

Help me if I can't

Citation for published version (APA):

Verga, L., & Kotz, S. A. (2017). Help me if I can't: Social interaction effects in adult contextual word learning. *Cognition*, 168, 76-90. <https://doi.org/10.1016/j.cognition.2017.06.018>

Document status and date:

Published: 01/11/2017

DOI:

[10.1016/j.cognition.2017.06.018](https://doi.org/10.1016/j.cognition.2017.06.018)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Original Articles

Help me if I can't: Social interaction effects in adult contextual word learning



Laura Verga^a, Sonja A. Kotz^{a,b,*}

^aMax Planck Institute for Human Cognitive and Brain Sciences, Dept. of Neuropsychology, Leipzig, Germany

^bFaculty of Psychology and Neuroscience, Dept. of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, The Netherlands

ARTICLE INFO

Article history:

Received 21 April 2016

Revised 16 June 2017

Accepted 16 June 2017

Available online 26 June 2017

Keywords:

Adult word learning
Contextual learning
Temporal coordination
Joint attention
Social interaction

ABSTRACT

A major challenge in second language acquisition is to build up new vocabulary. How is it possible to identify the meaning of a new word among several possible referents? Adult learners typically use contextual information, which reduces the number of possible referents a new word can have. Alternatively, a social partner may facilitate word learning by directing the learner's attention toward the correct new word meaning. While much is known about the role of this form of 'joint attention' in first language acquisition, little is known about its efficacy in second language acquisition. Consequently, we introduce and validate a novel visual word learning game to evaluate how joint attention affects the contextual learning of new words in a second language. Adult learners either acquired new words in a constant or variable sentence context by playing the game with a knowledgeable partner, or by playing the game alone on a computer. Results clearly show that participants who learned new words in social interaction (i) are faster in identifying a correct new word referent in variable sentence contexts, and (ii) temporally coordinate their behavior with a social partner. Testing the learned words in a post-learning recall or recognition task showed that participants, who learned interactively, better recognized words originally learned in a variable context. While this result may suggest that interactive learning facilitates the allocation of attention to a target referent, the differences in the performance during recognition and recall call for further studies investigating the effect of social interaction on learning performance. In summary, we provide first evidence on the role joint attention in second language learning. Furthermore, the new interactive learning game offers itself to further testing in complex neuroimaging research, where the lack of appropriate experimental set-ups has so far limited the investigation of the neural basis of adult word learning in social interaction.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Learning a new language is a complex task that an increasing number of adult learners is facing in the modern, multilingual world. New words—such as neologisms or English terms—can be encountered everyday when reading the newspaper or surfing the web. In this situation, the reader has to face a challenge, namely to assign a meaning to the new word (e.g., Horst, Scott, & Pollard, 2010; McMurray, Horst, & Samuelson, 2012). Despite the apparent simplicity of this task, every new word has multiple referents defined by cues that can be derived from the context a word is encountered in. However, next to a situational context, another

person may also provide these cues non-verbally; in fact, most often language learning contexts are social contexts, in which a more knowledgeable person facilitates the learner's efforts to acquire new words. This is certainly the case for infants learning their first language (Csibra & Gergely, 2009; Gleitman, Newport, & Gleitman, 1984; Kuhl, 2007; Kuhl, Tsao, & Liu, 2003), but it may also be crucial for adults acquiring a second language (Jeong et al., 2010; Verga, Bigand, & Kotz, 2015; Verga & Kotz, 2013).

First language (L1) acquisition studies suggest that social interaction with a caregiver is fundamental to successfully developing communication skills (Bruner, 1974, 1983). In particular, sharing visual attention with a caregiver is a sine qua non condition for successful infant word learning (Kuhl et al., 2003; Waxman & Gelman, 2009) as a caregiver may direct the infant's attention toward the correct referent among many possible referents (Dominey & Dodane, 2004; Tomasello, 2000; Verga & Kotz, 2013). This form of 'joint attention' (Tomasello & Akhtar, 1995)

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neuropsychology, Stephanstrasse 1a, Leipzig 04103, Germany.

E-mail address: kotz@cbs.mpg.de (S.A. Kotz).

during word learning may still apply in adult word learning (L2), as adult learners may also profit from joint attention, which could reduce the number of possible referents for a new word and improve learning (Louwerse, Dale, Bard, & Jeuniaux, 2012; Rader & Zukow-Goldring, 2012). Yet investigations in this domain are sparse.

Here, we investigated adult word learning taking on a social-interaction perspective. More specifically, we explored the possibility that adult learners—similarly to infants—may benefit from the presence of a knowledgeable partner and sharing attention when learning new words. In order to investigate this hypothesis, we developed and validated a learning game whose characteristics make it particularly well suited for social interaction studies even in complex learning settings (e.g. neuroimaging settings).

1.1. Social interaction in second language learning

Learning a new language in adulthood has long been considered as an imperfect process and strongly limited by the age of the learner; in other words, it has been hypothesized that only when a language is learned in early childhood is it possible to attain native-like proficiency (Lenneberg, Chomsky, & Marx, 1967; Penfield & Roberts, 1959). Decades of investigations in this domain allowed to discredit this reductionist view by highlighting how—in the right circumstances—even adult learners can attain native-like proficiency (Bongaerts, 1999; Bongaerts, Mennen, & Slik, 2000). For these learners, the pattern of activation when using L1 or L2 is remarkably similar (Abutalebi, 2008; Green, Crinion, & Price, 2006), leading to the hypothesis that L2 acquisition is based on an already specified L1 network, and receives convergent neural representation within the representations of the language learned as the L1 (Green, 2003). Importantly, the extent of convergence between L1 and L2 does not seem to depend upon the linguistic proximity between the two languages, as demonstrated in studies where L1 and L2 were, respectively, Italian and English or Catalan and Spanish (Perani et al., 1998) but also more distant languages such as English and Mandarin (Chee et al., 1999). Nevertheless, several factors have been proposed to contribute to a positive learning outcome, including age of acquisition (e.g., Bialystok & Hakuta, 1999; Birdsong, 1999), proficiency (Abutalebi, Cappa, & Perani, 2001; Perani et al., 1998), and exposure (Consonni et al., 2013; Perani et al., 2003).

Among these factors the role of exposure is perhaps the most elusive so far. Indeed, while it is self-evident that “exposure” includes social aspects, the extent to which these specifically contribute to L2 word learning remains largely unknown. A possible reason for this lies in the evidence that while social interaction may be important in word learning, we do know that adults may utilize other strategies when learning. For example, they often rely on a situational context when acquiring new words (e.g., they can obtain information from the situation the communication is taking place in—Laufer & Hulstijn, 2001; Nagy, Anderson, & Herman, 1987; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009; Swanborn & De Groot, 1999). In this scenario, the mapping of a word with its meaning is critically dependent upon how variable the sentence context a word is presented in is; however, as of yet there is still a debate whether a more variable or more consistent context is more beneficial for learning or not with evidence either showing a prevalence for learning effects in consistent contexts (Dempster, 1987; Hicks, Marsh, & Cook, 2005; Koffka, 1935; Steyvers & Malmberg, 2003; Young & Bellezza, 1982) or the importance of variable contexts as a successful mnemonic device (Hills, Maouene, Riordan, & Smith, 2010; Smith, 2000). In the latter case, variability is claimed to improve the generalization to novel items in particular (Perry, Samuelson, Malloy, & Schiffer, 2010).

Given the fact that adult learners may learn either alone or with a partner, the question that arises is: Which type of word learning—alone or with a social partner—is more beneficial for an adult learner? This question is indicative of two opposing theoretical accounts: On the one hand, adults could be considered as self-sufficient word learners, cognitively equipped to acquire any information they need; thus, a social partner could be expected to not or only minimally influence their learning behavior (Pickering & Garrod, 2006; Stephens, Silbert, & Hasson, 2010). Therefore, a situational context providing enough information regarding a referent’s characteristics (e.g., variable contexts, in which several cues suggest a word’s meaning—Borovsky, Kutas, & Elman, 2010; Hills et al., 2010; Perry et al., 2010; Smith, 2000) should suffice to identify the correct referent of a new word. On the other hand, research in social cognition and neuroscience suggests that not only is an adult’s behavior influenced by others, but also that this influence is qualitatively and quantitatively different when an adult is interacting with rather than merely observing someone (Bond & Titus, 1983; Ciaramidaro, Becchio, Colle, Bara, & Walter, 2014; Schilbach, 2014; Schilbach et al., 2013; Sebastiani et al., 2014; Zajonc, 1965).

One of the first studies investigating social word learning in adults—although not in a real-time social interaction—by Jeong et al. (2010), explored how adult Japanese speakers learned new Korean words. Participants watched movie clips depicting either a text-based learning context (new words were spoken by a person holding up their written translation) or a situation-based context (new words were exchanged between two actors). Post-learning functional magnetic resonance imaging (fMRI) revealed that the right supramarginal gyrus (rSMG) was involved in the retrieval of L2 words encoded in a social setting (Jeong et al., 2010). This result highlights the importance of considering social interaction not only as a context of a new word’s acquisition but also as a context, in which the word will be used in. Indeed, the context a newly learned language is used in is often a social one: Even when L2 is learned with a textbook or computer program, its final use is to communicate with others. As consistency between learning and testing environments has been suggested to facilitate recall (Godden & Baddeley, 1975; Polyn, Norman, & Kahana, 2009; Stein, 1978; Tulving, 1979), using an L2 acquired via text book learning in a social context may be more difficult than learning the language directly with a partner. Accordingly, new words learned and retrieved in mismatched conditions (for example, learned via text and retrieved socially) activate brain areas involved in conflict resolutions (such as the inferior frontal gyrus), while new words learned and used socially elicit brain activity similar to L1 words (Jeong et al., 2010). This latter evidence suggests that L1 and L2 words learned in a social context may exploit similar mechanisms during acquisition.

A powerful social mechanism employed by children when acquiring new words is sharing attention with a caregiver (Tomasello & Akhtar, 1995). When several possible meanings for a new word are available in a context, a social partner may direct the learner’s attention toward a new word’s correct referent, thus facilitating learning. Would adult learners also benefit from sharing attention with a knowledgeable partner? Theoretically, even for L2 this should significantly reduce the number of possible referents a word can take during learning especially when the context of a word presentation includes several possible meanings (Louwerse et al., 2012; Rader & Zukow-Goldring, 2012). Supporting this hypothesis, studies investigating social interaction often report the activation of the right temporo-parietal junction (TPJ) for social stimuli. While activation in this region is consistent in social neuroscience studies, this area also engages in joint attention and visuo-spatial attention (Decety & Lamm, 2007) and has

been attributed to the re-directing of attention toward new targets (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Krall et al., 2015).

1.2. Social interaction as an attentional device

Social learning contexts are substantially different from other types of word learning contexts (e.g., learning from a text book), as learning is facilitated by the partners' reciprocal influence to reach a shared goal (Rączaszek-Leonardi, Dębska, & Sochanowicz, 2014; Tylén, Weed, Wallentin, Roepstorff, & Frith, 2010). Indeed, whenever two people engage in a dialogue they exert a strong influence on each other at several levels; for example, they tend to use similar words and linguistic structures without being aware of it (Pickering & Garrod, 2004). Another good example is an asymmetric word learning setting—such as a parent and child interaction—, where the more experienced partner may adapt his/her behavior to guide the learner's attention toward the referent of a new word that is learned (Csibra & Gergely, 2009; Dominey & Dodane, 2004; Gogate, Bahrick, & Watson, 2000; Newman-Norlund, Noordzij, Meulenbroek, & Bekkering, 2007; Rader & Zukow-Goldring, 2012; Snow, 1972). Joint attention is known to significantly reduce the number of possible referents a new word can take on and to facilitate word learning (Louwerse et al., 2012; Rader & Zukow-Goldring, 2012).

Several L1 acquisition theories (e.g. Hollich et al., 2000; Markman, 1990; Tomasello, 2000) emphasize the role of a knowledgeable partner as an attentional enhancer and/or modulator. This has been ascribed to two main characteristics. First, a knowledgeable partner may be a particularly salient entity as s/he is conceived as a complex multi-modal stimulus (Gogate & Bahrick, 2001; Gogate, Walker-Andrews, & Bahrick, 2001). Second, a partner may attract attention by interacting with the learner in a contingent and timely manner compatible with her/his demands; in other words, the partner modifies her/his behavior to comply with the learner's needs (Hollich et al., 2000; Kuhl, 2007; Reddy & Uithol, 2015; Tomasello, 2000; Tomasello & Carpenter, 2007), which differs from other types of learning (e.g., learning from text books).

With regard to the first characteristic, the multi-modal nature of a social partner can work in favor of the learner. In just one interaction, s/he is exposed to several different cues (e.g., auditory, visual, somatosensory, linguistic, and meta-linguistic), each of which can be modulated in different ways (e.g., auditory stimuli can be modulated in prosody, content, noise made by movement, interjections, etc.). This abundance of information is difficult to find in other stimuli even when similarly complex (e.g., music; Verga et al., 2015; Wilson & Wilson, 2005). As redundancy of information coming from different modalities has been shown to facilitate the allocation of attention (Gogate & Bahrick, 2001; Schmidt-Kassow, Heinemann, Abel, & Kaiser, 2013), focusing on a human partner represents a major advantage for the learner (Fischer, Foth, Rohlfing, & Wrede, 2011; Gergely & Watson, 1999; Markova & Legerstee, 2006; Striano & Stahl, 2005; Szufnarowska, Rohlfing, Fawcett, & Gredebäck, 2014).

The second characteristic implies that a social partner attracts attention because of her/his timely adaptation to the learner's needs (Kuhl, 2007; Pereira, Smith, & Yu, 2008; Sage & Baldwin, 2010). Reports indicate that social partners tend to become temporally coupled (or coordinated) both in their behavior (Demos, Chaffin, Begosh, Daniels, & Marsh, 2012; Louwerse et al., 2012; Oullier, de Guzman, Jantzen, Lagarde, & Kelso, 2008; Pereira et al., 2008; Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Vesper, van der Wel, Knoblich, & Sebanz, 2012) as well as in their neural activity (Dumas, Lachat, Martinerie, Nadel, & George, 2011; Dumas, Nadel, Soussignan, Martinerie, & Garnero,

2010; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Jiang et al., 2012; Kawasaki, Yamada, Ushiku, Miyauchi, & Yamaguchi, 2013).

Importantly, these two characteristics are not mutually exclusive: The presence of a social partner may attract the learner's attention because it represents an economical way to obtain as much information as possible from a single source (Hollich et al., 2000; Markman, 1990; Tomasello, 2000); in turn, if the learner is paying attention to a knowledgeable partner, her/his reaction to the partner's behavior will be faster (Marinovic, Cheung, Riek, & Tresilian, 2014), leading to "time-locking" between the partner's action and the learner's response, and consequently to the evolution of coordinative dynamics. Hence, when a referent for a new word is presented, the learner will be maximally prepared to grasp its meaning. This supports the notion that sharing visual attention with a knowledgeable partner is a *sine qua non* condition for successful infant word learning (Kuhl et al., 2003; Waxman & Gelman, 2009), as a partner (e.g. caregiver) may direct the infant's attention toward the correct referent among many possible targets (Dominey & Dodane, 2004; Verga & Kotz, 2013).

1.3. The current study

Given the paucity of research on social second language learning, it is necessary to conceive and behaviorally validate a paradigm that is suitable for subsequent adult word learning studies in a social context. Such a paradigm should re-create a complex natural word-learning environment, yet it should be simple enough to comply with the requirements of experimental settings, including neuroimaging studies. Here we introduce such a paradigm and present its behavioral validation. The paradigm is a word learning game, in which participants are exposed to a visual learning environment that comprises several possible referents for a new word. The game is created in such a way that a new word can be learned either with or without a social partner by extracting information from a context that changes from trial to trial or stays consistent (e.g., providing more or fewer cues for a new word meaning). This set-up allows evaluating whether—and to what extent—a social partner facilitates the identification of a referent and, hence, word learning.

Concerning the identification of the referent, in order to label a new word with meaning the learner's attention needs to be guided to the correct referent at the correct point in time during the learning phase (Gogate et al., 2000, 2001; Pereira et al., 2008; Rader & Zukow-Goldring, 2012; Rolf, Hanheide, & Rohlfing, 2009); this can only be achieved if the learner's behavior is temporally attuned to a social partner. Operatively, this means that if a social partner facilitates the allocation of the learner's attention to a target referent, participants should be faster in identifying a target in a social learning condition. Furthermore, this should also be reflected in increased temporal coordination between partners in social learning compared to learning alone (e.g., computer learning). Indeed, temporal coordination between partners is pivotal to successfully join attention between the learner, the partner, and the new word referent. Importantly, this set-up exploits a contextual word learning approach in which the role of contextual variability is crucial (Glenberg, 1976, 1979): Every time the same word is encountered different contextual cues accumulate and ultimately facilitate the identification of the target referent and its association with the new word (Adelman, Brown, & Quesada, 2006; Lohnas, Polyn, & Kahana, 2011; Verkoeijen, Rikers, & Schmidt, 2004). Consequently, new words embedded in a same sentence context should elicit faster identification of the correct referent compared to new words embedded in a variable context. Indeed, if a new word is always repeated in the same context, the choice of a target word only needs to be replicated across repetitions after the first successful identification. Nevertheless, the mapping between words repeated

in a variable sentence context and their referents should improve: Each time the word is encountered, more cues will be available from a new context leading to an enriched representation of the word meaning (Adelman et al., 2006; Verkoeijen et al., 2004). Would the presence of a social partner influence the learner in these two sentence contexts similarly? We hypothesize that a social partner may maximally benefit the learner when the context varies as a priori the learner will not know where the target referent will be in the visual environment implemented by the game-like set-up. Conversely, words repeated in the same context do not require the experimenter to guide the learner's attention toward the target as the same target will be known from previous presentations in the same context. In addition, we hypothesize that the facilitation in the allocation of attention in the learning phase should also improve when learning with a social partner as reflect by better performance in the testing phase.

While the game is inspired by first language learning principles, it has been developed to specifically investigate second language learning from a social interaction standpoint. For this reason, we specifically tested adult learners who were instructed to learn new "Italian" words.

2. Materials and methods

2.1. Participants

Sixty-eight participants took part in the experiment (34 F, mean age 25.19 ± 2.88 years). All were native speakers of German recruited from the database of the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. Most participants only spoke one language (German), but English, French, and Russian were reported as second languages during schooling. All participants reported normal or corrected-to-normal vision, and none of them reported a history of hearing or neurological disorders. Right-handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants gave written informed consent and were paid for their participation. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Leipzig.

An experimenter (L.V., 26 years, female) was the game partner in the social condition. Using the same person to interact with all participants allowed controlling for interaction variability. Although it is possible that the experimenter adapted to each participant unintentionally, it is safe to expect a degree of stability in the experimenter's behavior. To further control for differences in interaction variability (social partner vs. computer), we used measures of variability (standard deviations of response times) as covariates in the statistical analyses of the data.

2.2. Stimuli

The stimuli to be learned consisted of twenty pseudo-words, representing the "Italian" name of 20 common objects (e.g., egg). The pseudo-words were presented to the participant as written words only if s/he was able to complete the task, which consisted in finding—among several possible referents—an object correctly completing a transitive sentence composed by subject and verb ("Sentence Context"), that could either change at each occurrence (different Sentence Context) or stay consistent (same Sentence Context). All elements of the sentence as well as the possible referents were represented by pictures inserted in a visual checkerboard.

The pseudo-words comprised 40 items (length range: min. 4 max. 6 letters) extracted from a set of Italian-based disyllabic

pseudo-words (Kotz et al., 2010). The selected sample of pseudo-words was balanced for syllabic complexity, initial letter, and final letter ("a" or "o"). We excluded words ending in "e" or "i" to avoid possible confounds with the Italian plural form as all the pictures contained singular elements. A list of the pseudo-words and their intended meaning is summarized in Table 1.

The pseudo-words were "hidden" in 220 checkerboards each containing nine images (330 width \times 245 height pixels, 72 dpi). Each image was centered in a different cell of the checkerboard (Fig. 1). The images depicted black-and-white drawings of objects, humans, animals, or actions selected from a validated database of pictures (Bates et al., 2003; Szekely et al., 2003, 2005; <http://crl.ucsd.edu/experiments/ipnp/>). A total of 79 images were selected, including 15 pictures representing humans or animals (category: Subject), 24 representing actions (category: Verb), and 40 representing objects, humans, or animals (category: Object). On each checkerboard, two nouns and an action were combined to form a simple transitive German sentence (Noun – Transitive Verb – Target Object. E.g., "Der Junge isst das Ei", "The boy eats the egg" – see Fig. 1). A total of 40 objects were chosen as Target Object ("das Ei") for the Sentence Contexts and associated with a different pseudo-word. Each target object and the associated pseudo-word could be presented a maximum of 11 times.

The pictures representing the elements of a sentence were arranged in such a way that the cells touched each other at least corner to corner. This constrained the game in the following way: (i) only one sentence could be created within each checkerboard, and (ii) only one object could be chosen to form a plausible German sentence based on the sentence context. The other six pictures on each checkerboard (excluding those belonging to the sentence) were distractor images chosen from the initial picture pool and were balanced between pictures representing nouns (either animals, humans, objects), and actions. These distractors were selected to ensure that none of them could be considered as an additional plausible object for a given Sentence Context. The checkerboards were further balanced for the mean naming frequency of the items depicted by the 9 pictures; moreover, each element of the target sentence (subject, verb, object) appeared a comparable number of times in each cell. All possible dispositions of the three target pictures were employed a comparable number of times.

Table 1

Summary of the target objects presented in the experiment and the pseudo-words (items to be learned) corresponding to each of them. Abbreviations: dSC = different Sentence Context, sSC = same Sentence Context.

dSC		sSC	
Target object	"Italian" name	Target object	"Italian" name
Brot (<i>bread</i>)	soro	Kuh (<i>cow</i>)	vicco
Buch (<i>book</i>)	lanto	Bär (<i>bear</i>)	recuo
Ei (<i>egg</i>)	tasna	Knochen (<i>bone</i>)	piusca
Glas (<i>glass</i>)	gualvo	Schwein (<i>pig</i>)	nago
Mädchen (<i>girl</i>)	zorta	Baum (<i>tree</i>)	suota
Pferd (<i>horse</i>)	renio	Schmetterling (<i>butterfly</i>)	zuccia
Schaf (<i>sheep</i>)	vonna	Möhre (<i>carrot</i>)	gubo
Baby (<i>baby</i>)	ciato	Clown (<i>clown</i>)	cagia
Brief (<i>letter</i>)	cavria	Eis (<i>ice-cream cone</i>)	lasdio
Hase (<i>rabbit</i>)	muga	Paket (<i>package</i>)	terto
König (<i>king</i>)	miervo	Getreide (<i>grain</i>)	fadia
Mais (<i>corn</i>)	zaso	Fahrrad (<i>bike</i>)	vuora
Pfarrer (<i>priest</i>)	noda	Birne (<i>pear</i>)	rupo
Vogel (<i>bird</i>)	sestia	Torte (<i>cake</i>)	pano
Karte (<i>map</i>)	pionto	Ziege (<i>goat</i>)	lafria
Blume (<i>flower</i>)	ruoca	Matrose (<i>sailor</i>)	goblio
Maus (<i>mouse</i>)	gispia	Kleid (<i>dress</i>)	tausa
Rose (<i>rose</i>)	rasuo	Ball (<i>ball</i>)	miecrha
Schüssel (<i>bowl</i>)	vedia	Brille (<i>glasses</i>)	mifro
Zwiebel (<i>onion</i>)	viato	Hose (<i>trousers</i>)	sialva

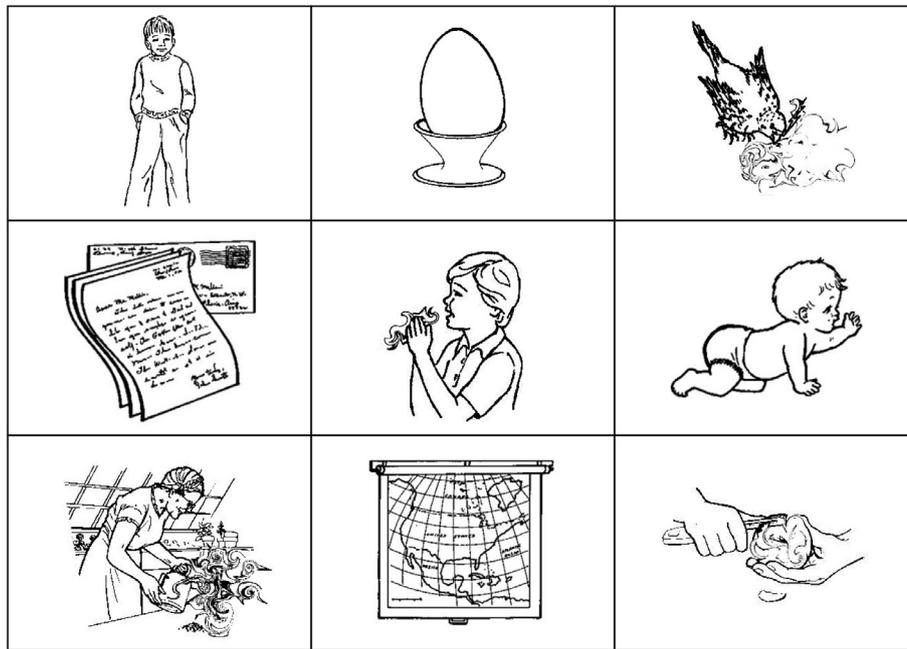


Fig. 1. Example of checkerboard used in the experiment. The hidden sentence is in this example composed of the pictures representing a young boy, the act of eating, and the object egg. The sentence “The boy eats the egg” is the only plausible sentence that can be constructed within the given constraints. Elements depicted in the checkerboard are in the first row (top) from left to right: boy (noun), egg (noun), to pick (verb); second row (middle) from left to right: letter (noun), to eat (verb), baby (noun). Third row (bottom) from left to right: to water (verb), map (noun), to peel (verb). Images are reproduced from <http://crl.ucsd.edu/experiments/ipnp/index.html>.

2.3. Experimental design

We manipulated two factors: two levels of social interaction (present, absent) and two levels of Sentence Context (same, different). These two factors were tested in a between-subjects design; thus, every participant was randomly assigned to one of four conditions: Social Interaction and different Sentence Context (S+, dSC – $n = 17$, 9 F, mean age 25.71 ± 3.12), Social Interaction and same Sentence Context (S+, sSC – $n = 17$, 8 F, mean age 25.53 ± 2.81), no Social Interaction and different Sentence Context (S–, dSC – $n = 17$, 9 F, mean age 24.06 ± 3.29), and lastly no Social Interaction and same Sentence Context (S–, sSC – $n = 17$, 8 F, mean age 25.47 ± 2.13). There was no age difference between participants in the four groups ($F(3,64) = 1.197$, $p = 0.318$, $\eta_p^2 = 0.053$).

To evaluate the influence of a social partner on the learning process, participants assigned to the social condition (S+) performed the task together with the experimenter; participants in the non-social condition (S–) performed the task alone at a computer (see “Task and Experimental Procedure”). To evaluate the effect of sentence context variability, we split the pool of target objects into two groups: half of the objects ($n = 20$) occurred repetitively within the same sentence context (sSC). For example, the image representing “cow” was always the correct ending for the same sentence context “the wolf bites”. The other half of the objects ($n = 20$) was presented within a different sentence context (dSC). For example, the image representing “egg” could be presented in either one sentence contexts such as “the woman cuts”, “the boy eats”, etc. (Table 2).

Although each sentence was repeated 11 times, the actual number of exposures to each pseudo-word was dependent on the number of correct responses given by the participants, as a pseudo-word was presented only in the case of a correct response.

2.4. Task and experimental procedure

The experiment consisted of three parts: practice trials, learning phase, and testing phase. 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 computer and used as response devices. The task specifics are described below and displayed in Fig. 2.

2.4.1. Practice trials and learning phase

In these phases, the participant’s task was to find amongst the pictures on the checkerboard the correct final object following a given sentence context. In all conditions, the trial began with the presentation of a fixation cross (500 ms), followed by a checkerboard. On each checkerboard the participant was provided with a sentence context (Fig. 2): In the S+ condition, the participant and the experimenter were sitting side by side in front of the same computer screen; each member of the dyad was controlling a differently colored cursor on the screen through his/her own computer mouse: The experimenter controlled a light blue squared-shaped cursor, while the participant’s cursor was green. When the checkerboard containing the nine images appeared, the experimenter selected the subject and verb of the target sentence, by left clicking on them with the mouse in rapid succession. A light blue frame appeared around each picture immediately after its selection and remained on the screen until the participant had provided an answer. The experimenter tried not to adapt specifically to each participant, but instead to keep a constant behavior. This decision was made to ensure maximal compatibility with the non-social conditions, in which there was no adaptation to the participant’s behavior by the computer.

In the S– condition, the sentence context was automatically selected by the computer, with a constant delay of 1000 ms between the markings of the “Subject” and “Verb” pictures. As in the S+ condition, a light blue frame appeared around the picture immediately after its selection and remained on the screen until the participant had provided an answer. In both conditions, the selection of the verb represented a go-signal for the participant to identify the correct target object by left clicking on it with the mouse. There was no time limit. The picture selected by the

Table 2

Example of sentences conveyed by the images used in the experiment. Each sentence target object was repeated 11 times during the experiment. Abbreviations: dSC = different Sentence Context; sSC = same Sentence Context.

Repetition	dSC		sSC	
	Sentence context	Target object	Sentence context	Target object
1	Der Junge isst ... <i>The boy eats ...</i>		Der Wolf beißt ... <i>The wolf bites ...</i>	
2	Der Kellner schält ... <i>The waiter peels ...</i>		Der Wolf beißt ...	
3	Der Junge kocht ... <i>The boy cooks ...</i>		Der Wolf beißt ...	
4	Der Mann isst ... <i>The man eats ...</i>		Der Wolf beißt ...	
5	Der Soldat kocht ... <i>The soldier cooks ...</i>		Der Wolf beißt ...	
6	Der Hund zerbricht ... <i>The dog breaks ...</i>	... das Ei ... <i>the egg</i>	Der Wolf beißt die Kuh ... <i>the cow</i>
7	Die Frau isst ... <i>The woman eats ...</i>		Der Wolf beißt ...	
8	Die Katze zerbricht ... <i>The cat breaks ...</i>		Der Wolf beißt ...	
9	Die Frau schält ... <i>The woman peels ...</i>		Der Wolf beißt ...	
10	Die Frau schneidet ... <i>The woman cuts ...</i>		Der Wolf beißt ...	
11	Der Soldat isst ... <i>The soldier eats ...</i>		Der Wolf beißt ...	

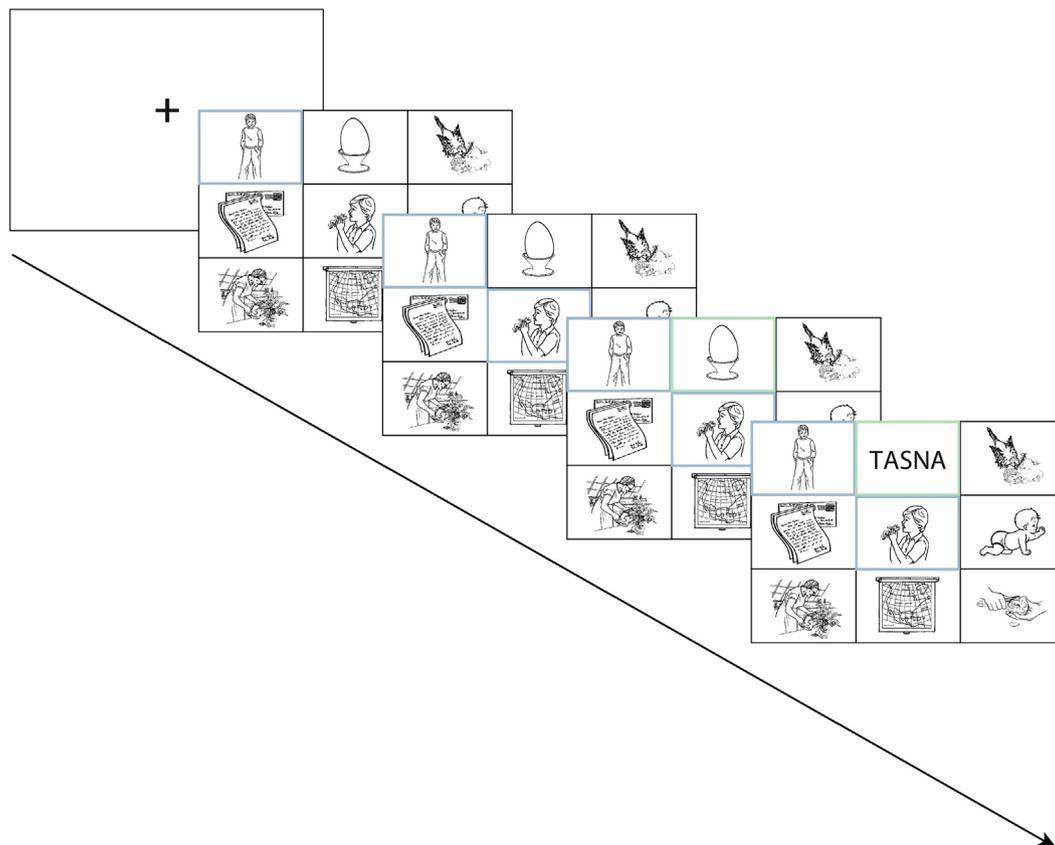


Fig. 2. Example of an experimental trial. In the social condition, the experiment selected the Subject and Verb pictures. In the non-social condition, a computer program selected these pictures.

participant was immediately marked with a green square; when the provided answer was correct, the selected image was substituted by a pseudo-word providing the “Italian name” of the object. The pseudo-words were presented in black capital letters (font Arial, size 40 points) on a white background in the cell selected

by the participant, and remained on the screen for 1000 ms. In case of an incorrect response, no “Italian name” was displayed and the following trial began immediately (Fig. 2).

After performing ten practice trials, participants started the learning phase, whose procedure was identical to the training

phase. Two hundred and twenty trials (20 objects \times 11 repetitions) were presented during the experiment. See Picture 2 for a graphical representation of a trial.

2.4.2. Testing phase

The testing phase consisted of two tasks widely employed in learning and memory research: a recognition task and a recall task (Glenberg, 1976, 1979; Lohnas et al., 2011; Polyn et al., 2009; Swanborn & De Groot, 1999; Verkoeijen et al., 2004; Fig. 3). In the recognition task, an object picture was presented together with a pseudo-word that participants had seen during the learning phase; participants were asked to indicate whether the two elements belonged to each other or not based on what they had learned in the learning phase. Picture-word associations were correct in 70% of the trials and incorrect in the remaining 30% of the trials. In the recall task, participants were presented with one of the target objects and asked to type in the pseudo-word associated with the object during the learning phase. No time limit was imposed. Both in the recognition as well as in the recall task, all 20 objects were presented in a randomized order that was different for each participant. The two tasks were presented in counterbalanced order across participants: Half of the participants were first tested with the recognition task and then with the recall task, while the order was inverted for the other half. All participants underwent the same testing phase individually (i.e., without the experimenter), irrespective of the condition they were exposed to during the learning phase.

2.5. Data analyses

Statistical analyses were performed using MATLAB R2013a (The Mathworks Inc., Natick, USA) and IBM SPSS Statistics 18 (IBM Corporation, New York, USA). Behavioral data were first corrected for outliers; trials with response times exceeding the mean response times (RTs) \pm 2 standard deviations (SDs) were excluded from further analysis (mean of rejected trials across participants = 6.33%).

For the learning phase, response times were calculated as the time between the appearance of the “verb” picture and the participant’s answer. Accuracy scores (proportion of correct responses), response times of correct responses, and their standard deviations were calculated for each repetition of the object and each participant. To evaluate the degree of temporal coordination of the learner and the social partner during the learning phase, we used the SDs of response times as an index of stability in the participants’ performance; in other words, the higher the SDs, the less stable (or more variable) the performance. Further, we calculated the lag-0 cross correlation coefficients between the inter-trial intervals produced by the participant (i.e., the time delay between the highlighted verb picture and the selection of the object picture) and

those produced by the social partner (S+ condition) or computer (S– condition; i.e., the time delay required from the appearance of the checkerboard to the selection of the subject of the sentence). These measures indicate how much the behavior of the participant in one trial is temporally related to the behavior of the partner (experimenter/computer) in the same trial. 2×2 repeated measures ANCOVAs were conducted on the outcome measures of the learning phase, with the between factors Sentence Context (sSC, dSC) and Social Context (S+, S–), and the within factor Repetition (11 repetitions). To account for differences in the variability of trial presentation in the different conditions (i.e., variable timing for social partner and fixed for the computer), the SDs of the social partner’s response/computer times were used as a covariate in the data analysis of the learning phase. We did not covary for SDs in the cross-correlation analyses as SDs account for the variability in the computer/experimenter RTs series based on which the correlation coefficients are calculated.

To evaluate the outcome measures of the testing phase (either recognition or recall), we performed two separate 2 (Social Context: S+ vs S–) \times 2 (Sentence Context: sSC vs dSC) ANCOVAs on the accuracy scores, with the covariate being the mean number of repetitions participants were exposed to during the training phase. For the recognition task, response times were calculated as the time between the appearance of the word/picture combination and the participant’s response; accuracy scores were defined as the proportion of correct responses. For the Recall task, response times of correct responses were calculated as the time between the appearance of the image and the button press to move on to the next trial. Only words perfectly recalled (i.e., the recalled pseudo-word was identical to the one presented in the learning phase) were considered correct. In both testing tasks, we used the number of exposures during the learning phase as a covariate. This number is directly related to the number of correct responses and it takes into account the mean number of times pictures were repeated during the learning phase, ranging from a minimum of 0 (no correct responses) to a maximum of 11 times (no errors).

When the assumption of sphericity was not met, the 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. We used an alpha level of $p < 0.05$ to ascertain significance for all statistical tests.

3. Results

3.1. Learning phase

Participants achieved an average accuracy of 93%. Accuracy scores increased over time (main effect of Repetition, linear trend, $F(5.78, 364.35) = 13.06$, $p = 0.000$, $\eta_p^2 = 0.172$) similarly for all

Testing Phase	Event	Task
Recognition	 TASNA	Say if the word is the correct “Italian name” of the represented object (yes/no)
Recall		Type in the “Italian name” of the represented object.

Fig. 3. Example of stimuli used during the testing phase.

experimental conditions (i.e., no interactions, all p s > 0.102). We observed a marginally significant difference between sSC and dSC ($F(1,63) = 3.390$, $p = 0.070$, $\eta_p^2 = 0.051$), with greater accuracy when words were repeated in same Sentence Contexts ($M = 0.956$, $SD = 0.048$) compared to different Sentence Context ($M = 0.934$, $SD = 0.046$). There were no other significant results or interactions (all p s > 0.155; Fig. 4).

RTs of correct responses significantly decreased over the course of the learning phase (main effect of Repetition, linear trend, $F(4.02, 253.31) = 31.538$, $p = 0.000$, $\eta_p^2 = 0.334$) similarly for all experimental conditions (no interactions, all p s > 0.521). Furthermore, participants were faster when learning socially ($S+$, $M = 3.013$, $SD = 1.310$) than non-socially ($S-$, $M = 3.711$, $SD = 1.486$; $F(1,63) = 9.496$, $p = 0.003$, $\eta_p^2 = 0.131$). For Sentence Context, RTs were faster in the sSC ($M = 2.567$, $SD = 1.165$) as compared to the dSC ($M = 4.157$, $SD = 1.232$) condition ($F(1,63) = 17.345$, $p = 0.000$, $\eta_p^2 = 0.216$). There was no interaction between the factors ($p = 0.361$; Fig. 5).

Standard deviations of response times decreased over time (linear trend, $F(6.30, 396.56) = 7.610$, $p = 0.000$, $\eta_p^2 = 0.108$) in a comparable way for all experimental conditions (no interactions, all p s > 0.377). Participants trained in a social context ($S+$, $M = 1.560$, $SD = 0.738$) were less variable than participants trained non-socially ($S-$, $M = 0.201$, $SD = 0.958$) ($F(1,63) = 7.434$, $p = 0.008$, $\eta_p^2 = 0.106$). There were no other significant main effects or interactions (all p s > 0.113; Fig. 6).

The ANOVA conducted on the lag-0 cross-correlations coefficients shows that this index did not change over time (no main effect of Repetition, $F(10,640) = 0.309$, $p = 0.979$, $\eta_p^2 = 0.005$). Participants trained in the social condition ($S+$, $M = 0.301$, $SD = 0.028$) had higher correlation coefficients compared to $S-$ participants ($M = 0.129$, $SD = 0.028$; $F(1,64) = 97.212$, $p = 0.000$, $\eta_p^2 = 0.603$). There was no significant difference between sSC and dSC ($F(1,64) = 0.200$, $p = 0.656$, $\eta_p^2 = 0.003$). There was, however, a significant interaction between Sentence Context and Social Context ($F(1,64) = 6.178$, $p = 0.016$, $\eta_p^2 = 0.088$). Thus, we computed a simple effect analysis showing that there was no difference between the levels of sentence context in the non-social group ($F(1,64) = 2.078$, $p = 0.154$). However, in the social group, the correlation coefficients were higher in the different Sentence Context condition ($M = 0.425$, $SD = 0.033$) than in the same Sentence Context condition ($M = 0.329$, $SD = 0.031$; $F(1,64) = 4.300$, $p = 0.042$) (Fig. 7).

To summarize, while accuracy levels were comparable across conditions, significant differences emerged in the temporal behavior of participants as a function of the type of interaction (social or non-social) to which they were exposed to: Standard deviations,

response times, and lag-0 cross-correlations all showed significant differences between participants trained socially and non-socially. Sentence context variability influenced response times, with lower RTs for words embedded in a consistent sentence context. Further, only for the participants trained socially, words embedded in a more variable sentence context yielded higher values of lag-0 cross-correlations.

3.2. Testing phase

3.2.1. Recognition task

The mean accuracy rate of the recognition task was 87%, significantly above chance level (1 sample t -test against a chance level of 0.50, $t(67) = 22.772$, $p = 0.000$, 95% C.I. (0.333, 0.397)).

Accuracy scores did not differ between participants trained socially and participants trained on a computer, nor between sSC and dSC (all p s > 0.320). However, the interaction between the factors was significant ($F(1,63) = 5.965$, $p = 0.017$, $\eta_p^2 = 0.086$). Therefore, an analysis of simple effects was carried out. This analysis revealed that sSC words were recognized significantly better by participants trained non-socially ($M = 0.901$, $SD = 0.304$) as compared to participants trained socially ($M = 0.793$, $SD = 0.034$; $F(1,63) = 6.111$, $p = 0.016$). Rather, recognition accuracy for dSC words was comparable for participants trained socially and non-socially ($F(1,63) = 1.022$, $p = 0.316$). However, participants trained socially recognized dSC words better ($M = 0.904$, $SD = 0.036$) than words repeated in the same sentence context ($M = 0.793$, $SD = 0.034$; $F(1,63) = 4.114$, $p = 0.047$), an effect not found in the non-social group ($F(1,63) = 0.761$, $p = 0.386$) (Fig. 8).

The ANCOVA conducted on the RTs did not show any significant effect or interaction (all p s > 0.389).

In summary, recognition accuracy scores revealed a significant interaction between sentence context and social interaction: Participants trained non-socially recognized words encoded in a consistent context better, while participants trained socially recognized words encoded in a variable context better.

3.2.2. Recall task

The recall task was more challenging for the participants as revealed by the overall low number of correctly recalled items (40%). Accuracy scores differed for Sentence Context, with dSC revealing higher scores ($M = 0.521$, $SD = 0.064$) than sSC ($M = 0.298$, $SD = 0.067$; $F(1,63) = 4.671$, $p = 0.034$, $\eta_p^2 = 0.069$). However, there was no effect of Social Context ($F(1,63) = 2.232$, $p = 0.140$, $\eta_p^2 = 0.034$), nor an interaction between the two factors ($F(1,63) = 2.456$, $p = 0.122$, $\eta_p^2 = 0.038$) (Fig. 9). There were no significant differences for RTs in the recall task for any of the

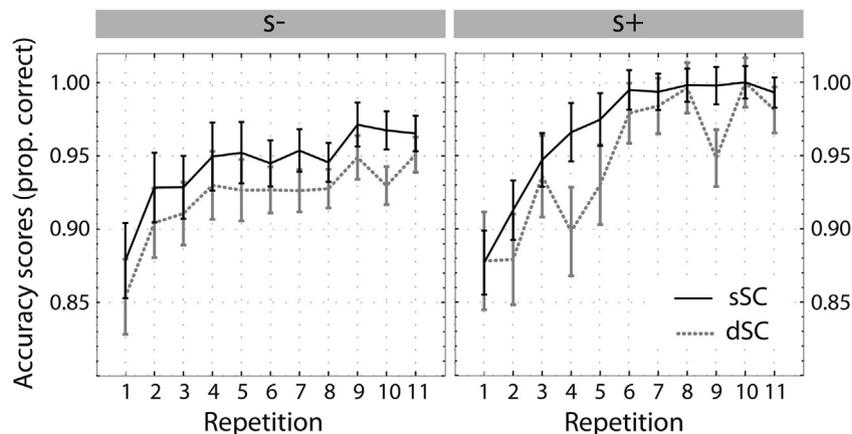


Fig. 4. Accuracy scores of responses in the learning phase plotted as a function of item repetitions while controlling for experimenter/computer variability. Vertical lines represent the standard error of the mean. Abbreviations: $S-$ = non-social interaction; $S+$ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

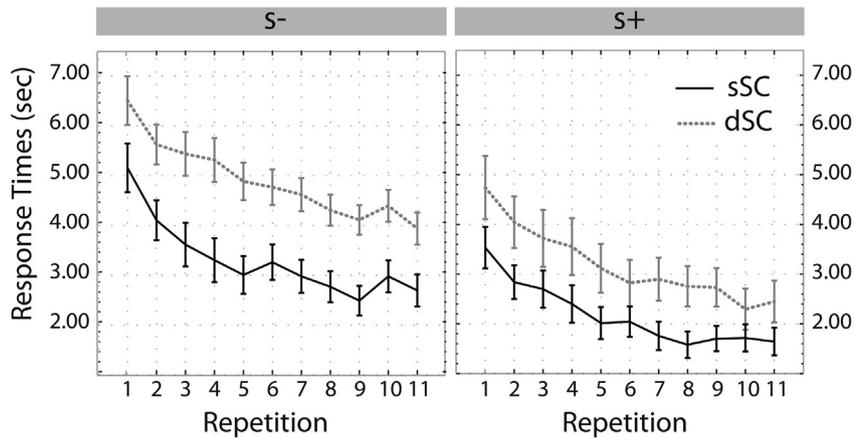


Fig. 5. Response times of correct responses plotted as a function of item repetitions while controlling for experimenter/computer variability. Vertical lines represent the standard error of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

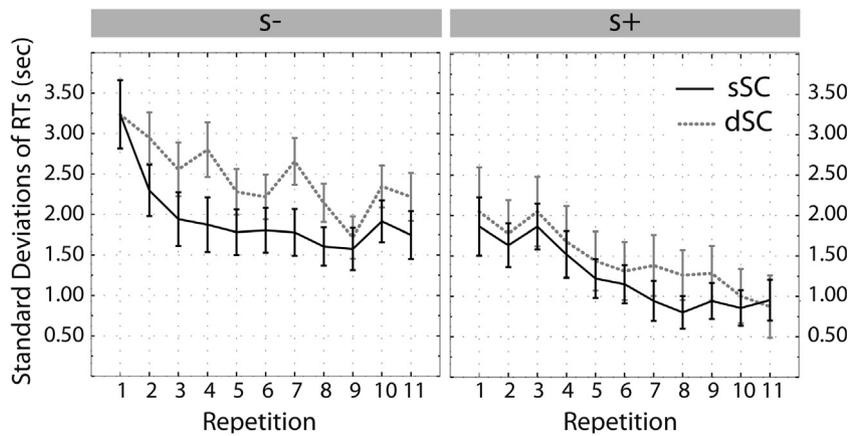


Fig. 6. Standard deviations of response times indicating participants' variability plotted as a function of item repetitions while controlling for experimenter/computer variability. Vertical lines represent the standard error of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

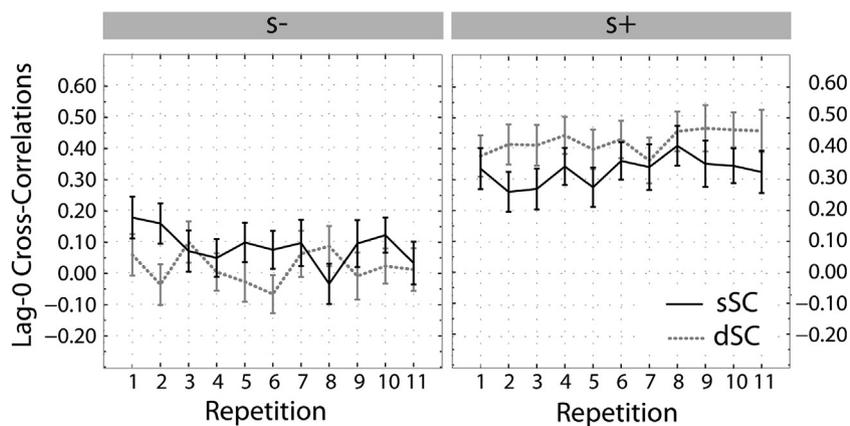


Fig. 7. Lag-0 cross-correlations coefficients are plotted as a function of item repetitions separately for non-social and social conditions (respectively: left and right panel). Vertical lines represent the standard error of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

experimental conditions nor for the interactions between the factors (all $p > 0.311$).

In summary, in the recall task, more words encoded in a variable context were correctly recalled than words encoded in the same context.

4. Discussion

The aim of the current study was to investigate adult second language word learning while playing a language learning game with a social partner. More specifically, we explored whether a

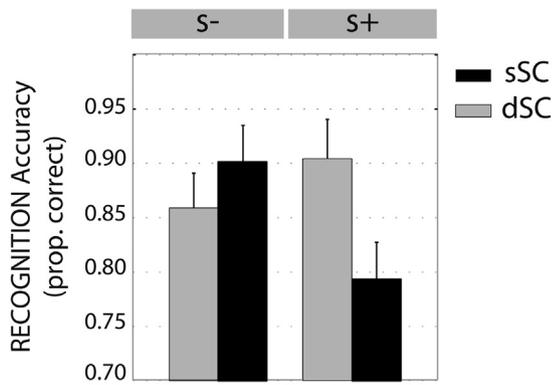


Fig. 8. Accuracy scores for the recognition task. Vertical lines represent the standard error of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

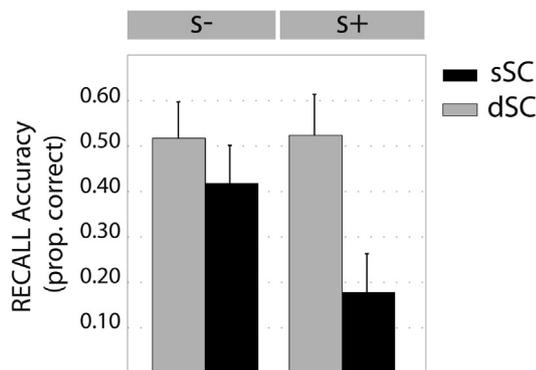


Fig. 9. Accuracy scores for the recall task. Vertical lines represent the standard error of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

social partner may facilitate the new word learning of adults, or whether contextual information may suffice. Given the paucity of research in this field, we introduced and validated a learning game based on contextual learning principles: Participants learned—either alone or with a social partner—new words embedded in sentence contexts with different degrees of variability. Results show that during the learning phase, social interaction significantly influenced the participants' behavior: Response times were faster, standard deviations of response times smaller, and temporal correlations higher when participants were learning socially compared to non-socially. Furthermore, in the social condition, temporal correlations with the partner were higher when words were presented in variable sentence contexts. In addition, variable contexts elicited slower response times compared to a repetitive context. In a subsequent recognition task, participants, who learned socially, recognized more words when they were originally embedded in a variable context. Participants who learned new words on a computer recognized more words embedded in a same context compared to participants who learned together with a social partner. Lastly, in a recall task, more words embedded in different contexts during learning were recalled than words repeated in the same context.

Learning a new language is a rather complex task; however, most often people are able to learn words of a new language relatively easily and effortlessly. How is this possible? The literature on first language acquisition emphasizes the role of the caregiver, who directs a child's attention to the correct referent of a new word thus speeding up the identification of a new word's meaning

(Csibra & Gergely, 2009; Frith & Frith, 2006; Louwerse et al., 2012; Rader & Zukow-Goldring, 2012). In this process, temporal coordination between the child and the caregiver is crucial, as the child's attention needs to be guided to the correct referent at the correct point in time (Pereira et al., 2008; Rolf et al., 2009). Adults learning new words in a social context face a problem remarkably similar to that of children: When a new word is encountered, it is necessary to understand what the word means, by identifying its referent among many possible meanings. In this process, is a social partner helpful in allocating the learner's attention to the correct referent? Results of the learning phase suggest that this is the case: Participants were significantly faster in identifying the correct referent of a new word when interacting with a social partner than with the computer. This means that participants needed less time to identify the target referent as attention was already allocated to it. This interpretation parallels findings in first language acquisition studies and suggests that social interaction between a learner and a more knowledgeable person directs the learner's attention toward correct new word referents (Pereira et al., 2008). It may be argued that this effect merely reflects "speeding up", possibly due to social pressure (or the "social facilitation" effect—Allport, 1924; Strube, 2005; Triplett, 1898); however, a simple increase in response speed would predict an increased error rate, which was not found in the learning phase. Rather, the current result is compatible with the interpretation that a social partner may orient the learner's attention toward the correct target referent, thus speeding up her/his response.

In line with this interpretation, temporal coordination was larger during social learning than during non-social learning. This result extends previous findings that partners involved in social interactions tend to coordinate their behavior to achieve common goals (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Schmidt, Fitzpatrick, Caron, & Mergeche, 2011; Stivers et al., 2009; Wilson & Wilson, 2005). Such temporal coordination is likely supported by the human ability to infer a social partner's intention during an interaction (Frith & Frith, 2012). Indeed, understanding a partner's intention is essential to predict what will happen next (Frith & Frith, 2006; Verga & Kotz, 2013) and, consequently, to adjust the time-course of one's own action (Pecenka, Engel, & Keller, 2013; Vesper, van der Wel, Knoblich, & Sebanz, 2011). However, human interactions entail much more than simply predicting a next event and adjusting to it. If this were the case, temporal coordination would have been maximal with a perfectly predictable partner. Participants performing the computer-based learning phase faced this situation; yet, participants in this learning condition showed significantly lower temporal coordination than participants dealing with a less predictable human social partner. Furthermore, participants not interacting socially during learning were also less precise in their learning behavior as indicated by larger standard deviations in their response times. Reduced variability (i.e., increased precision) in the participants' performance is a typical observation in joint action studies and is interpreted as an indication of the fine-tuning of one's own action necessary to ensure smooth coordination between partners (Repp, 2005; Vesper et al., 2011, 2012).

The results of the learning phase suggest that a social partner may significantly influence adult word learning. But do adults also use other sources of information during word learning? In line with the literature (Laufer & Hulstijn, 2001; Nagy et al., 1987; Rodríguez-Fornells et al., 2009; Swanborn & De Glopper, 1999), we hypothesized that adults may use information derived from the context a word is used in to identify its meaning. This was confirmed as participants learning alone on a computer were equally able to identify a new word's meaning. This was seen in the comparable accuracy rates during the learning phase in social and non-social learners. However, important differences emerged

between groups when considering the degree of variability in the information provided by the context: Temporal coordination between the participant and the social partner was higher when new words appeared in different sentence contexts, and lower when words were embedded and repetitively presented in the same context; however, no such differences were observed for participants interacting with the computer. A possible interpretation for this result is that if a new word is repeated in different contexts, it is not possible to apply a priori knowledge of the correct word referent. Consequently, the learner may have to rely on the partner to obtain cues as to which referent may be the correct one. This situation is remarkably similar to that of a child learning his first words: The learner is faced with a constantly changing environment (in the present case, the checkerboard), in which multiple referents are present. In such a situation, the help of the caregiver is critical for children in their effort to learn new words (Csibra & Gergely, 2009; Kuhl, 2007; Pereira et al., 2008). Similarly, adult learners may rely more on a social partner when they have to find a referent for a new word in a new learning environment. On the other hand, words repeated in the same sentence context do not require a social partner to guide the learner's attention toward the target as the target will be known from previous presentations in the same sentence context. Accordingly, in the recognition task, social learners recognized words originally encoded in a variable sentence context better than words encoded in a repeated context. Hence, the presence of a partner may, in this instance, facilitate the allocation of attention and the integration of information pivotal to successfully learn new words (Lagarde & Kelso, 2006; Rolf et al., 2009; Schmidt-Kassow et al., 2013). However, despite clear differences emerging between groups in the recognition accuracy scores as a function of sentence context, we did not find similar differences in the recall task. In addition, no differences were observed in the accuracy rates during the learning phase. This may suggest that both learning strategies (with a social partner or without) are equally effective in adult learners during word learning. Alternatively, it is possible that the learning game used in the current investigation was too easy to uncover how a social partner can facilitate learning per se. In other words, as the task was fairly easy for all participants, the use of specific learning strategies when learning with a partner may not have constituted an advantage. Third, and most informative, accuracy scores in the learning phase do not measure learning outcome per se, which is rather measured in the recognition and recall tasks of the testing phases. Indeed, in the learning phase participants are simply required to find a third picture that correctly matches a sequence of two pictures presented before. In this phase, knowing the new name of the word is of no advantage to proceed with the following learning trials. Hence, accuracy scores in the learning phase do not represent a measure of learning per se, but merely monitor the learner's ability to perform the visuo-spatial task. On the other hand, measures of coordination (reaction times, standard deviations, cross-correlations) are crucial in this phase as they represent – according to the theoretical framework we work with – the basis for successful learning.

During the learning phase, words repeated in the same sentence context elicited faster response times compared to words embedded in variable sentence contexts. This result can be explained as a contextual cueing effect: attention is guided by repetitive contextual cues; consequently, repeated sentence contexts facilitate the orienting of attention to the correct target (Chun & Jiang, 1998; Conci & Müller, 2012). However, words embedded in variable sentence contexts led to higher recall scores in the testing phase. We consider that when words appear in variable sentence contexts each time the target word is presented, the number of available cues that enrich its representation increases and strengthens the new word's meaning (Adelman et al., 2006; Verkoeijen et al.,

2004). Evidence on increased recall rates for words originally embedded in variable contexts compared to words repeated in the same context is therefore in line with this consideration.

What remains to be discussed are the differences regarding the recognition and recall tasks in the testing phase. Recognition and recall tasks hinge on different aspects of the memory process; recognition is easier and does not require a deep level of encoding, while recall requires the access and retrieval of a stored item (Craik & Lockhart, 1972; Moscovitch & Craik, 1976). It is possible that the inconsistency between the learning phase (performed with a partner) and the testing phase (non social) may have hindered retrieval in the recall task in the social learners' group: Participants who learned socially may have been disadvantaged as they experienced a contextual inconsistency between the learning phase (performed with a social partner) and the testing phase (performed alone on a computer). Consistency between learning and testing environments has been suggested to facilitate recall (Godden & Baddeley, 1975; Polyn et al., 2009). Accordingly, participants engaged in a non-social learning environment had an advantage in the testing phase, as recall during testing was similar to the learning phase. However, it is important to note that the context of language use is most often a social one: Even when learned from a textbook, language is most often used to communicate with other people. Because of this mismatch, learners may experience a conflict when using L2 learned via text, while they do not when new words are learned in a social context. For this reason, it has been suggested that learning new words in a social context should be encouraged as compared to learning via text books (Jeong et al., 2010).

Collectively, the results of the current experiment show how the presence of a social partner significantly influences the learner's behavior by directing the learner's attention toward the referent of a new word. While the notion of the caregiver as an attentional enhancer has been clearly outlined in L1 models of language acquisition (Hollich et al., 2000; Kuhl, 2007; Markman, 1990; Tomasello, 2000), the implications for L2 learners have not been explicitly postulated. The data presented here allow extending L1 learning models to account for an important characteristic of adult learners: Social interaction—even in adults—works as an attentional beacon, which can direct the learner's attention toward relevant elements in the environment; however, if the context allows the learner to identify the referent by her/himself, the presence of another person does not help. This evidence characterizes the adult learner as neither a self-sufficient entity, nor as a socially grounded one, but rather as a “utilitarian learner”. Indeed, while social presence influences adult learners, the impact of the exerted influence depends on the task at hand. In our word-learning game, the participants' task was to resolve the indeterminacy of the referent in order to learn new word meanings. In this context, the idea of the “utilitarian learner” predicts that a social partner's influence will be maximal when the referent is more difficult to find. In other words, when information provided by the sentence context is sufficient for the learner to ascertain a new word's meaning, social interaction will not be particularly influential. In the game, a difficult learning situation is presented when new words are embedded in variable sentence contexts. Contrary, new words repeated in the same sentence context represent an easier learning situation, as every time a word is encountered its referent may immediately be identified on the basis of previous encounters with that word. In this latter case, it is possible for the learner to directly map a new word onto the referent, possibly by guessing strategies based on contextual cues. However, in more difficult situations, when many referents are encountered, a social partner may facilitate the mapping of a word and a referent. In this scenario, the partner orients the learner's attention toward the correct referent and the crucial information in the learning context (Fig. 10).

Importantly, the current experiment introduced and validated an easy-to-use learning game to investigate a complex phenomenon—word learning during social interaction—even in experimental settings, in which simplicity is particularly crucial such as neuroimaging. We have previously discussed the role of the parietal cortex in social interaction. Consistent activation of the right inferior parietal lobe has reported in social neuroscience studies and may point to a general-purpose function of this area (Carter & Huettel, 2013; Decety & Lamm, 2007); that is, the right parietal cortex may serve to direct attention, for example in social interaction. According to this proposal, “changing from a non-social to a social context should increase activation within the TPJ and its functional connectivity with other task-relevant regions; (...) should the TPJ indeed be critical for social contexts, then (...) when social information becomes irrelevant for behavior, the TPJ should be disengaged even if a social agent is still present” (Carter & Huettel, 2013). Hence, in line with the idea of the adult as “utilitarian learner”, we predict that the current task would lead to the engagement of the right TPJ when learners acquire new words in different sentence context (when social information would be relevant) while engaging with a social partner, but not for words in consistent sentence contexts (when social information is irrelevant, although the partner may still be present).

Lastly, we would like to address some limitations of the current study. First and foremost, we were unable to show that social interaction affects the learning performance (accuracy scores) in the recall task, while sentence context effects emerge during the recognition task in the social group. It may be possible that the game was too easy to uncover differences between the two groups of participants. Further studies using this paradigm should try to test this hypothesis by making the game more difficult (for example, by reducing the number of repetitions for each target object).

Second, unlike other experiments investigating social interaction we did not collect video recordings of the interaction between participants and a social partner. This choice was made to obtain the most naturalistic interactive behavior as the presence of a camera—in addition to the experimenter—might have put too much pressure on the learners. Nevertheless, video recordings may reveal aspects of an interaction that the experimenter may have missed by being directly involved in the task and these aspects could have been quantified and used to refine the statistical analysis.

In addition, in the current study we used a real person—the experimenter—as the learner’s partner in the social condition. This allowed evaluating the coordinative dynamics naturally occurring in a social learning situation, in which, however, variability was controlled by the presence of the same partner (i.e., the experimenter) for all participants. Other studies have employed a different strategy, consisting in using a computer controlled virtual partner whereas participants are told they are in fact interacting with another human (see Schilbach et al., 2009; Wilms et al., 2010). This approach has the advantage of enabling more technical and controlled experimental manipulations across participants, and represents an alternative way to investigate social word learning in future studies employing the paradigm introduced here.

Lastly, in this study we focused on a possible way through which a social partner may influence a learner: By directing the learner’s attention, the social partner may reduce uncertainty. However, this may not be the only way how social interaction influences learning. Another important factor to consider would be the motivational value of a social partner: A study by Schilbach et al. (2009) demonstrated that joining attention with a partner engages reward-related brain areas, possibly reflecting increased motivation in sharing with another person. It is plausible that even in the learning game participants playing with a social partner were more motivated than participants playing alone with the computer, leading to better performance when learning. This aspect should be further investigated in future studies, for example, by asking the participants to fill in a questionnaire concerning how much they liked the experiment.

To conclude, the current study aimed to investigate second language word learning in adults during social interaction. More specifically, we explored whether a social partner may facilitate the learning of new words in adults or whether contextual information may suffice to do so. To this aim, we introduced and validated a novel word learning game. The results suggest that a social partner may help adult learners to learn new words by directing attention towards a correct new word referent. However, cues derived from the situational learning context a new word is presented in are essential to determine the learning outcome: A social partner only directs the learner’s attention when the context does not provide enough information to identify a new word’s meaning. While providing evidence on how social interaction may influence word learning, this study also opens a major venue

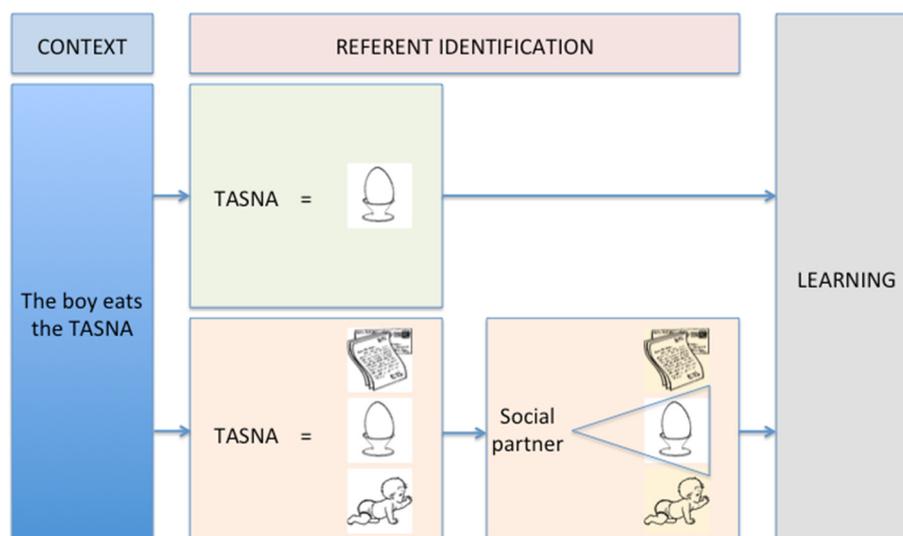


Fig. 10. The “utilitarian learner” model. When learning a new language, an adult learner refers to the context a new word is presented in. When the context provides enough information, the learner does not need help from another person to identify the word’s referent. On the contrary, a social partner is needed to disentangle the correct meaning when several possible referents are available.

for future research in this domain. In particular, the word learning game may be particularly helpful for future investigations on the role of the attentional network in social word learning.

Acknowledgments

Laura Verga received funding from the European Community's Seventh Framework Program under the EBRAMUS project—grant agreement no. 238157 granted to S.A.K. The authors would like to thank Marieke Van Der Steen for insightful comments and suggestions in the process of data analysis, and Jan Bergmann for help with the original Presentation scripts.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2017.06.018>.

References

- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, 128(3), 466–478.
- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4(02), 179–190. <http://dx.doi.org/10.1017/S136672890100027>.
- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814–823. <http://dx.doi.org/10.1111/j.1467-9280.2006.01787.x>.
- Allport, F. H. (1924). *Social psychology*. New York: Houghton Mifflin.
- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... Pléh, C. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2), 344–380.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In *Second language acquisition and the critical period hypothesis* (pp. 161–181).
- Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. David: Birdsong.
- Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, 94(2), 265.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In *Second language acquisition and the critical period hypothesis* (pp. 133–159).
- Bongaerts, T., Mennen, S., & Slik, F. van der (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, 54(2), 298–308. <http://dx.doi.org/10.1111/1467-9582.00069>.
- Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, 116(2), 289–296.
- Bruner, J. S. (1974). From communication to language—A psychological perspective. *Cognition*, 3(3), 255–287. [http://dx.doi.org/10.1016/0010-0277\(74\)90012-2](http://dx.doi.org/10.1016/0010-0277(74)90012-2).
- Bruner, J. S. (1983). *Child's talk*. New York: Norton.
- Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, 17(7), 328–336.
- Chee, M. W., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, 23(1), 127–137.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71. <http://dx.doi.org/10.1006/cogp.1998.0681>.
- Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, 9(7), 909–916. <http://dx.doi.org/10.1093/scan/nst062>.
- Conci, M., & Müller, H. J. (2012). Contextual learning of multiple target locations in visual search. *Visual Cognition*, 20(7), 746–770. <http://dx.doi.org/10.1080/13506285.2012.694376>.
- Consonni, M., Caffero, R., Marin, D., Tettamanti, M., Iadanza, A., Fabbro, F., & Perani, D. (2013). Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex*, 49(5), 1252–1258. <http://dx.doi.org/10.1016/j.cortex.2012.04.009>.
- 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. *Nature Neuroscience*, 3(3), 292–297. <http://dx.doi.org/10.1038/73009>.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [http://dx.doi.org/10.1016/S0022-5371\(72\)80001-X](http://dx.doi.org/10.1016/S0022-5371(72)80001-X).
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153.
- Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *The Neuroscientist*, 13(6), 580–593.
- Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, 141(1), 49.
- Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, 79(2), 162.
- Dominey, P. F., & Dodane, C. (2004). Indeterminacy in language acquisition: The role of child directed speech and joint attention. *Journal of Neurolinguistics*, 17(2–3), 121–145. [http://dx.doi.org/10.1016/S0911-6044\(03\)00056-3](http://dx.doi.org/10.1016/S0911-6044(03)00056-3).
- Dumas, G., Lachat, F., Martinerie, J., Nadel, J., & George, N. (2011). From social behaviour to brain synchronization: Review and perspectives in hyperscanning. *IRBM*, 32(1), 48–53.
- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS One*, 5(8), e12166. <http://dx.doi.org/10.1371/journal.pone.0012166>.
- Fischer, K., Foth, K., Röhlfing, K. J., & Wrede, B. (2011). Mindful tutors: Linguistic choice and action demonstration in speech to infants and a simulated robot. *Interaction Studies*, 12(1), 134–161.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313.
- Gergely, G., & Watson, J. S. (1999). Early socio-emotional development: Contingency perception and the social-biofeedback model. *Early Social Cognition: Understanding Others in the First Months of Life*, 60, 101–136.
- Gleitman, L., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, 11(01), 43–79. <http://dx.doi.org/10.1017/S0305000900005584>.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15(1), 1–16.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7(2), 95–112.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66(3), 325–331. <http://dx.doi.org/10.1111/j.2044-8295.1975.tb01468.x>.
- Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations. *Infancy*, 2(2), 219–231.
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4), 878–894.
- Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension: An ecological-dynamic systems view. *Developmental Science*, 4(1), 1–18.
- Green, D. W. (2003). The neural basis of the lexicon and the grammar in L2 acquisition. In *The interface between syntax and the lexicon in second language acquisition* (pp. 197–208).
- Green, D. W., Crinon, J., & Price, C. J. (2006). Convergence, degeneracy, and control. *Language Learning*, 56, 99–125. <http://dx.doi.org/10.1111/j.1467-9922.2006.00357.x>.
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2), 114–121.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664), 1634–1640.
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1160–1164. <http://dx.doi.org/10.1037/0278-7393.31.5.1160>.
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63(3), 259–273.
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, 65(3), i–135.
- Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, 13(5), 706–713. <http://dx.doi.org/10.1111/j.1467-7687.2009.00926.x>.
- 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(2), 802–809. <http://dx.doi.org/10.1016/j.neuroimage.2009.12.038>.
- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L. I., Lu, C., & Jing (2012). Neural synchronization during face-to-face communication. *The Journal of Neuroscience*, 32(45), 16064–16069. <http://dx.doi.org/10.1523/JNEUROSCI.2926-12.2012>.

- Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports*, 3(1692). <http://dx.doi.org/10.1038/srep01692>.
- Koffka, K. (1935). *Principles of Gestalt psychology* (Harcourt: Brace and Company). New York: Routledge.
- Kotz, S. A., D'Ausilio, A., Raettig, T., Begliomini, C., Craighero, L., Fabbri-Destro, M., ... Fadiga, L. (2010). Lexicality drives audio-motor transformations in Broca's area. *Brain and Language*, 112(1), 3–11. <http://dx.doi.org/10.1016/j.bandl.2009.07.008>.
- Krall, S. C., Volz, L. J., Oberwilling, E., Grefkes, C., Fink, G. R., & Konrad, K. (2015). The right temporoparietal junction in attention and social interaction: A transcranial magnetic stimulation study. *Human Brain Mapping*. <http://dx.doi.org/10.1002/hbm.23068>.
- Kuhl, P. K. (2007). Is speech learning "gated" by the social brain? *Developmental Science*, 10(1), 110–120.
- 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. *Proceedings of the National Academy of Sciences*, 100(15), 9096–9101. <http://dx.doi.org/10.1073/pnas.1532872100>.
- Lagarde, J., & Kelso, J. (2006). Binding of movement, sound and touch: Multimodal coordination dynamics. *Experimental Brain Research*, 173(4), 673–688.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
- Lenneberg, E. H., Chomsky, N., & Marx, O. (1967). *Biological foundations of language* (Vol. 68) New York: Wiley.
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, 64(3), 249–255. <http://dx.doi.org/10.1016/j.jml.2010.11.003>.
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404–1426. <http://dx.doi.org/10.1111/j.1551-6709.2012.01269.x>.
- Marinovic, W., Cheung, F., Riek, S., & Tresilian, J. R. (2014). The effect of attention on the release of anticipatory timing actions. *Behavioral Neuroscience*. <http://dx.doi.org/10.1037/bne0000007>.
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1), 57–77.
- Markova, G., & Legerstee, M. (2006). Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology*, 42(1), 132–141. <http://dx.doi.org/10.1037/0012-1649.42.1.132>.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4), 831–877. <http://dx.doi.org/10.1037/a0029872>.
- Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, 15(4), 447–458. [http://dx.doi.org/10.1016/S0022-5371\(76\)90040-2](http://dx.doi.org/10.1016/S0022-5371(76)90040-2).
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237–270.
- Newman-Norlund, R. D., Noordzij, M. L., Meulenbroek, R., & Bekkering, H. (2007). Exploring the brain basis of joint action: Co-ordination of actions, goals and intentions. *Social Neuroscience*, 2(1), 48–65.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, 3(2), 178–192.
- Pecenka, N., Engel, A., & Keller, P. E. (2013). Neural correlates of auditory temporal predictions during sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7.
- Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ, US: Princeton University Press.
- Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, 19(3), 170–182. <http://dx.doi.org/10.1002/hbm.10110>.
- Perani, D., Paulesu, E., Galle, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., ... Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, 121(10), 1841–1852.
- Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: Interacting systems of perception and action. *Connection Science*, 20(2–3), 73–89. <http://dx.doi.org/10.1080/09540090802091891>.
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally. *Psychological Science*, 21(12), 1894–1902. <http://dx.doi.org/10.1177/0956797610389189>.
- Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, 27(02), 212–225.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language & Computation*, 4(2), 203–228.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129.
- Rączaszek-Leonardi, J., Dębska, A., & Sochanowicz, A. (2014). Pooling the ground: Understanding and coordination in collective sense making. *Frontiers in Psychology*, 5. <http://dx.doi.org/10.3389/fpsyg.2014.01233>.
- Rader, N. deV., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences*, 34(5), 559–568. <http://dx.doi.org/10.1016/j.langsci.2012.03.011>.
- Reddy, V., & Uithol, S. (2015). Engagement: Looking beyond the mirror to understand action understanding. *British Journal of Developmental Psychology*. Retrieved from <<http://onlinelibrary.wiley.com/doi/10.1111/bjdp.12106/pdf>>.
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6), 969–992. <http://dx.doi.org/10.3758/BF03206433>.
- 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. *Human Movement Science*, 26(6), 867–891.
- Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3711–3735.
- Rolf, M., Hanheide, M., & Rohlfing, K. J. (2009). Attention via synchrony: Making use of multimodal cues in social learning. *IEEE Transactions on Autonomous Mental Development*, 1(1), 55–67.
- Sage, K. D., & Baldwin, D. (2010). Social gating and pedagogy: Mechanisms for learning and implications for robotics. *Neural Networks*, 23(8–9), 1091–1098. <http://dx.doi.org/10.1016/j.neunet.2010.09.004>.
- Schilbach, L. (2014). On the relationship of online and offline social cognition. *Frontiers in Human Neuroscience*, 8, 278. <http://dx.doi.org/10.3389/fnhum.2014.00278>.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Voegeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(04), 393–414.
- Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., ... Voegeley, K. (2009). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, 22(12), 2702–2715. <http://dx.doi.org/10.1162/jocn.2009.21401>.
- Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, 30(5), 834–845. <http://dx.doi.org/10.1016/j.humov.2010.05.014>.
- Schmidt-Kassow, M., Heinemann, L. V., Abel, C., & Kaiser, J. (2013). Auditory-motor synchronization facilitates attention allocation. *NeuroImage*, 82, 101–106.
- Sebastiani, V., de Pasquale, F., Costantini, M., Mantini, D., Pizzella, V., Romani, G. L., & Della Penna, S. (2014). Being an agent or an observer: Different spectral dynamics revealed by MEG. *NeuroImage*. <http://dx.doi.org/10.1016/j.neuroimage.2014.08.031> (in press).
- Smith, L. B. (2000). Learning how to learn words: An associative crane. In *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80).
- Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 43(2), 549–565. <http://dx.doi.org/10.2307/1127555>.
- Stein, B. S. (1978). Depth of processing reexamined: The effects of the precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior*, 17(2), 165–174. [http://dx.doi.org/10.1016/S0022-5371\(78\)90128-7](http://dx.doi.org/10.1016/S0022-5371(78)90128-7).
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430. <http://dx.doi.org/10.1073/pnas.1008662107>.
- Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 760–766. <http://dx.doi.org/10.1037/0278-7393.29.5.760>.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Yoon, K. E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592.
- Striano, T., & Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Developmental Science*, 8(4), 333–343. <http://dx.doi.org/10.1111/j.1467-7687.2005.00421.x>.
- Strube, M. J. (2005). What did Triplet really find? A contemporary analysis of the first experiment in social psychology. *The American Journal of Psychology*, 271–286.
- Swanborn, M. S. L., & De Groot, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261–285.
- Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods*, 35(4), 621–633.
- Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... Bates, E. (2005). Timed action and object naming. *Cortex*, 41(1), 7–26.
- Szafarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, 4. <http://dx.doi.org/10.1038/srep05304>.
- Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, 10(4).
- Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, 10(2), 201–224.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121–125.
- Triplet, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, 9(4), 507–533. <http://dx.doi.org/10.2307/1412188>.
- Tulving, E. (1979). Relation between encoding specificity and levels of processing. *Levels of Processing in Human Memory*, 405–428.

- Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind & Language*, 25(1), 3–29. <http://dx.doi.org/10.1111/j.1468-0017.2009.01379.x>.
- Verga, L., Bigand, E., & Kotz, S. A. (2015). Play along: Effects of music and social interaction on word learning. *Frontiers in Psychology*, 6. <http://dx.doi.org/10.3389/fpsyg.2015.01316>.
- Verga, L., & Kotz, S. A. (2013). How relevant is social interaction in second language learning? *Frontiers in Human Neuroscience*, 7(550). <http://dx.doi.org/10.3389/fnhum.2013.00550>.
- Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 796.
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3), 517–530.
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2012). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Experimental Brain Research*, 211, 517–530.
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, 13(6), 258–263.
- Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes—Using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, nsq024. <http://dx.doi.org/10.1093/scan/nsq024>.
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12(6), 957–968.
- Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(6), 545.
- Zajonc, F. (1965). Social facilitation. *Science*, 149(3681), 269–274.