Working memory for object concepts relies on both linguistic and simulation information

Dymarska, Agata

Connell, Louise

Banks, Briony

Working memory for object concepts relies on both linguistic and simulation information

Agata Dymarska (a.dymarska@lancaster.ac.uk)
Louise Connell (l.connell@lancaster.ac.uk)
Briony Banks (b.banks@lancaster.ac.uk
Department of Psychology, Fylde College,
Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK

Abstract

The linguistic-simulation approach to cognition predicts that language can enable more efficient conceptual processing than sensorimotor-affective simulations of concepts. We proposed that this has implications for working memory, whereby use of linguistic labels enables more efficient representation of concepts in a limited-capacity store than representation via full sensorimotor simulation. In two pre-registered experiments, we asked participants to remember sequences of real-world objects, and used articulatory suppression to selectively block access to linguistic information, which we predicted would impair accuracy and latency of performance in an object memory recognition task. We found that blocking access to language at encoding impaired memory performance, but blocking access at retrieval unexpectedly facilitated speed of responding. These results suggest that working memory for object concepts normally relies on language but people can flexibly adapt their memory strategies when language is unavailable. Moreover, our data suggest that a sequence of up to 10 object concepts can be held in working memory when relying on sensorimotor information alone, but this capacity increases when linguistic labels are available.

Keywords: working memory; concepts; linguistic information; simulation; embodied cognition

Introduction

Although traditionally conceptual representations were considered amodal and removed from perceptual experience (Tranel, Damasio & Damasio, 1997), more recent evidence suggest that concepts are grounded in sensorimotor and linguistic experience (Barsalou et al., 2008; Connell & Lynott, 2014; Vigliocco et al., 2009). Simulated representations engage the neural subsystems involved in sensorimotor, affective, introspective, and other situated experiences of a concept. For example, the concept *"dog"* includes its visual shape and colour, the action and feel of patting its fur, the sound of its bark, walking it on a leash, and the positive feelings towards a pet. The neural activation patterns involved in processing these experiences can later be partially re-activated (i.e., *simulated*) to represent a concept. Linguistic representations, on the other hand, comprise word (and phrase) labels associated with these sensorimotor-affective simulations, and the distributional patterns between them (statistical co-occurrences of words in language). For instance, seeing a terrier or hearing a bark will activate the label "dog", and words that frequently appear in similar contexts, like "tail" or "leash". These two components are interrelated and mutually supportive, and recent theories argue that both are intrinsic to conceptual representation (Connell & Lynott, 2014; Louwerse, 2011). That is, linguistic labels are part of concepts and conceptual processing uses simulation and linguistic information to varying extents depending on task demands, available resources, and other factors (Connell, 2018; Connell & Lynott, 2014).

The role of simulation and linguistic components in cognition is illustrated by a range of empirical evidence. Neuroimaging research has shown that processing of action words (e.g. "pick", "kick") activates body part-specific motor areas (Hauk, Johnsrude, & Pulvermüller, 2004). Critically, processing of such words is selectively impaired in patients with neurodegeneration of the motor system – Parkinson's disease (Boulenger et al., 2008). Behavioural experiments also show evidence for use of simulations: for example, people were faster to recognise a horizontally-oriented nail after reading "He pounded the nail into the wall" than "He pounded the nail into the floor" (Stanfield & Zwaan, 2001). Participants were also quicker to make a size judgment of manipulable objects than when the objects were too big to be physically manipulated (Connell, Lynott, & Dreyer, 2012). As for the linguistic component, information from language alone is powerful enough to inform responses across diverse conceptual tasks. Evidence comes from a range of paradigms, including property verification and generation (Louwerse & Connell, 2011; Santos et al., 2011), spatial iconicity judgements (Louwerse & Jeuniaux, 2010) and spatial cuing of attention (Goodhew, McGaw, & Kidd, 2014). Frequency of words co-occurring in the same context can predict how easily they are understood as a novel conceptual combination (Connell & Lynott, 2013). These findings show that both sensorimotor and linguistic information is functionally important to conceptual processing.

Much evidence for the linguistic component centres on the usefulness of distributional information (i.e., co-occurrence relationships between words/phrases) in cognition. However, that is not its full role. Language is a unique human characteristic which allows us to communicate something in the past, future, or hypothetical existence (Barsalou, 2005), and allows us to concisely name a complex multimodal experience. The idea that language is beneficial for our cognitive processing has been around for a while (e.g.: Paivio, 1971), but recent theories have developed the role of linguistic labels in a number of new directions (e.g., Borghi et al., 2018; Connell, 2018; Lupyan, 2012). Most relevant to our present purposes, Connell and Lynott (2014) propose that having labels for concepts enables a process of *linguistic bootstrapping*, whereby words and phrases act as *linguistic placeholders* in an ongoing representation when there are insufficient resources to maintain a sensorimotor simulation in full. These linguistic placeholders can later be fleshed out into a simulation again at any time if resources become available. To date, the linguistic bootstrapping hypothesis has remained theoretical and has not been tested directly but there is indirect support for the idea in the wider literature. Working memory (WM) is necessarily limited in capacity – there are only so many concepts that can be maintained and manipulated at once – and recent evidence does suggest that linguistic information is more economical in representation (i.e., occupies less "space" in working memory) than sensory information (Langerock, Vergauwe, & Barrouillet, 2014). Further, explicitly labelling simple visual stimuli seems to increase memory capacity (Zormpa et al., 2018). It is possible that when working memory capacity is strained to its limit, as when trying to maintain a representation of numerous concepts, a linguistic label could deputise for its referent sensorimotor information (e.g., word "dog" replaces simulation of *dog*) to free up space.

It is currently unknown how many concepts (i.e., representations of real-world objects, events, and situations, such as *dog*, *running*, or *holiday*) can be maintained in working memory at once. Research on memory from the linguistic-simulation perspective concentrated on the role of sensorimotor simulation in memory (Dutriaux, Dahiez, & Gyselinck, 2018; Vermeulen et al., 2013) rather than the interplay of simulated and linguistic information in capacity limits. Working memory research has established a central capacity limit of 4 items (Cowan, 2010), but research informing this has used simple, artificial stimuli (e.g., feature conjunctions such as *red triangle*; random word pairs such as *desk-ball*). Such stimuli do not generalise to naturalistic, real-world concepts that comprise rich sensorimotor and linguistic information from long-term memory, and that are typically represented in broader situated simulations where concepts to reinforce and cue one another (e.g., a *dog* that is *running* with a *ball*). Baddeley's (2000) episodic buffer, a finite-capacity buffer that allows information from long-term memory to be integrated and manipulated goes some way to address these issues. For instance, participants remember sequences of words better when they are presented in meaningful sentences (i.e., that exploit interconnections between words) than in unstructured lists, which Baddeley and colleagues attribute to long-term knowledge retrieved to support representations in the episodic buffer. Nonetheless, not much is known about the role of simulated and linguistic information in representing concepts in working memory.

The current study

Our aim was to examine the role of linguistic and simulation components in working memory for real-world object concepts. In two pre-registered experiments, we presented ecologically valid sequences of object pictures (e.g., ingredients for a novel recipe) in a nonverbal paradigm, and tested recognition memory by asking participants to select the previously-presented objects from arrays of distractors. Critically, participants performed articulatory suppression (repeating aloud "the") during item encoding and/or retrieval to block access to linguistic information. Articulatory suppression has been widely used in WM research (Baddeley, 1992), where it has been shown to interfere with verbal encoding but to have little effect on general processing in the central executive (e.g., De Rammelaere, Stuyven, & Vandierendonck, 2001; Jaroslawska et al., 2018). We used a no-suppression condition as a baseline instead of an alternative suppression task, such as spatial tapping or visual interference, because such tasks would have interfered with the sensorimotor representation of concepts and therefore could not provide a useful control in our experiment. Thus, we expected the articulatory suppression task to block participants' access to the linguistic component of their conceptual representations.

We hypothesised that storage of object concepts in working memory will normally rely on language (i.e., linguistic placeholders), and that speed and accuracy of performance will be impaired when access to language is blocked. We expected performance to be best with no articulatory suppression at either encoding or retrieval, when participants are free to use both linguistic and sensorimotor information. We expected performance to be worst with articulatory suppression at retrieval only, when participants can employ linguistic placeholders to encode more objects, but lose access to those objects at retrieval when access to linguistic information is blocked. We planned to estimate working memory capacity for sensorimotor representations of concepts by calculating the average number of objects correctly retrieved with articulatory suppression at both encoding and retrieval (i.e., when linguistic information was unavailable).

Experiment 1

In this first experiment (pre-registration, data, analysis code, and full results are available as supplemental materials at https://osf.io/acv3m/?view_only=c1799106289a4063abf2eaa490eae009), we presented six objects per sequence, based on estimates from Langerock et al. (2014) that only 2-3 complex representations can be maintained in the episodic buffer. Participants viewed images of objects in each sequence one at a time during the encoding stage, and then had to select an alternative image of each target object from an array of distractors during the retrieval stage. Articulatory suppression took place half the time at encoding (to block access to object labels and prevent the use of linguistic placeholders when storing objects in WM) and half the time at retrieval (to block access to linguistic placeholders stored in WM).

Method

Participants Thirty-two native speakers of English (27 females; mean age = 21.2 years, *SD* = 3.2 years) were recruited from Lancaster University, for which they received course credit or a payment of £3.50. One participant was replaced due to a procedural error during testing.

The sample size was determined using sequential hypothesis testing with Bayes Factors (Schönbrodt et al., 2017). We stopped at the minimum sample size $N = 32$ when the Step 3 models for accuracy and Response Time (RT) cleared the specified threshold of evidence $BF_{10} < 0.2$. (model details in Design & Analysis section, full statistics in the Results section).

Materials Test items comprised of 72 target objects, divided into 12 sequences, each designed to be an ecologically valid order of objects that would be plausibly used in a real-world setting, such as ingredients used in the process of making a cake, or a set of tools used in order to hang a picture. Each sequence therefore consisted of 6 target objects for study during the encoding stage, and each target object was assigned five distractor items for display in an object array during the retrieval (testing) stage. Five distractor objects were selected from the same category (e.g., food items, clothing) of which three were chosen based on colour, shape or function of the target object. Target and distractor items could all be plausibly used in a particular sequence (e.g. recipe) so that the task maintained ecological validity, and it would not be obvious which item in the array was the correct one.

To ensure the order of objects within each sequence was ecologically valid, we asked 9 naïve participants (who did not take part in the experiment) to rank-order the items according to how they are used in a given context. We used mean rank per object to finalise each sequence. For example: in the scenario *Tools for hanging a picture on the wall*, participants decided on the following order: *spirit level, drill, screw plug, screw, screwdriver, picture frame.*

We sourced photographic images for all objects from license-free online resources and edited them to appear on a uniform transparent background. To ensure that participants were tested on memory for object concepts, and not perceptual matching of a specific image, we prepared two different images for each target object: one for study during encoding and one for display in the distractor array during retrieval (e.g., showing an object from a different perspective or in a different colour). Images were scaled to 840 pixels along the longest dimension for objects presented during the encoding stage, and 470 pixels along the longest dimension for objects (targets and distractors) presented in the object arrays during retrieval. This resulted in a total of 504 object images: 72 target objects presented at encoding, 72 target objects presented at retrieval, and 360 distractor objects presented at retrieval. Figure 1 shows sample stimuli.

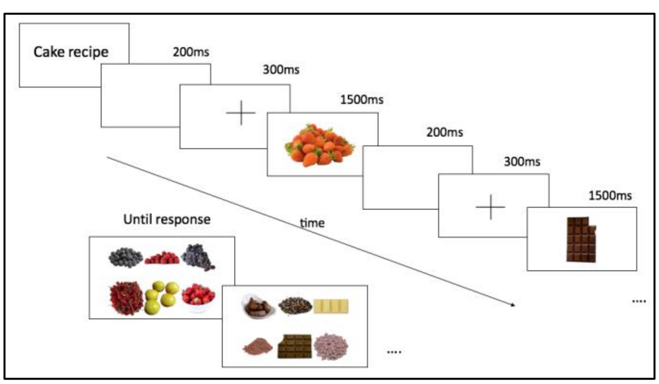

Figure 1: Diagram showing trial sequence and example stimuli at encoding and retrieval stages.

Procedure Participants were tested individually. After signing the consent form, the experimenter explained and demonstrated articulatory suppression, asking the participant to practice it. Once the participant confirmed that they understood and could perform articulatory suppression correctly, they sat in front of a computer, provided demographic information and proceeded to instructions. Participants were told they would see a sequence of everyday objects appear one by one onscreen, and their task was to remember the objects; later, they would see groups of objects onscreen and they should click on the object that belongs to the sequence they had been asked to remember. Participants then commenced a practice sequence without any articulatory suppression at encoding or retrieval. When the participant confirmed that they understood the task and were happy to continue, they were instructed on the articulatory suppression condition for both encoding and retrieval (i.e. when to start and stop) and commenced the experimental trials. Articulatory suppression was manipulated between participants at encoding and within participants at retrieval. The order of retrieval conditions was counterbalanced, and six sequences were presented in a randomised order within each condition. Experiment presentation was controlled by PsychoPy software (version 1.84.1; Peirce, 2009).

In the encoding stage, target objects were presented individually in a sequence (see Figure 1 for display times). Each sequence was preceded by a label (e.g. "cake recipe"). Once a full sequence of six target objects had been presented, participants saw a "wait" screen of 3 asterisks ("***") for 10 seconds before the retrieval stage began. In the articulatory suppression condition at encoding, participants repeated "the" aloud until this wait screen timed out. In the retrieval stage, participants saw a 2×3 array of six objects (one target object and five distractors) in random locations within the array (see Figure 1). Response times were measured from the onset of the array display until the participant clicked on an object using the mouse. After the retrieval of all six target objects had been tested, a message on the screen asked participants to press space when they were ready to proceed to the next object sequence.

After completing encoding and retrieval of six sequences, participants were instructed to take a self-paced break. They were then asked to switch to the counterbalanced articulatory suppression condition at retrieval (articulatory suppression at encoding remained constant). Participants then completed encoding and retrieval for six further object sequences.

Design and Analysis We analysed accuracy (dummy coded: incorrect = 0, correct = 1) with a mixed-effects hierarchical logistic regression (binomial, logit link). Step 1 entered participants and items as crossed random effects, where items were defined as objects nested within sequences. Step 2 added encoding and retrieval as fixed effects (no articulatory suppression = 0, articulatory suppression = 1). Step 3 added the interaction of encoding and retrieval as a fixed effect. We ran Bayesian model comparisons between steps, with Bayes Factors (BF) calculated via Bayesian Information Criteria (Wagenmakers, 2007). Similarly, we analysed RT for correct responses in a mixed-effects linear regression with the same effects and model comparisons as above. All analyses were run in R software (lme4 package, R version 3.4.1, 2017).

Results and Discussion

No trials were excluded on the basis of accuracy results[1]. For analysis of correct RTs, one trial was excluded as a motor error (faster than 300ms), and 27 trials were removed as outliers more than 3 standard deviations from the individual participant's mean (total 0.015% data removed).

Accuracy Bayesian model comparison showed equivocal evidence for Step 2 over Step 1, $BF_{10} = 1.58$; the data *very* weakly favoured the model containing articulatory suppression as fixed effects at encoding and retrieval over a model containing only random effects. There was strong evidence at Step 3 *against* the effect of the encoding-retrieval interaction on accuracy, $BF_{10} = 0.02$: the data were 47 times more likely under the Step 2 model without the interaction than the Step 3 model with the interaction.

We used the coefficients in Step 3 model to estimate the marginal accuracy for each condition of encoding × retrieval articulatory suppression (see Table 1). As predicted, accuracy was best in the no-suppression/no-suppression condition (no articulatory suppression at encoding or retrieval), with participants correctly recognising 5.6 out of 6 objects per sequence on average. However, accuracy was worst in the suppression/suppression condition: object memory was least accurate when language access was blocked at both encoding and retrieval (5.0 objects per sequence).

Finally, in an exploratory analysis not specified in the pre-registration, we examined the individual coefficients in the most likely model of fixed effects (i.e., Step 2)[2]. Articulatory suppression at encoding had a negative effect on response accuracy, beta = -0.567, $SE = 0.275$, $z = -2.06$, $p = .039$, as did articulatory suppression at retrieval, beta = -0.436, $SE = 0.121$, $z = -3.59$, $p < .001$. That is, as we predicted, removing access to language impaired object memory accuracy. When access to labels was blocked at the point of *encoding* objects, people were 76% more likely to make an error when later asked to recognise the object. Independently, when access to labels was blocked at the point of *retrieving* objects, people were 55% more likely to make an error. However, the inconsistency between equivocal Bayesian model comparison and significant regression parameters for Step 2 suggests that these effects should be treated cautiously.

Response Times Model comparisons showed very strong evidence at Step 2 for the effects of articulatory suppression at encoding and retrieval, $BF_{10} = 1808.04$. However, there was strong evidence at Step 3 *against* the effect of the encoding-retrieval interaction on RT, $BF_{10} = 0.03$: the data were 33 times more likely under the Step 2 model without the interaction than the Step 3 model with the interaction.

We took the coefficients in the Step 3 model to estimate the marginal mean RT for each condition of encoding × retrieval articulatory suppression (see Table 1). Against our expectations, recognition of target objects was best (fastest) in the no-suppression/suppression condition (language available at encoding but not at retrieval), and worst (slowest) in the suppression/no-suppression condition (language available at retrieval but not at encoding). People had most difficulty recognising the objects when language was blocked at the point of encoding but was available at retrieval.

We report an exploratory analysis of the coefficients in the most likely model of fixed effects (i.e., Step 2). As expected, articulatory suppression at encoding increased RT, beta = 408.57, $SE = 172.32$, $t(31.99) = 2.371$, $p = .024$. However, articulatory suppression at retrieval unexpectedly *reduced* RT, beta = -219.57, $SE = 43.85$, $t(1770.74) = -5.007$, $p < .001$). Closer examination of RT and accuracy suggested that this pattern was due to a speed-accuracy trade-off rather than facilitation of memory. When participants were asked to perform articulatory suppression at retrieval, response times were faster, but this was accompanied by lower accuracy, relative to no-suppression conditions (see Table 1). We discuss possible reasons for this trade-off below.

Summary Overall, the results support the hypothesis that memory for object concepts normally relies on language. Blocking language access when encoding a real-world object sequence affected memory: speed and accuracy were both impaired relative to no suppression. This is consistent with the idea that object concept is stored in WM via sensorimotor simulation *and* its linguistic label, and memory is impaired when only sensorimotor simulation is available.

However, blocking access to language while retrieving objects had unexpected effects. Rather than straightforward impairment, there was a speed-accuracy trade-off (low RT and accuracy), suggesting that articulatory suppression at retrieval caused participants to adopt an alternative, heuristic strategy that led to fast but inaccurate responding. Thus, the hypothesis that memory performance would be worst in the no-suppression/suppression condition was not supported. This may be because participants knew, before they studied the object sequence, whether they would perform articulatory suppression at retrieval. Knowing that language would be unavailable during retrieval could have led participants to strategically rely on sensorimotor information even when they had language access at encoding. Another possibility is that performance was subject to ceiling effects. When language was not available, participants correctly recognised 5.0 items per sequence on average, indicating that they were able to represent five object concepts in working memory from sensorimotor simulation alone (more than the suggested episodic buffer capacity of 2-3 items, Langerock et al., 2014). Thus, working memory capacity may not have been under particular strain, and participants did not have to replace some of the sensorimotor information with linguistic placeholders to remember the full sequence. We examine these possibilities in the next experiment.

[1] Exclusion criteria detailed in pre-registration

[2] All coefficients for all models available in supplemental materials

Table 1: Marginal accuracy (%) from logistic mixed effect regression, and marginal mean RT (ms, with standard errors in parentheses) from linear mixed effect regression, for each articulatory suppression condition in Experiments 1 and 2.

	Retrieval			
	No suppression		Suppression	
Encoding	%	RT (SE)	%	RT (SE)
Experiment 1				
No suppression	92.9	2499 (137)	89.6	2288 (138)
Suppression	88.3	2916 (138)	82.7	2687 (139)
Experiment 2				
No suppression	92.0	2786 (144)	90.2	2635 (144)
Suppression	83.4	2854 (141)	82.7	2675 (141)

Experiment 2

In our second experiment (pre-registration, data, analysis code, and full results are available as supplemental materials at https://osf.io/acv3m/?view_only=c1799106289a4063abf2eaa490eae009), we presented 12 objects per sequence rather than 6, and made methodological improvements to the design. Our hypotheses remained the same.

Method

Participants As in Experiment 1, we used Bayesian sequential hypothesis testing to determine sample size. Bayes Factors for Step 3 cleared the evidence threshold for the null at Nmin = 32 for both RT (BF_{10} = 0.02) and accuracy BF_{10} = 0.03). However, sequential analysis plots for the Step 2 model (the best-fitting model in Experiment 1) suggested that the level of evidence was still unstable for RT (i.e., BFs fluctuating between evidence for the null and the alternative, and equivocal evidence). We tested additional participants until it stabilised at $N = 44$. We therefore report results for 44 participants (33 female; mean age = 20.3 years, SD = 5.4). All other BF inferences and parameter estimates were consistent between $N = 32$ and $N = 44$ (full data and statistics in supplementals). Three participants were replaced due to failure to follow instructions.

Materials and procedure We used materials and procedure from Experiment 1 with the following methodological changes: to reduce the risk of ceiling effects, we paired sequences from Experiment 1, which resulted in six lists of 12 items each. Instead of a label for each list, participants were given brief information about the context (e.g.: "*You are making dinner and need to remember your shopping list for a meal and tea. Press space to proceed to the list of ingredients.*"), to provide a real-life, ecologically valid situation.

To prevent participants' knowledge of the articulatory suppression condition from affecting their encoding strategies, we altered the presentation of retrieval conditions. Instead of verbal instructions on articulatory suppression at the start of the experiment, participants saw an image of a mouth on the screen indicating the start of articulatory suppression, and the same image crossed out to indicate no articulatory suppression, before the encoding and retrieval stages on each trial. We then randomised the order of lists across retrieval conditions, so that participants did not know whether the trial involved articulatory suppression until encoding was complete. We also altered some of the distractors (N = 8; 0.015% of all items) to ensure that the target items were not easy to guess without relying on memory.

The experimental design remained the same (i.e., articulatory suppression manipulated between participants at encoding and within participants at retrieval).

We changed the "wait" screen to reduce the possibility of participants relying on perceptual matching (instead of memory for object concepts). Rather than passively looking at the screen, participants had to click on 4 dots appearing in random points on the screen to "calibrate the mouse". Additionally, object presentation time during encoding was prolonged to 2000ms, to give participants more time to encode the concept.

Results and Discussion

No trials were excluded on the basis of accuracy results. For RT analysis of correct responses, 31 trials (0.012% of data) were removed as being more than 3 SDs above the individual participant's mean.

Accuracy Bayesian model comparison showed strong evidence *against* Step 2 over Step 1, BF_{10} = 0.017; the data were 57 times more likely under the Step 1 model containing only random effects than a model containing articulatory suppression as fixed effects at encoding and retrieval. There was also strong evidence at Step 3 *against* the effect of the encoding-retrieval interaction on accuracy, BF_{10} = 0.025: the data were 40 times more likely under the Step 2 model with no interaction than the Step 3 model with the interaction.

We then used the coefficients in the Step 3 model to estimate the marginal accuracy for each encoding × retrieval articulatory suppression condition (see Table 1). As in Experiment 1, accuracy was best in the no-suppression/no-suppression condition (no articulatory suppression in either encoding or retrieval), with participants correctly recognising 11.0 out of 12 objects per sequence on average. However, accuracy was worst in the suppression/suppression condition (9.9 objects remembered). Object memory was least accurate when access to language was blocked at both encoding and retrieval, in line with Experiment 1 but not our predictions.

Although the BFs showed evidence against both models, we ran an exploratory analysis of the individual coefficients in the Step 2 model to make a comparison with Experiment 1. Articulatory suppression at encoding had a negative effect on response accuracy, beta = -0.736, SE = 0.273, z = -2.69, p = .007. As predicted, and replicating Experiment 1, removing access to language impaired object memory accuracy: when access to labels was blocked at encoding, people were 109% more likely to make an error when later attempting to recognise the object. Unlike Experiment 1, articulatory suppression at retrieval had little effect, beta = -0.117, SE = 0.100, z = -1.17, p = .243, decreasing the probability of a correct response by only 12%. However, the NHST effect of articulatory suppression at encoding was not consistent with the Bayesian model comparison at Step 2 (which added encoding and retrieval effects simultaneously), and so should be treated cautiously.

Reaction Time Bayesian model comparison showed equivocal evidence for Step 2 over Step 1, $BF_{10} = 0.61$. As in Experiment 1, there was strong evidence at Step 3 *against* the effect of the encoding-retrieval interaction, $BF_{10} = 0.02$.

We used the coefficients in the Step 3 model to estimate the marginal mean RT for each condition of encoding × retrieval articulatory suppression (see Table 1). Against our predictions, but in line with Experiment 1, recognition was best (fastest) in the no-suppression/suppression condition (language available at encoding but not retrieval), and worst (slowest) in the suppression/no-suppression condition (language available at retrieval but not encoding). People had most difficulty remembering objects when language was blocked at the point of encoding but was available at retrieval.

We report an exploratory analysis of the most likely model of fixed effects (Step 2). Articulatory suppression at encoding had no effect on speed of recognition, unlike Experiment 1, beta = 54.22, SE = 149.60, $t(43.68) = 0.36$, $p = .719$. Against our expectations but in line with Experiment 1, articulatory suppression at retrieval *reduced* RT, beta = -164.91, SE = 43.26, $t(2430.48) = -3.81$, $p < .001$. People were *faster* to recognise target objects if language was blocked at retrieval. Closer examination of RT and accuracy suggested that this may be due to a speed-accuracy trade-off, as in Experiment 1. Faster RTs due to articulatory suppression at retrieval were accompanied by a trend towards poorer accuracy, relative to no-suppression conditions (see Table 1).

Summary The results were broadly in line with Experiment 1 and support the hypothesis that memory for object concepts typically relies on language. Blocking language access while encoding a real-world object sequence impaired performance in terms of accuracy (but not latency), in line with the idea that an object concept is typically stored in WM via sensorimotor simulation *and* its linguistic label, and when only sensorimotor simulation is available, memory is adversely affected.

However, blocking language access while retrieving objects from working memory resulted in faster speed of responding that was not completely eliminated by methodological changes. These results suggest that participants adopted a heuristic strategy for responding while performing articulatory suppression at retrieval. For instance, participants may have adopted a satisficing approach to selecting the target object, based on a rapid assessment of superficial sensorimotor similarity between a concept in WM and the objects in the array, to compensate for lack of access to linguistic information. Alternatively, perhaps the articulatory suppression task itself made participants want to get through the task faster, which resulted in emphasis on speed at the cost of accuracy. We plan to follow up these possibilities in future work.

General Discussion

The study is the first to take a linguistic-sensorimotor approach to working memory. We used real-world object sequences to account for the complex nature of naturalistic concepts that can draw upon information in long-term memory, and an articulatory suppression task to investigate the role of linguistic labels in working memory for such objects. We found that blocking language access at encoding impairs memory performance (poorer speed and accuracy in Experiment 1; poorer accuracy in Experiment 2), whereas blocking access to language during retrieval leads to an apparent speed-accuracy trade-off (faster speed and poorer accuracy in Experiment 1; faster speed in Experiment 2).

The results support the sensorimotor-linguistic theories of conceptual processing that argue the importance of language in conceptual representation (Connell, 2018; Louwerse, 2011), and the linguistic bootstrapping hypothesis that proposes word labels act as placeholders in mental representations when resources are insufficient to maintain a full simulation (Connell & Lynott, 2014). When language is available, people encode objects in WM with linguistic labels and sensorimotor simulations, and when storage is at capacity, the linguistic placeholder can free up resources by allowing to drop a simulation. Experiment 2 results suggest that linguistic bootstrapping allows people to remember one extra concept in a sequence of 12 (11 rather than 10).

We expected memory performance to be impaired the most when participants could use linguistic bootstrapping at encoding but had no access to object labels at retrieval (no-suppression/suppression condition). This effect did not appear. Instead, when language access was blocked at retrieval, participants adopted an alternative, heuristic strategy to compensate for it, which resulted in a trade-off between speed and accuracy. Participants may have relied on an incomplete sensorimotor simulation in working memory (e.g., only the shape or the colour of the target object), which allowed them to respond quickly, but not always correctly.

The results highlight the importance of language in working memory performance for real-world object sequences. We plan to explore encoding and retrieval processes in more detail by testing complex stimuli in sequences of varying lengths and under time constraints.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 682848.

References

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, and A. C. Graesser (Eds), *Symbols, Embodiment, and Meaning.* Oxford: Oxford University Press.

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2018). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*. Advance Online Publication

Connell, L. (2018). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *0*(0), 1–11. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin and Review*, *20*(3), 542–550. https://doi.org/10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. https://doi.org/10.1111/tops.12097

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033321

Dutriaux, L., Dahiez, X., & Gyselinck, V. (2018). How to change your memory of an object with a posture and a verb. *Quarterly Journal of Experimental Psychology*, 174702181878509. https://doi.org/10.1177/1747021818785096

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Langerock, N., Vergauwe, E., & Barrouillet, P. (2014). The maintenance of cross-domain associations in the episodic buffer. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(4), 1096–1109. https://doi.org/10.1037/a0035783

Louwerse, M., & Connell, L. (2011). A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words. *Cognitive Science*, *35*(2), 381–398. https://doi.org/10.1111/j.1551-6709.2010.01157.x

Louwerse, M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. *Symbols and Embodiment: Debates on Meaning and Cognition*, 309–326. https://doi.org/10.1093/acprof:oso/9780199217274.003.0015

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002

Vermeulen, N., Chang, B., Mermillod, M., Pleyers, G., & Corneille, O. (2013). Memory for words representing modal concepts. *Experimental psychology*, *60*(4), 293-301.

Stanfield, R. A. & Zwaan, R. A. (2001). The Effect of Implied Orientation Derived from Verbal Context on Picture Recognition. *Psychological Science*, *12*(2), 153–156.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, *35*(10), 1319-1327.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(02), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values, *Psychonomic bulletin & review*, *14*(5), 779–804.

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2018). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 1–13. https://doi.org/10.1080/09658211.2018.1510966