvalinka, I., 2017. Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. Soc. Neurosci. 0:17470919.2017.1400463.

Kotz, S.A., D'Ausilio, A., Raettig, T., Begliomini, C., Craighero, L., Fabbri-Destro, M., Zingales, C., Haggard, P., Fadiga, L., 2010. Lexicality drives audio-motor transformations in Broca's area. Brain Lang. 112, 3–11.

Kuhl, P.K., Tsao, F.M., Liu, H.M., 2003. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. Proc. Natl. Acad. Sci. Unit. States Am. 100, 9096–9101.

Macaluso, E., Frith, C.D., Driver, J., 2000. Modulation of human visual cortex by crossmodal spatial attention. Science 289 (80-), 1206–1208.

Mechelli, A., Humphreys, G.W., Mayall, K., Olson, A., Price, C.J., 2000. Differential effects of word length and visual contrast in the fusiform and Lingual gyri during. Proc. R. Soc. Lond. Ser. B Biol. Sci. 267, 1909–1913.

Mestres-Misse, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., Münte, T.F., 2008. Functional neuroanatomy of meaning acquisition from context. J. Cogn. Neurosci. 20, 2153–2166.

Mestres-Misse, A., Rodriguez-Fornells, A., Münte, T.F., 2007. Watching the brain during meaning acquisition. Cerebr. Cortex 17, 1858–1866.

Mundy, P., Jarrold, W., 2010. Infant joint attention, neural networks and social cognition. Neural Network. 23, 985–997.

Nardo, D., Santangelo, V., Macaluso, E., 2014. Spatial orienting in complex audiovisual environments. Hum. Brain Mapp. 35, 1597–1614. https://doi.org/10.1002/hbm.22276.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Pfeiffer, U.J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A.L., Bente, G., et al., 2014. Why we interact: on the functional role of the striatum in the subjective experience of social interaction. Neuroimage 101, 124–137. https://doi.org/10.1016/j.neuroimage.2014.06.061.

Richardson, M.J., Marsh, K.L., Isenhower, R.W., Goodman, J.R., Schmidt, R.C., 2007. Rocking together: dynamics of intentional and unintentional interpersonal coordination. Hum. Mov. Sci. 26, 867–891.

Rodriguez-Fornells, A., Cunillera, T., Mestres-Misse, A., de Diego-Balaguer, R., 2009. Neurophysiological mechanisms involved in language learning in adults. Phil. Trans. Biol. Sci. 364, 3711–3735.

Sacchet, M.D., LaPlante, R.A., Wan, Q., Pritchett, D.L., Lee, A.K.C., Hamalainen, M., Moore, C.I., Kerr, C.E., Jones, S.R., 2015. Attention drives synchronization of alpha and beta rhythms between right inferior frontal and primary sensory neocortex. J. Neurosci. 35, 2074–2082.

Schilbach, L., 2014. On the relationship of online and offline social cognition. Front. Hum. Neurosci. 8, 278.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, K., 2013. Toward a second-person neuroscience. Behav. Brain Sci. 36, 393–414.

Sebanz, N., Bekkering, H., Knoblich, G., 2006. Joint action: bodies and moinds moving together. Trends Cognit. Sci. 11, 70–76.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A.L., Vargha, A., Bates, E., 2005. Timed action and object naming. Cortex 41, 7–25.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Bates, E., 2003. Timed picture naming: extended norms and validation against previous studies. Behav. Res. Methods Instrum. Comput. 35, 621–633.

Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C.C., Pechmann, T., Pléh, C., Wicha, N., 2004. A new on-line resource for psycholinguistic studies. J. Mem. Lang. 51, 247–250.

Thiel, C.M., Zilles, K., Fink, G.R., 2004. Cerebral correlates of alerting, orienting and reorienting of visuospatial attention: an event-related fMRI study. Neuroimage 21, 318–328.

Thoma, V., Henson, R.N., 2011. Object representations in ventral and dorsal visual streams: FMRI repetition effects depend on attention and part-whole configuration. Neuroimage 57, 513–525. https://doi.org/10.1016/j.neuroimage.2011.04.035.

Tomasello, M., 2000. The social-pragmatic theory of word learning. Pragmatics 10, 401–413.

Tomasello, M., Carpenter, M., 2007. Shared intentionality. Dev. Sci. 10, 121–125.

Verga, L., Bigand, E., Kotz, S.A., 2015. Play along: effects of music and social interaction on word learning. Front. Psychol. 6, 1316.

Verga, L., Kotz, S.A., 2013. How relevant is social interaction in second language learning? Front. Hum. Neurosci. 7, 550.

Verga, L., Kotz, S.A., 2017. Help me if I can't: social interaction effects in adult contextual word learning. Cognition 168, 76–90.

Verga, L., Kotz, S.A., 2018. Putting language back into ecological communication contexts. Lang. Cognit. Neurosci. https://doi.org/10.1080/23273798.2018.1506886.

Verghese, A., Kolbe, S.C., Anderson, A.J., Egan, G.F., Vidyasagar, T.R., 2014. Functional size of human visual area V1: a neural correlate of top-down attention. Neuroimage 93, 47–52.

Waxman, S.R., Gelman, S.A., 2009. Early word-learning entails reference, not merely associations. Trends Cognit. Sci. 13, 258–263.

Whiten, A., 2017. Social learning and culture in child and chimpanzee. Annu. Rev. Psychol. 68, 129–154.