Engaging Nonverbal Theory-of-Mind Boosts Novel Word Retention in Adults

Katherine M. Trice (trice.k@northeastern.edu)
Department of Psychology, 360 Huntington Ave. Forsyth Building Rm 133
Boston, MA 02115 USA

Emily Cohen (cohen.emil@northeastern.edu)
Department of Behavioral Neuroscience, 360 Huntington Ave. Forsyth Building Rm 133
Boston, MA 02115 USA

Zhenghan Qi (z.qi@northeastern.edu)
Department of Communication Sciences and Disorders; Department of Psychology, 360 Huntington Ave. Forsyth Building Rm 133
Boston, MA 02115 USA

Abstract

Previous research has argued that theory-of-mind (ToM) plays a critical role to motivate early language acquisition. Vocabulary learning is a lifelong journey from infancy to adulthood. As ToM and vocabulary grows in tandem through development, little is known whether ToM continues to be important for word learning in adults, and if so, whether there is a causal interplay between the ToM state and adults' word learning outcomes. This study tests whether engaging nonverbal ToM assists novel word encoding and retention. We examined young adults (N = 51)' word learning outcomes for meanings that were either introduced in a word-to-world direct mapping context or a pragmatically inferred context. Preceding each block of word learning for each context was a brief, non-verbal animation that either primed the ToM system or did not engage ToM. We found that initial meaning mapping via pragmatic inference was assisted specifically by priming ToM prior to word learning. Furthermore, word retention, tested 20 minutes after initial encoding, was strengthened for both pragmatically inferred and directly mapped words, when learning was preceded with a ToM video, as opposed to a non-ToM video. These results demonstrate that ToM causally interacts with word learning processes and facilitates both encoding and retention. Our findings strengthen previous arguments that ToM plays a critical role in language development and broaden this to encompass lifelong vocabulary learning.

Keywords: theory of mind; word learning; pragmatic inference; cognitive priming; memory

Introduction

Theory of Mind and Word Learning

For decades, researchers have argued that social cognition, and within this, theory of mind (ToM) – the ability to intuit what another person believes, intends, or feels – serve as a critical gatekeeper for word learning (Tomasello, 2003; Kuhl, 2011). Here, researchers such as Kuhl posits that social interactions can 1) increase arousal – and through this attention, motivation and encoding – 2) often provide coordinated information and cues that enhance learning – such as eye gaze and joint attention – 3) have an interlocuter who is perceived to be like the self and is thus more easily recognized as being an intentional agent, and 4) make use of predispositions towards aspects such as human faces, interactions, and patterns of action (Kuhl, 2011). These factors can in turn support stronger and more accurate mapping and retention of novel words. And indeed, various studies have found that such aspects of social interaction do indeed support varying aspects of language development, including in vocabulary (Çetinçelik, Rowland, & Snijders, 2021; Kuhl, Tsao, & Liu, 2003; Tanaka, Cicourel, & Movellan, 2007). Furthermore, the ToM system provides necessary underpinnings for comprehending social interactions and inferring speaker's intent, making it a necessary component for sound-meaning mapping (Brosseau-Liard, Penney, & Poulin-Dubois, 2015; Gollek & Doherty, 2016; Jaswal, 2004; Sabbagh & Baldwin, 2001; Saylor & Carroll, 2008).

Our understanding of the role ToM plays in word learning is largely shaped by developmental work. However, from lectures to job training, in-group slang to new technologies, vocabulary acquisition occurs across the lifespan as vocabulary size continues to expand (Kavé, 2024). Furthermore, within the realm of language processing, adults can make incredibly rapid and sophisticated judgements of speaker meaning and intent rooted in their ToM abilities to map the form present to the abstract, intended meaning. Research has associated individual differences in ToM abilities and adults' success in pragmatic processing (Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2014; Fairchild & Papafragou, 2021; Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012). And we further know that social interaction, as opposed to solo learning situations, seems to support language learning, though whether this stems from ToM has not been studied (Verga & Kotz, 2017). ToM should thus also guide how adults learn new words in incredibly sophisticated ways. However, current theories regarding the role of ToM fail to consider whether this skill is a gatekeeper during the language development of young

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

children or whether the ToM is necessarily and consistently involved during vocabulary acquisition across the lifespan (Richardson, Lisandrelli, Riobueno-Naylor, & Saxe, 2018).

Thus, work that probes both degree of causal relation between ToM and word learning and whether any such relation exists in adults with fully mature ToM systems is critical to understanding the fundamental mechanism of vocabulary acquisition.

Cognitive Priming, Word Learning, and Theory of Mind

One way to behaviorally investigate whether there is a cause-and-effect interplay between ToM and vocabulary acquisition is through a priming paradigm.

The linguistic sphere of knowledge has been shown to be extremely sensitive to priming effects across multiple domains of linguistic representation (ex: auditory semantic and phonological lexical representations: Arias-Trejo, Angulo-Chavira, Avila-Varela, Chua-Rodriguez, & Mani, 2022; phonological and orthographic representations: Chen, Zhang, He, Wei, Zou, Li, & Zhao, 2023; cognitive control and language processing/error monitoring and correction: Hsu & Novick, 2016; semantic and orthographic lexical network size representations: Yap, Tse, & Balota, 2009).

However, within word learning, its application is most frequently restricted to implicitly testing the degree of integration into memory, with priming success during retention as a measure of linguistic depth of encoding (Batterink, & Neville, 2011; Rabovsky, Sommer, & Abdel Rahman, 2012). Studies that depart from this, such as Fritz, Schütte, Steixner, Contier, Obrig, & Villringer, (2019), most often make use of simultaneous rather than sequential priming, for instance accompanying novel word exposure with semantically congruent or incongruent music. This study in particular demonstrated that associated, evoked meaning across domains – linguistic vs non-linguistic – supports learning success, indicating that word learning remains malleable to other cognitive mechanisms.

If ToM plays a causal role in facilitating adult word learning, then priming ToM system should result in a differential learning and memory profile. A prior study, Saratsli, Trice, Papafragou, & Qi (2023), found not only correlation between ToM and word learning outcomes in adults, but also reported altered word memories when learning took place after an intensive, linguistic ToM task. These findings provide preliminary evidence for the intrinsic role that ToM plays, at least in a subset of word learning situations, even into adulthood, indicating that not only is it a relevant skill but that it is malleable on a state level with that state affecting subsequent learning.

Pragmatic Inferences, Word Learning, and Theory of Mind

One critical finding of Saratsli et al. (2023) is that ToM seems to be only necessary for word learning that took place in scenarios that require pragmatic inference, but not the context of direct mapping. This context-specific requirement of ToM may not be too surprising. The natural world is full of competing referents, and thus it is often the case that the intended referent of a given speaker is ambiguous. In such scenarios, it may be necessary to rely on one's internal mapping of the speaker – their knowledge, beliefs, and intentions – to infer the meaning of a novel word (Akhtar, Carpenter, & Tomasello, 1996; Nadig & Sedivy, 2002; Nurmsoo & Bloom, 2008). This is what is referred to as a pragmatic inference, and it stands in contrast to the unambiguous one-to-one mapping that requires no such presuppositions about ones interlocuter that direct mapping conditions serve (Halberda, 2006). Thus, it is no surprise that pragmatics and pragmatic inferences have been correlated with and mediated by ToM skills both within behavioral and neuroimaging literature (Matthews, Biney, & Abbot-Smith, 2018; Fairchild & Papafragou, 2021; Spotorno et al., 2012).

However, it is critical to note that the lack of modulation of direct mapping retention offers an interesting counterpoint to the hypothesis placing ToM as a fundamental mechanism of word learning. While both word mapping scenarios contains social interchange – a speaker informing the listener of objects she likes or wants, which the listener was then asked to select – this alone was not enough for state changes in ToM to have an effect. Thus, it may be the case that only more complex social interactions or word learning scenarios in which ToM is critical for mapping the word itself rely on the ToM mechanism as a gateway to greater learning success.

And indeed, the primary findings of Saratsli et al. (2023) appear to offer preliminary support for the role of the ToM system specifically as a significant force for adult vocabulary acquisition in the form of successful word retention. Here, it was found that words learned via pragmatic inference – and thus related to ToM – were remembered significantly better than those directly mapped. This dovetails well with both emerging literature demonstrating that successful mapping of words alone is not enough to support encoding and memory long-term (Axelsson, Churchley, & Horst, 2012; Horst & Samuelson, 2008). Instead, it is the presence of other mechanisms at play during the learning experience that govern successful retention (Vlach & Sandhofer, 2012). ToM engagement may thus be one such factor at play. However, the past research did not manipulate ToM state within subjects. As a result, the dynamic causal interplay between ToM and meaning acquisition process is poorly understood. In addition, the linguistic nature of the ToM task prevented us from teasing apart the contribution of language system from that of the ToM system.

Current Study

In the current study, we expand our understanding on the interplay between the ToM system and word learning in adults, focusing specifically on the role of ToM system activation in immediate mapping and retention of words that rely on complex ToM-based inferences as well as those that can be acquired through direct mapping. To do so, the above word learning contexts will each be primed in turn by stimuli designed to either evoke or not evoke ToM. Here, we predict that the boosting effect of complex ToM-based inferences over direct mapping during mapping on retention seen in previous literature will be replicated, and that ToM priming will result in an increase in both initial mapping and retention accuracy for these inferences. If boosting effects of ToM activation are restricted to scenarios where they play a critical role in mapping itself, then we may only observe the boosting effects on memory in pragmatic inference resolution. If ToM facilitates meaning consolidation in general, regardless of its interplay with word mapping mechanisms, a boosting effect would be seen for both word learning contexts.

Here, the stimuli for word mapping and retention in each learning context have been reported in Saratsli et al. (2023). We use non-linguistic movie stimuli that allows for real-time ToM engagement (or lack of engagement) on a trial-by-trial basis. This addresses two critical gaps in previous findings. First, the confounding effect of using a linguistic-based ToM task in the past literature means that it remains unclear whether the link seen stems from ToM itself, from pre-activation of the interplay between the ToM and language system, or from primarily the language component. Second, interweaving primed and non-primed blocks of each condition allows us to more fully capture the impact of real-time ToM engagement in word learning within subject.

Methods

Participants

Fifty-one undergraduates (18-33, M_{age}=20.14, 15 male) completed this experiment. All participants were monolingual native speakers of English with no exposure to a second language before the age of 5. All were compensated with course credits.

Stimuli and Procedure

The experiment was administered asynchronously through the online experimental platform Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Participants completed a word learning experiment, with learning assessed at three timepoints: in-the-moment mapping, immediate recall, and retention. This consisted of an introduction/practice phase, a word learning task consisting of 4 blocks each followed by an immediate recall task, and a retention task beginning ~ 20 minutes after leaning start. This ensures that learned words have already shifted from short to long-term memory consolidation (Radvansky, Doolen, Pettijohn, & Ritchey, 2022). Two individual-difference tasks were administered between word learning and retention phases (see **Individual Difference Measures**). These tasks maintain a consistent interval between between learning and retention across participants.

Word Learning Task The study began with an introduction where participants were immersed in a toy store setting and administered two practice trials in which they listened to a cartoon tour guide that directed them to choose a toy holding an object that the tour guide preferred. The introduction ended with a short introductory video clip from the Pixar Short *Partly Cloudy* that set the background context to understand subsequent video clips.

Next, participants completed the word learning task which consisted of four randomized blocks. **Each block** was either in the ToM video priming (ToM) or non-ToM video priming (Non-ToM) *condition*, and thus the subsequent block of trials was preceded by a video that was known to either activate or not activate the ToM system. Each block was either in the pragmatic inference (PI; Fig 1A) or direct mapping (DM; Fig 1B) learning *context*. For each learning context, four novel words were taught twice for a total of 8 trials per block. The blocks thus consisted of: ToM with PI, ToM with DM, Non-ToM with PI, non-ToM with DM (Fig 2).

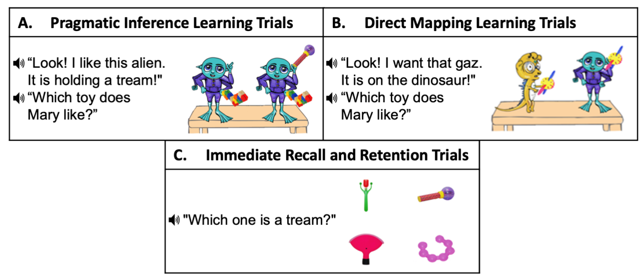

Figure 1: Word learning paradigm with example trials in each learning context with given visual and auditory stimuli.

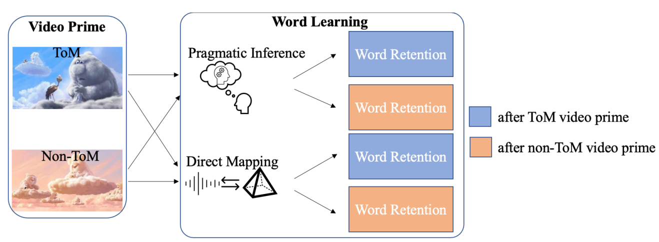

Figure 2: ToM priming paradigm. Note that all participants completed all prime-learning combinations.

Immediately after each block, the participants completed a four-alterative-forced-choice-task to test the immediate recall of the novel words learned in the preceding block. The same task was used to later test retention of each novel word (Fig 1C).

Video primes were 30 seconds in length and each contained a full story sequence. Critically, all videos were fully non-verbal, with no dialogue present at any point. Each was segmented from *Partly Cloudy* based on Richardson et al. (2018)'s analysis based on whether the clip reliably activated canonical ToM brain regions for more than half of the length and occurred at least in the second half of the clip (ToM) or never activated canonical ToM regions during any portion of the clip (Non-ToM). Valence was controlled between the ToM and the Non-ToM videos. Note that in all videos, social interaction between characters occurred. The

ToM clips lead viewers to spontaneously mentalize the characters' thoughts, emotions, and intent. The Non-ToM videos have clips either with characters experiencing pain for negative valence or through viewing pleasant scenes for positive valence. Social interaction is still present, but the interplay is not the focus of the plot. Thus, any difference in effect should stem from evocation of underlying cognitive mechanism of ToM rather than being induced by the observation of a social scene. All movie-watching is passive, with the social interaction and mental states of the characters being designed to be inferred even by young children without any explicit tasks; thus, any boosting effect seen by ToM is unlikely to result purely from cognitive task-demand differences in the priming stimuli.

Trials in the PI context showed two identical aliens or two identical dinosaurs, in which one held both the target unique novel object and the competitor novel object while the other toy held only the same competitor (Fig. 1A). For each trial, participants heard a phrase such as (1) below. The final sentence fails to resolve the ambiguity, as on purely a lexical and semantic level the novel word could refer to either the target or competitor. Instead, the listener must make a pragmatic inference based on internal computations of speaker assumptions of intent, on the principle that the speaker seeks to be informative and disambiguate between the two identical figures and that they believe they have given the listener enough information to do so (Grice, 1975). Thus, here, the novel word would be mapped to the disambiguating and thus unique novel object that is the target.

1. Pragmatic Inference: "Look! I like this alien. It is holding a tream!"

Trials in the DM context showed one alien and one dinosaur, each holding the same novel object. For each trial, participants heard a phrase such as (2) below (Fig 1B). The final sentence allows participants to directly map the novel word onto the sole novel object, thus removing the necessity to make an effortful and uncertain computation as to the speaker's intent to map the novel label to the novel object.

2. Direct Mapping: "Look! I want that gaz. It's on the dinosaur!"

Note in both these cases that the sentences contained the same amount of social interactive component – the participant was told that the character likes or wants a toy, and then selected the desired object. This, like in the video task, allows for consistence in the social interaction aspect across context while still modulating necessity of ToM for the mapping of novel words themselves.

Individual Difference Measures Both individual difference measure tasks occurred between the final learning block and the retention task in counterbalanced order.

ToM Task This task was an adapted web-based Mind-in-the-Eyes task (created by Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Each trial presented participants with a picture of different set of eyes set above four different words, each describing a feeling or mental state. From these choices, they chose the words that best described what the person pictured in the photo was thinking or feeling.

Filler Task A Flanker task from the *Gorilla* platform was used as a Filler task in this study to test participants' inhibitory control, i.e., their ability to suppress responses that are inappropriate in a particular context. In the task there were 5 fish that either all looked towards the same direction (congruent trial) or all the but the middle fish looked in the same direction and the middle in the opposite (incongruent trial). Participants had to choose a button that correlated with the direction that the middle fish was looking towards, disregarding the directions of the surrounding fish.

Results

Effect of Priming Condition on Learning Context

Because of the difference in task demand during learning – select the novel object with the clear one-to-one mapping during DM vs making a complex mapping inference with continued uncertainty during PI – contexts were analyzed separately using generalized linear mixed effects models via the lme4 package (Baayen, 2008; Baayen, Davidson, & Bates, 2008) in R (R Core Team, 2021). The models had binomial trial-by-trial mapping accuracy as the dependent variable. A fixed effect of priming condition – ToM vs Non-ToM prime – was used for both contexts. An additional fixed effect of ToM skills measured by the mean accuracy of the Mind-in-the-Eyes ToM Task was included for the PI context to examine if relationships between ToM skill and overall learning outcome seen in prior work was found. Contrasts of 0.5 vs -0.5 were used for priming such that the mean across the different levels for the given effect formed the baseline. Random intercepts for items and participants were included. Full results for all models results can be found in Table 1.

For the PI context, priming with ToM videos resulted in significantly greater accuracy of in-the-moment mapping than priming with non-ToM videos (M_{ToM}=0.89, $M_{Non\text{-}ToM}$=0.74; β=3.03, *SE*=0.97, z=3.14, *p*=0.002, Fig. 3). Greater ToM skill, based on the ToM individual difference measure, predicted more accurate mapping for the PI context (β=6.37, *SE*=2.15, z=2.96, *p*=0.003).

For the DM context, priming with ToM resulted only a marginal effect and accuracy was at ceiling (M_{ToM}=0.99, $M_{Non\text{-}ToM}$=1; *p*=0.10). This is expected, as there is no ambiguity present in the item to select.

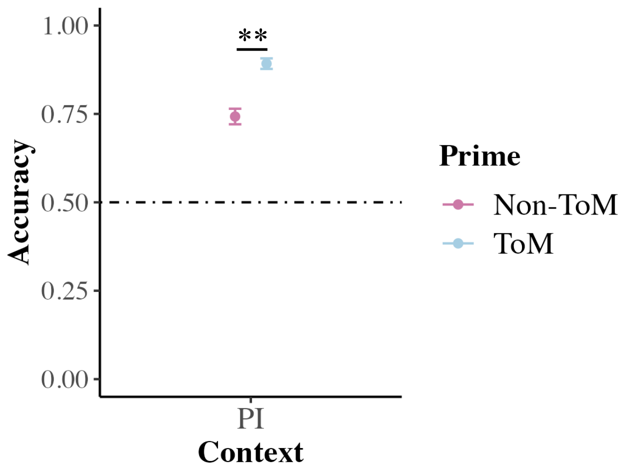

Figure 3: In-the-moment mapping accuracy for PI based on priming condition. Dashed line represents chance, error bars within-subject standard error. Significance levels: . p < 0.1, * p < .05, ** p < .01, ***p < .001.

Table 1: Parameter estimates for learning accuracy scores

Effects	Estimate	SE	Z
Pragmatic Inference			
Intercept	-1.34	1.50	-0.89
Condition (ToM vs Non-ToM)	3.03	0.97	3.14 **
ToM Skill	6.37	2.15	2.95 **
Direct Mapping			
Intercept	57.03	7.60	7.51 ***
Condition (ToM vs Non-ToM)	-20.32	12.21	-1.66 .

Note: Significance levels: . p < 0.1, * p < .05, ** p < .01, ***p < .001.

Effect of Priming Condition on Recall & Retention

Separate generalized linear mixed effects models were also constructed to analyze immediate recall and retention. Here, as in learning, the models had binomial trial-by-trial mapping accuracy as the dependent variable. However, contexts were not split into separate models. Instead, a fixed effect of context – PI vs DM – and of condition – ToM vs Non-Tom prime – as well as their interaction was used. ToM task accuracy formed an additional co-variate. Contrasts 0.5 vs -0.5 were used for both fixed effects. Random intercepts for items and participant and by-subject random intercepts and slopes for context type were included. Full results for all models results can be found in Table 2.

There was no significant difference in immediate recall between PI and DM contexts, ToM and Non-ToM primes, or the interaction between the two (DM: M_{ToM}=0.80, $M_{Non\text{-}ToM}$=0.78; PI: M_{ToM}=0.89, $M_{Non\text{-}ToM}$=0.81; *p's*>0.10). However, like during word learning, a participant's ToM ability had a significant effect on immediate recall (β=6.23, *SE*=1.60, z=3.91, *p*<0.001).

For retention, there were significant effects of both context and condition such that words mapped during PI were better retained than words mapped during DM. Words primed by ToM videos were remembered with greater precision than those primed by non-ToM videos (DM: M_{ToM}=0.70, $M_{Non\text{-}ToM}$=0.64; PI: M_{ToM}=0.78, $M_{Non\text{-}ToM}$=0.72; Context: β=0.55, *SE*=0.21, z=2.65, *p*=0.008; Condition: β=0.37, *SE*=0.17, z=2.12, *p*=0.03; Fig. 4). However, no interaction between the Prime and Context was found (*p*>0.10). Participants' ToM ability once again had a significant effect on accuracy (β=4.44, *SE*=1.70, z=2.61, *p*=0.009).

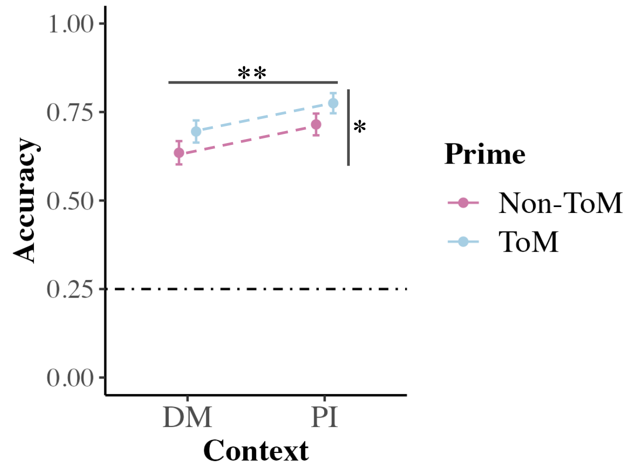

Figure 4: Retention accuracy for DM and PI based on priming condition. Dashed line represents chance, error bars within-subject standard error. Significance levels: . p < 0.1, * p < .05, ** p < .01, ***p < .001.

Table 2: Parameter estimates for immediate recall and retention accuracy scores

Effects	Estimate	SE	Z
Immediate Recall			
Intercept	-2.15	1.06	-2.02 *
Context (PI vs DM)	0.24	0.45	0.54
Condition (ToM vs Non-ToM)	0.52	0.39	1.35
Context * Condition	0.75	0.78	0.97
ToM Skill	6.23	1.60	3.91 ***
Retention			
Intercept	-1.86	1.16	-1.61
Context (PI vs DM)	0.55	0.21	2.65 **
Condition (ToM vs Non-ToM)	0.37	0.17	2.12 *
Context * Condition	0.07	0.35	0.19
ToM Skill	4.44	1.70	2.62 **

Note: Significance levels: . p < 0.1, * p < .05, ** p < .01, ***p < .001.

Conclusions

Here, we replicated prior work in adults demonstrating that stronger ToM abilities, as indexed by a language-based ToM task, predict better retention of words mapped via ToM. Furthermore, we extended these results to in-the-moment mapping, indicating that theory of mind abilities – or, at minimum, the strength of association between linguistic and ToM knowledge – play a critical role in resolving pragmatic inferences.

More critically, however, we have found that *non-verbal* activation of the ToM system prior to learning not only has a boosting effect on accurate mapping of words via ToM but results in significantly stronger long-term retention of words, regardless of whether ToM played a heavy role in their initial mapping. The prime used here is naturalistic and specific to the ToM domain: activation of ToM is based on a combination of inferences for the characters in the video, stemming from expressions, actions, and contextual cues without a language component. As such, this illuminates the role that domain-general ToM may play in word learning in daily life.

Furthermore, by controlling the degree of social salience across both primes and word-learning contexts, confounding effects of the social but not ToM context are accounted for. Thus, the impact of ToM on word mapping is shown to stretch beyond the social element of the learning scenario. However, this does not mean that any causal nature of ToM indicated by the priming effect is direct. While it is possible that the ToM system may be tied to memory such that greater activation results in deeper encoding or consolidation, it may also be the case that greater ToM activation necessarily results in downstream effects such as greater arousal, which may in turn be the instrumental factor in better memory.

This distinction is furthered by our selected primes. We know from prior work that the movie clips used as our ToM primes reliably activate ToM-related brain regions (Richardson et al, 2018). However, due to the lack of behavioral measures for ToM before and after the prime videos, it is possible that increased arousal and motivation rather than boosted ToM led to memory benefits. Future work that monitors the physical systems at play, both neurological and physiological, and that tests the effects of the prime on ToM function, would be necessary to tease these possibilities apart. Additionally, implicated by the larger effect sizes at later than earlier retention stages in our data, more work with sophisticated memory measures would address how the priming memory effect manifests over the course of memory consolidation.

Overall, we demonstrated a dynamic cause-and-effect role of ToM in word mapping via pragmatic inference. We also showed a significant positive relationship between ToM and memory consolidation of newly acquired meanings even when ToM doesn't play a necessary role in the initial mapping itself. This in turn provides targeted empirical evidence that strengthens previous arguments that social cognition plays a critical role in language acquisition. Furthermore, our work suggests that social cognition can be a fundamental support for word learning beyond early language development.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The Role of Discourse Novelty in Early Word Learning. *Child Development*, *67*(2), 635. https://doi.org/10.2307/1131837

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arias-Trejo, N., Angulo-Chavira, A. Q., Avila-Varela, D. S., Chua-Rodriguez, F., & Mani, N. (2022). Developmental changes in phonological and semantic priming effects in Spanish-speaking toddlers. *Developmental psychology*, *58*(2), 236–251.

Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The Right Thing at the Right Time: Why Ostensive Naming Facilitates Word Learning. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00088

Baayen, R. H. (2008). Analyzing Linguistic Data. *Cambridge University Press.* https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 42*(2), 241–251.

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning. *Cerebral Cortex, 24*(10), 2572–2578. https://doi.org/10.1093/cercor/bht112

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *Journal of cognitive neuroscience*, *23*(11), 3181–3196. https://doi.org/10.1162/jocn_a_00013

Brosseau-Liard, P., Penney, D., & Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *The British journal of developmental psychology*, *33*(4), 464–475. https://doi.org/10.1111/bjdp.12107

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development. *Frontiers in psychology*, *11*, 589096. https://doi.org/10.3389/fpsyg.2020.589096

Chen, Y., Zhang, C., He, W., Wei, S., Zou, K., Li, X., & Zhao, L. (2023). The phonological congruency modulated long-term form priming of Chinese characters. Memory & cognition, 10.3758/s13421-023-01462-y. Advance online publication. https://doi.org/10.3758/s13421-023-01462-y

Fritz, T. H., Schütte, F., Steixner, A., Contier, O., Obrig, H., & Villringer, A. (2019). Musical meaning modulates word acquisition. *Brain and language, 190,* 10–15. https://doi.org/10.1016/j.bandl.2018.12.001

Gollek, C., & Doherty, M. J. (2016). Metacognitive developments in word learning: Mutual exclusivity and theory of mind. *Journal of Experimental Child Psychology, 148,* 51–69. https://doi.org/10.1016/j.jecp.2016.03.007

Grice, H. P. (1975). Logic and conversation. In R. Cole & J. Morgan (Eds.), Syntax and semantics: speech acts. *Academic Press.*

Jaswal, V. K. (2004). Don't believe everything you hear: preschoolers' sensitivity to speaker intent in category induction. *Child development*, *75*(6), 1871–1885. https://doi.org/10.1111/j.1467-8624.2004.00822.x

Kavé, G. (2024). Vocabulary changes in adulthood: Main findings and methodological considerations. *International journal of language & communication disorders*, *59*(1), 58–67. https://doi.org/10.1111/1460-6984.12820

Kuhl, P. K. (2011). Social mechanisms in early language acquisition: Understanding integrated brain systems supporting language. In J. Decety & J. Cacioppo (Eds.), *The Oxford handbook of social neuroscience*. Oxford UK: Oxford University Press.

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 of the United States of America*, *100*(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

Fairchild, S., & Papafragou, A. (2021). The Role of Executive Function and Theory of Mind in Pragmatic Computations. *Cognitive science*, *45*(2), e12938. https://doi.org/10.1111/cogs.12938

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology, 53(*4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Hsu, N. S., & Novick, J. M. (2016). Dynamic Engagement of Cognitive Control Modulates Recovery From Misinterpretation During Real-Time Language Processing. *Psychological science*, *27*(4), 572–582. https://doi.org/10.1177/0956797615625223

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual Differences in Children's Pragmatic Ability: A Review of Associations with Formal Language, Social Cognition, and Executive Functions. *Language Learning and Development, 14*(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of Perspective-Taking Constraints in Children's On-Line Reference Resolution*. Psychological Science, 13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' Perspective Taking in Word Learning. *Psychological Science, 19(*3), 211–215. https://doi.org/10.1111/j.1467-9280.2008.02069.x

R Core Team (2021). R: A language and environment for statistical computing*. R Foundation for Statistical Computing.*

Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). Implicit word learning benefits from semantic richness: electrophysiological and behavioral evidence. *Journal of experimental psychology. Learning, memory, and cognition*, *38*(4), 1076–1083. https://doi.org/10.1037/a0025646

Radvansky, G. A., Doolen, A. C., Pettijohn, K. A., & Ritchey, M. (2022). A new look at memory retention and forgetting. *Journal of experimental psychology. Learning, memory, and cognition*, *48*(11), 1698–1723. https://doi.org/10.1037/xlm0001110

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature communications*, *9*(1), 1027. https://doi.org/10.1038/s41467-018-03399-2

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*(4), 1054–1070. https://doi.org/10.1111/1467-8624.00334

Saratsli, D., Trice, K. M., Papafragou, A., & Qi, Z. (2023, September 20). Pragmatics and social cognition in learning and remembering words. https://doi.org/10.31234/osf.io/gv4an

Saylor, M. M., & Carroll, C. B. (2009). Direct and indirect cues to knowledge states during word learning. *Journal of child language*, *36*(3), 685–695. https://doi.org/10.1017/S0305000908009136

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage, 63*(1), 25–39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17954–17958. https://doi.org/10.1073/pnas.0707769104

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

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

Vlach, H. A., & Sandhofer, C. M. (2012). Fast Mapping Across Time: Memory Processes Support Children's Retention of Learned Words. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00046

Yap, M. J., Tse, C. S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and

word frequency: The role of lexical integrity. *Journal of memory and language*, *61*(3), 303.
https://doi.org/10.1016/j.jml.2009.07.001