UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Dymarska, A., Connell, L., & Banks, B. (2022). Linguistic Bootstrapping Allows More Real-world Object Concepts to Be Held in Mind. *Collabra: Psychology*, *8*(1). https://doi.org/10.1525/collabra.40171

Cognitive Psychology

Linguistic Bootstrapping Allows More Real-world Object Concepts to Be Held in Mind

Agata Dymarska[1] [a], Louise Connell[1,2], Briony Banks[1]

[1] Department of Psychology, Lancaster University, Lancaster, UK, [2] Department of Psychology, Maynooth University, Maynooth, Ireland

Keywords: concepts, sensorimotor simulation, mental representation, linguistic labels, linguistic bootstrapping

https://doi.org/10.1525/collabra.40171

Collabra: Psychology

Vol. 8, Issue 1, 2022

The linguistic-simulation approach to cognition predicts that language can enable more efficient conceptual processing than purely sensorimotor-affective simulations of concepts. We tested the implications of this approach in memory for sequences of real-world objects, where use of linguistic labels (i.e., words and phrases) could enable more efficient representation of object concepts than representation via full sensorimotor simulation; a proposal called *linguistic bootstrapping*. In three pre-registered experiments using a nonverbal paradigm, we asked participants to remember sequences of contextually-situated, real-world objects (e.g., the ingredients for a recipe), and later asked them to select the correct objects from arrays of distractors. Critically, we used articulatory suppression to selectively suppress implicit activation of linguistic labels, which we predicted would impair performance by reducing the number of objects that could be held in mind simultaneously. We found that suppressing access to language when learning the sequences impaired accuracy of object recognition, though not latency, and that this impairment was not simply dual-task load. Results show that a sequence of up to 10 contextually-situated object concepts can be held in mind when language is inhibited, but this increases to 12 objects when language is available. The findings support the linguistic bootstrapping hypothesis that representing familiar object concepts normally relies on language, and that implicitly-retrieved object labels, used as linguistic placeholders, can increase the number of objects that can be simultaneously represented beyond what sensorimotor information alone can accomplish.

1. Introduction

There is a broad consensus in the cognitive sciences that the conceptual system consists of simulation- and linguistic-based components (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008; Vigliocco et al., 2009). Simulated representations engage the neural subsystems involved in sensorimotor, affective, introspective, and other situated experience of a concept (e.g., Barsalou, 1999; Martin, 2007). For example, the experience of a *dog* may include its visual shape and colour, the action and feel of patting its fur, the sound of its bark, the broader situation of walking it on a leash, and the love and positive feelings towards a pet. The neural activation patterns involved in processing these experiences can be partially re-activated (i.e., simulated) at a later time when representing a concept. Linguistic representations of concepts, on the other hand, comprise word (and phrase) labels associated with these sensorimotor-affective simulations; for instance, seeing a terrier or hearing a bark will activate the label "dog", as well as other associated words that represent experiences in related contexts, such as "tail", "walkies", or "leash" (e.g., Louwerse, 2011; Wingfield & Connell, 2022). These simulated and linguistic components are interrelated and mutually supportive, and recent theories argue that both are intrinsic to conceptual representation (e.g., Connell & Lynott, 2014; Louwerse, 2011). That is, linguistic labels are part of concepts, and conceptual processing utilises simulation and linguistic information to varying extents depending on task demands, available resources, and other factors (Connell, 2018; Connell & Lynott, 2014).

The role of both simulation and linguistic components in conceptual processing is supported by a range of empirical evidence. Support for sensorimotor simulation comes from neuroimaging of sensory and motor cortices during word processing (e.g., Goldberg et al., 2006; Hauk et al., 2004), neuropsychology of motor impairment (e.g., Boulenger et al., 2008; Fernandino et al., 2013), and a variety of behav-

a Correspondence concerning this article should be addressed to Agata Dymarska, Department of Psychology, Fylde College, Lancaster University, Lancaster LA1 4YF, UK. Email: a.dymarska@lancaster.ac.uk

ioural paradigms involving perceptual or action manipulations (e.g., Bidet-Ildei et al., 2017; Connell et al., 2012; Davis et al., 2020). Support for the linguistic component comes from computational modelling of conceptual information captured in language (e.g., Banks et al., 2021; Riordan & Jones, 2011; Wingfield & Connell, 2022) and from behavioural paradigms showing that information from language alone can inform responses in diverse conceptual tasks (e.g., Connell & Lynott, 2013; Goodhew et al., 2014; Louwerse & Jeuniaux, 2010; see Connell, 2018, for review). For example, because the linguistic component has a relative speed advantage over the simulation component (Barsalou et al., 2008; Connell, 2018; Louwerse, 2011), responses that rely on language tend to be faster and less effortful than those that rely on sensorimotor simulation (e.g., Louwerse & Connell, 2011; Santos et al., 2011). However, much evidence for the linguistic component centres on the usefulness of linguistic distributional knowledge (i.e., the statistical patterns of how words/phrases co-occur across language: see Wingfield & Connell, 2022, for review), which does not encompass the full role of language in conceptual processing. The very existence of linguistic labels – that is, being able to concisely name a complex multimodal experience with a word or phrase, and use this label to activate the complex experience at a later time – provides another means for the linguistic component to enhance the efficiency of conceptual processing.

The idea that language is beneficial for our cognitive processing has existed for some time (e.g., Paivio, 1971; Vygotsky, 1934/1986). Recent theories have, however, 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 the present article, Connell and Lynott (2014, p. 7) 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, thus enhancing the achievable size and complexity of what can be held in mind. That is, when the sheer scale or complexity of what one is trying to simulate at a given moment outstrips the limited resources of the human cognitive system, replacing a portion of the simulation with a linguistic placeholder can preserve structure in the representation while freeing up resources to maintain or extend the simulation as needed. These linguistic placeholders can later be fleshed out into a simulation again at any time if resources become available or the task requires it, which has implications for our ability to remember and manipulate multiple concepts in everyday life. For example, when following a recipe to make a cake, knowing the next steps and what comes after the current ingredient is beneficial for performing the task efficiently, and using labels potentially increases the number of steps which can be planned ahead. When keeping in mind labels of the ingredients (e.g., *flour*), rather than their complex, situated sensorimotor representations (e.g., a fine white powder in a red-and-white paper bag), a person can maintain an economical representation of the object concept. When it is time to find flour on the shelf, they can keep the label in mind to compare with the implicitly-activated labels of objects in front of them, *or* they can flesh out the sensorimotor simulation of flour and compare with the actual objects in front of them, until flour is located and the task can be completed successfully. Either way, because a label occupies less representational "space" than a sensorimotor simulation, having language available to label concepts could increase the number of items that can be represented during the task at a given time.

To date, the linguistic bootstrapping hypothesis has remained theoretical and has not been tested directly. Nonetheless, there is indirect support for the idea in the wider literature, particularly in working memory research. According to the most recent versions of the multi-component working memory model (Baddeley, 2012; Baddeley et al., 1984), when processing complex stimuli, information from multiple modalities is integrated with conceptual representations from long-term memory and stored in the episodic buffer. This episodic buffer storage is necessarily limited in capacity: that is, there are only so many concepts that can be maintained and manipulated at once. Empirical studies estimate the capacity of the episodic buffer (for combination of letters/digits and their spatial location) to be from 3 items (Langerock et al., 2014) to 5 or 6 items (Allen et al., 2015). Critically, other studies suggest that linguistic information is more economical in representation (i.e., may occupy less "space" in working memory) than sensory information, and may thus allow this capacity to be increased. For example, explicitly labelling simple visual stimuli (e.g., dots of different colours) increases memory capacity compared to unlabelled stimuli (Forsberg et al., 2020; Souza et al., 2021; Souza & Skóra, 2017; Zormpa et al., 2019). Additionally, language appears to be the form of object representation that people rely on by default, at least some of the time. That is, participants tend to remember category labels rather than specific item characteristics (Brandimonte et al., 1992), are faster to respond to objects when hearing a label instead of a nonverbal cue (e.g., "cat" vs the sound of a meow; Lupyan & Thompson-Schill, 2012), and can detect an object faster and more accurately when they are given its label (Lupyan & Ward, 2013; Ostarek & Huettig, 2017). Linguistic bootstrapping theory suggests that when working memory capacity is strained to its limit, such as when trying to maintain a representation of numerous concepts, a linguistic label could deputise for its referent sensorimotor information (e.g., the word "dog" could replace the simulation of a *dog*) in order to free up space and thereby increase the number of objects that can be represented simultaneously. Examining memory capacity for concepts may thus provide a means for directly testing the linguistic bootstrapping hypothesis, as well as for estimating the potential benefit to working memory capacity afforded by linguistic labels. To return to our earlier example, how efficient could it be to select the correct ingredients for a recipe when labels are not available, and one has to hold in mind complex sensorimotor representations of flour, eggs, milk, etc.? In an experimental setting, this question can be tested by asking participants to perform a

secondary verbal task, which will prevent them from being able to rely on linguistic labels.

Repeating a syllable or word (such as "la" or "the") while performing a concurrent task makes it harder to access linguistic representations, since the language processing system is occupied (Baddeley et al., 1984; Richardson & Baddeley, 1975), but it does not affect processing of non-verbal information (e.g., Jaroslawska et al., 2018; Wood et al., 2020). Articulatory suppression is commonly used in working memory research (see Baddeley, 2012) when investigating the role of language in, for example, remembering a list of random words or letters (Baddeley et al., 1984), a set of geometrical shapes or colours (Souza & Skóra, 2017), or in mentally manipulating a visual image (Brandimonte et al., 1992). Nevertheless, studies from the memory literature that have examined the role of language have predominantly focused on explicit verbal stimuli or subsidiary visual features (i.e., unimodal surface characteristics such as colour or shape), and – to the best of our knowledge – have not examined the benefits of labels in remembering multimodal, real-world object concepts. Studies from the linguistic-simulation perspective have examined memory for real-world objects, finding that performing a secondary task in a particular sensory modality or with a particular motor effector impairs memory for concepts associated with that perceptual or motor dimension (e.g., Dutriaux et al., 2018; Vermeulen et al., 2013; see also Shebani & Pulvermüller, 2013). However, these studies focused on *sensorimotor* representations of objects – that is, the ability to simulate their perceptual characteristics or motor affordances, such as visual and auditory information or hand and feet movements – rather than the *linguistic* representation of objects via their labels. These findings therefore show that sensorimotor information is important to how object concepts are represented in memory during conceptual processing, but they do not address the role of linguistic labels in these representations.

Moreover, all the above studies use the typical laboratory paradigm of (pseudo-)random sets of unrelated stimuli, which do not easily generalise to the real-world circumstances that would require people to represent multiple concepts simultaneously in working memory. A more naturalistic situation of holding multiple representations in mind would involve concepts that comprise rich sensorimotor and linguistic information from long-term memory, that are typically embedded in broader situated simulations that allow concepts to reinforce and cue one another (e.g., a *dog* that is *running* with a *ball*; a *cookie* which is *served* with a *jug* of *milk*, etc.). As outlined earlier, during a task such as following a recipe, the order of a sequence and the relationship between the ingredients matters as much as simply being able to list them. Critically, it has not yet been investigated whether the use of labels to represent these concepts can increase the number of objects which can be successfully held in mind, by reducing the size and complexity of each representation (i.e., via linguistic bootstrapping). The maximum sequence length of contextually-related, real-world object concepts that people can hold in mind therefore remains unknown.

1.1. The Current Study

The present study had two main aims: to test the linguistic bootstrapping hypothesis in the context of real-world object concepts, and to establish the number of objects, presented in a contextually-situated sequence, that can be held in mind both when access to linguistic labels is available and when it is not. In three pre-registered experiments using a nonverbal paradigm, we presented ecologically valid sequences of object pictures from naturalistic situations (e.g., ingredients for a novel recipe) and then asked participants to select the previously-presented objects from arrays of distractors. Critically, in Experiment 1, participants performed articulatory suppression (i.e., repeated aloud "the") while learning the sequences of objects and/or while being tested on the objects, in order to inhibit access to linguistic information (i.e., object labels; Baddeley et al., 1984; Richardson & Baddeley, 1975) while leaving access to sensorimotor simulation unaffected. We hypothesised that, even in ostensibly nonverbal paradigms, object concepts would normally be held in mind using language (i.e., as implicitly-retrieved object labels, used as linguistic placeholders), and that suppressing access to language would impair speed and accuracy, and reduce the number of objects which can be simultaneously represented. In Experiment 2, we included an additional control condition of foot tapping in order to compare performance in the articulatory suppression condition with a secondary task that had similar attentional demands but would not affect access to linguistic labels. Finally, in Experiment 3, we addressed the possibility of ceiling effects in earlier studies by using sequences of increasing length to determine the maximum number of contextually-situated, real-world object concepts that participants could hold in mind when linguistic labels were fully available (i.e., without articulatory suppression).

2. Experiment 1: Object Memory with Articulatory Suppression

In this study (pre-registration, materials, data, analysis code, and full results are available as supplemental materials at https://osf.io/mwzfh) we presented participants with images of natural and artifact objects, arranged in sequences of twelve items that would plausibly be experienced in a real-world setting, and asked them to remember each sequence. After each sequence, we tested memory for the objects by asking participants to choose each remembered object from an array of related distractors, and measured speed and accuracy of performance. Participants performed articulatory suppression while learning the sequences and/or while being tested on the objects. Following the linguistic bootstrapping hypothesis, we predicted that performance would be impaired when access to language was inhibited, such that articulatory suppression at either stage would lead to slower responses and more errors in identifying remembered objects. We expected performance to be best with no articulatory suppression at all, where participants would be free to utilise both linguistic and sensorimotor information to remember the objects. In addi-

tion, we expected performance to be worst with articulatory suppression during the test stage only, due to participants employing linguistic placeholders to replace sensorimotor information when learning the sequences, and then losing access to those placeholders (and thus the object representations) when access to linguistic information was suppressed in the test stage. Finally, by calculating the average number of objects that participants correctly recognised when performing articulatory suppression both while learning and while they were being tested on the sequences (i.e., when linguistic information was suppressed throughout the entire task), we planned to estimate the maximum sequence length of real-world concepts which can be held in mind using sensorimotor representations alone.

2.1. Method

2.1.1. Participants

Forty-four native speakers of English (33 female; mean age = 20.3 years, *SD* = 5.4) were recruited from Lancaster University, and received course credit or a payment of £3.50 for participation. Three initially-recruited participants were replaced: one due to not following instructions correctly, one who had previously participated in a pilot study, and one due to not being a native speaker of English. The sample size was determined using sequential hypothesis testing with Bayes Factors (Schönbrodt et al., 2017), and the threshold for inference was BF_{10} = 5 or its reciprocal BF_{01} = 0.2. Bayes Factors for Step 3 cleared the evidence threshold for the null hypothesis at *N*min = 32 for both RT (BF_{01} = 0.02) and accuracy (BF_{01} = 0.03). However, sequential analysis plots for the Step 2 model suggested that the level of evidence was still unstable for RT (i.e., BFs fluctuated with successive participants between evidence for the null, equivocal evidence, and evidence for the alternative), so we opted to deviate from the pre-registered stopping rule and tested additional participants until Bayesian evidence stabilised at *N* = 44 (see Results section for statistics). We therefore report results for 44 participants, but full analyses at the original *N*min = 32 are available in supplementary materials, and we note that parameter estimates remained consistent between *N* = 32 and *N* = 44.

2.1.2. Materials

Test items comprised a total of 72 target objects, divided into 6 sequences which were designed to be ecologically valid orders of objects that would be plausibly used in a real-world setting, such as a shopping list for lunch (ingredients used in the process of making a cake and a sandwich), items used for decorating an office (office supplies to move in and a set of tools used in order to hang a picture), or a packing list for a camping trip (camping equipment and an outfit one might dress in). We initially intended to use sequences of 6 objects, but a pilot study (see supplemental materials) revealed that it resulted in ceiling effects; we therefore paired these 6-object sequences to create longer sequences of 12 objects to remember, with each pair forming a naturalistic situated context (e.g., the ingredients for a cake and a sandwich formed a single shopping list). Each sequence therefore consisted of 12 target objects for study, and each target object was assigned five distractor items for display in an object array during the testing stage. The five distractor objects were selected from the same semantic category as the target (e.g., food items, clothing). In order to make sure that participants could not rely on any individual perceptual characteristic when remembering objects (e.g., encoding only a red colour instead of a *tomato* object), we ensured that at least one of the distractors shared its colour with the target object, at least one shared the shape, and at least one shared its function. We considered these three characteristics the most salient and the most likely to influence participant responses. Although not in a systematic way, further distractors shared perceptual or function characteristics with either the target or other distractors, to avoid any distractor being easily eliminated as the odd-one-out. The target and distractor items in each sequence (e.g., a recipe) could all plausibly be used for similar activities, so that the task maintained ecological validity, and it would not be obvious from the nature of the sequence which item in the array was the correct one (see sample stimuli in Figure 1). Before each sequence, participants were given some general information about the context (e.g., "You are preparing lunch for the next day, and need to remember the shopping list to make a cake and a sandwich. Press space to proceed to the list of ingredients."), to provide a plausible real-life situation that was ecologically valid.

We sourced photographic images for all objects from license-free online resources and edited them to appear on a uniform transparent background. Critically, in order 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 and one for display in the distractor array. Both images represented good examples of the target object and differed only in minor aspects (e.g., showing a vegetable from a different perspective, or a piece of clothing in a different colour). Images were scaled to be 840 pixels along the longest dimension for target objects presented during the learning stage, and 470 pixels along the longest dimension for objects (targets and distractors) presented in the object array during test. This process resulted in a total of 504 object images: 72 target objects to learn in sequences, 72 target objects, representing the same concepts, to be recognised in the test stage, and 360 distractor objects. Figure 1 shows sample stimuli in a trial sequence diagram.

To ensure the order of target objects within each sequence was ecologically valid, we asked 9 naïve participants (who did not take part in the main studies) to rank-order the items in each subset according to how they would be used in the given situated context. For example, in the context of decorating an office where a number of objects related to hanging a picture on the wall, participants had to decide the order in which they would use the following objects: "spirit level", "drill", "screw plug", "screw", "screwdriver", "picture frame". We then finalised each sequence

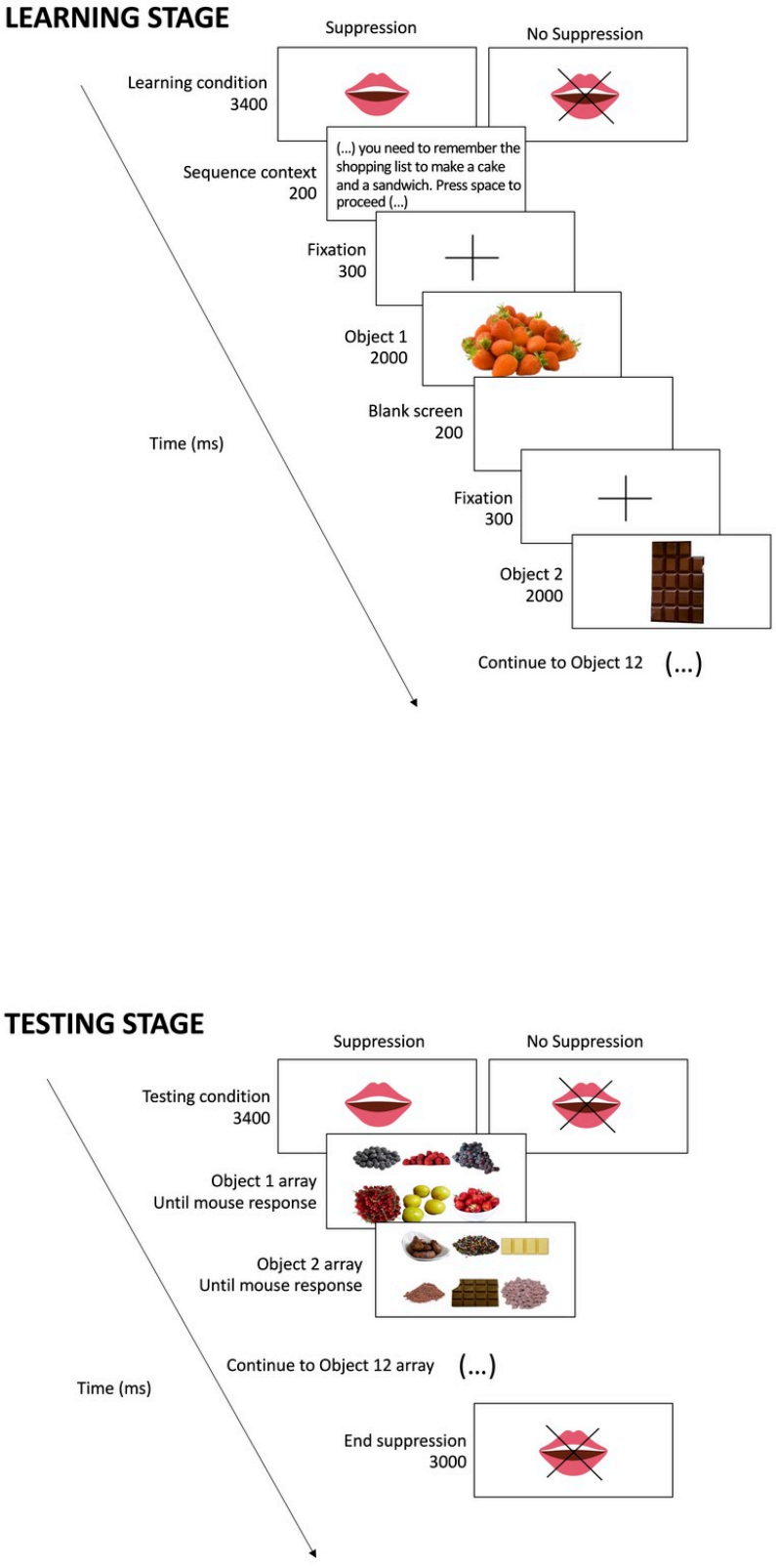

Figure 1. Diagram showing trial sequence and example stimuli at learning (above) and test (below) stages in Experiment 1.

according to the mean rank per object. Target objects were always presented in the same ecologically-valid order during both the learning and test stages.

2.1.3. Procedure

Participants were tested individually. After signing the consent form, which included consent to publicly share their anonymised data, they sat in front of a computer and were informed that they would perform a memory task, and that they would be asked to repeat the word "the" at some point during the task. We chose to use the word "the" for the articulatory suppression task (as opposed to pseudo-nonsense syllables such as "la" or numbers such as "one") because it was a real word that participants were practiced at articulating, but, as a function word, was semantically empty in isolation and so unlikely to activate any linguistic or sensorimotor information that could interfere with the task (see Weywadt & Butler, 2013). Participants were also shown an image of a mouth, which indicated when they should start performing articulatory suppression, and the same image crossed out to indicate that they should not perform articulatory suppression. This image cue was presented before both the learning and testing stages for each sequence. The experimenter then explained and demonstrated articulatory suppression, and asked the participant to practice it. Once the participant confirmed that they understood the instructions and the purpose of the image cues, and could perform articulatory suppression correctly, they provided demographic information and read the instructions onscreen.

Participants were instructed that 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 belonged to the sequence they had been asked to remember. Participants then commenced a practice sequence of twelve items (not used in the main experiment), without any articulatory suppression. After the practice session, when the participant confirmed that they understood the task and were happy to continue, they commenced the experimental trials. Articulatory suppression was manipulated between-participants while learning the sequences and within-participants while being tested on the objects (i.e., object recognition), producing four crossed experimental conditions: no-suppression/no-suppression, no-suppression/suppression, suppression/no-suppression, and suppression/suppression. The test stage of half the sequences was performed with articulatory suppression and half without, in random order, so that participants did not know whether articulatory suppression applied at testing until they had already learned the whole sequence. Experiment presentation was controlled by PsychoPy software (version 1.84.1; Peirce, 2009).

In the learning stage, participants in the articulatory suppression condition commenced repeating "the" aloud before each sequence began. The context for the sequence was first presented onscreen until the participants read the text and pressed the space bar to continue. Target objects were then presented individually in their fixed sequence, starting with a blank screen for 200 ms, followed by a central fixation cross for 300 ms, and then the target object for 2000 ms (see Figure 1). Once a full sequence of twelve target objects had been presented, and before the testing stage began, participants were shown a fake distractor task, where they had to click on 4 dots appearing in 4 corners on the screen in a random order to "calibrate the mouse". This dot-clicking subtask was to eliminate covert rehearsal in the no-suppression condition (which would give an advantage compared to the suppression condition, where verbal rehearsal was not possible), as well as to reduce the possibility that participants were using visualising strategies and focusing on specific perceptual features from the presented image (e.g., a round thing) instead of relying on memory for the holistic object concept (e.g., a tomato). If participants were performing articulatory suppression when learning the sequences, they continued repeating "the" aloud until this dot-clicking task timed out.

In the test stage, for sequences in the articulatory suppression condition, participants commenced (or continued) repeating "the" aloud before the first array appeared. On each trial, participants saw a 2x3 array of six objects, comprising one target object and five distractors in random locations within the array. Response times were measured from the onset of the array display until the onset of the mouse click. There was no time limit for the response. After twelve arrays had been displayed (for recognition of twelve target objects), a message appeared on the screen asking participants to press space when they were ready to proceed to the next sequence of objects. Participants could take a self-paced break after every sequence. The entire experimental procedure took approximately 15-20 minutes.

2.1.4. Ethics and Consent

All studies received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee. All participants read information detailing the purpose and expectations of the study before giving informed consent to take part. Consent included agreement to publicly share all alphanumeric data in anonymised form.

2.1.5. Design and Analysis

We analysed accuracy (with incorrect responses coded as 0, and correct responses coded as 1) with a mixed-effects hierarchical logistic regression (binomial, logit link). Participants and items (nested within sequences) were included as crossed random effects. We included fixed effects of articulatory suppression at learning and at test (dummy coded: no-suppression coded as 0, articulatory suppression coded as 1), and their interaction. Response times (RT; ms) for correct responses were analysed in a mixed-effects hierarchical linear regression with the same random and fixed effects as above.

In all regression analyses, Step 1 entered random intercepts, Step 2 added learning and test stages as fixed effects, and Step 3 added the interaction of learning and test. We ran Bayesian model comparisons between steps, with

Bayes Factors (BF) calculated via Bayesian Information Criteria (Wagenmakers, 2007), in order to quantify the evidence for or against the added step (threshold for inference was $BF_{10} = 5$ or its reciprocal $BF_{01} = 0.2$). We also report null hypothesis significance testing (NHST) statistics for parameter coefficients in the Step 3 model, which we used to estimate the marginal average accuracy for each condition of articulatory suppression. The marginal average accuracy was then used to calculate the maximum number of objects which can be successfully held in mind when language was or was not fully available at object learning and at object testing. All analyses were run in R software (lme4 package, Bates et al., 2015; lmerTest package, Kuznetsova et al., 2017; R version 3.4.1, 2017).

2.2. Results

Based on pre-registered exclusion criteria, no trials were excluded for accuracy analysis. For analysis of correct RTs, no trials were removed as motor errors, but 31 trials (1.2% of data) were removed where RTs were more than 3 SDs above the individual participant's mean.

2.2.1. Confirmatory Analysis

Accuracy. Bayesian model comparison showed strong evidence *against* Step 2 over Step 1, $BF_{10} = 0.02$; that is, the data were $BF_{01} = 57.40$ times more likely under the Step 1 model containing only random intercepts than a model containing articulatory suppression at learning and test. There was also strong evidence at Step 3 *against* the effect of the learning*test interaction on accuracy, $BF_{10} = 0.03$: that is, the data were $BF_{01} = 40.45$ times more likely under the Step 2 model without the interaction than the Step 3 model with the interaction.

We then used the coefficients in the Step 3 model (Table 1) to estimate the marginal accuracy for each condition of learning*test articulatory suppression (see Figure 2). Both learning and test parameters had negative coefficients but only learning had a significant effect in NHST terms, indicating that performing articulatory suppression while learning sequences impaired participants' accuracy. The highest accuracy score was in the no-suppression/no-suppression condition (i.e., no articulatory suppression at either learning or test), with participants correctly recognising on average 11 ± 0.2 (M ± SE) out of 12 objects per sequence, and was lowest in the suppression/suppression condition (9.9 ± 0.4 objects remembered) rather than in the no-suppression/suppression condition as we had hypothesised. That is, object memory was least accurate when access to language was suppressed at both learning and test.

Response Times. Bayesian model comparison showed equivocal evidence against Step 2 over Step 1, $BF_{10} = 0.61$, that is, the RT data were $BF_{01} = 1.65$ times more likely under a model with only random intercepts than a model that contained fixed effects of articulatory suppression at learning and test. There was strong evidence at Step 3 *against* the presence of a learning*test interaction, $BF_{10} = 0.02$; that is, data were $BF_{01} = 46.99$ times more likely under the non-interaction model than the interaction model (see Table 1).

As before, we used the coefficients in the Step 3 model (Table 1) to estimate the marginal mean RT for each articulatory suppression condition (see Figure 2). The test coefficient was negative and significant in NHST terms (i.e., unexpectedly faster performance under articulatory suppression), but this time articulatory suppression at learning had no effect. Against our predictions, selection of target objects was fastest in the no-suppression/suppression condition (i.e., when language was available at the point of learning sequences, but not when being tested on target objects), and slowest in the suppression/no-suppression condition (i.e., when language was available at the point of testing objects, but not learning them). That is, participants had most difficulty remembering objects when language access was suppressed at the point of learning only but was available at test.

2.2.2. Exploratory Analysis

Because our confirmatory analysis produced some unexpected results, we ran exploratory analyses to determine the best-fitting model for accuracy and RT. We first explored alternative random effects structures for the null model using Restricted Maximum Likelihood (i.e., random learning and/or test slopes for participants and/or items) and selected the best model using Bayes Factors calculated via BIC. We then examined (using Maximum Likelihood) whether the data favoured a learning-only (suppression at learning stage), test-only (suppression at testing stage), or learning+test model (suppression at both stages; the original Step 2 in confirmatory analysis) in comparison to the null model. Since the NHST results at Step 3 indicated that articulatory suppression elicited an effect at the learning stage on accuracy, and at the test stage on RT, this exploratory analysis allowed us to clarify whether there was Bayesian evidence for these effects when articulatory suppression at each stage was examined separately. For NHST coefficient statistics, we Bonferroni-corrected the p-values for each parameter by multiplying by 3. Model comparisons results are presented in Table 2.

Best-Fitting Model of Accuracy. Attempts to model random slopes led to non-convergence in all models (see supplementary materials for full results). As a result, we used models without random slopes (i.e., random intercepts only for participants and items, as per confirmatory analysis) as the null model in explorations of fixed effects on accuracy.

Bayesian model comparisons showed that the best-fitting model was an equivocal tie between the learning-only model and the null model, where the learning-only model was $BF_{10} = 16.67$ times better than the next-best alternative model (test-only). NHST results from the learning-only model indicated that articulatory suppression at learning impaired accuracy (b = -0.735, SE = 0.273, z = -2.691, p =.007), as predicted: participants were 109% more likely (i.e., more than twice as likely) to make an error during object testing if their access to labels had been inhibited when learning the object sequences. However, although this ef-

Table 1. Experiment 1 unstandardized regression coefficients, standard errors, and associated statistics from Step 3 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for articulatory suppression effects at learning, test, and their interaction.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
Accuracy	Intercept	2.437	0.252	-	9.678	<.001
	Learning	-0.824	0.294	-	-2.808	.005
	Test	-0.220	0.159	-	-1.388	.165
	Learning*Test	0.171	0.205	-	0.835	.404
					t	
RT	Intercept	2786.07	139.23	30.08	20.011	<.001
	Learning	68.73	155.66	51.22	0.442	.661
	Test	-150.49	60.87	2429.19	-2.472	.014
	Learning*Test	-29.11	86.48	2430.96	-0.337	.736

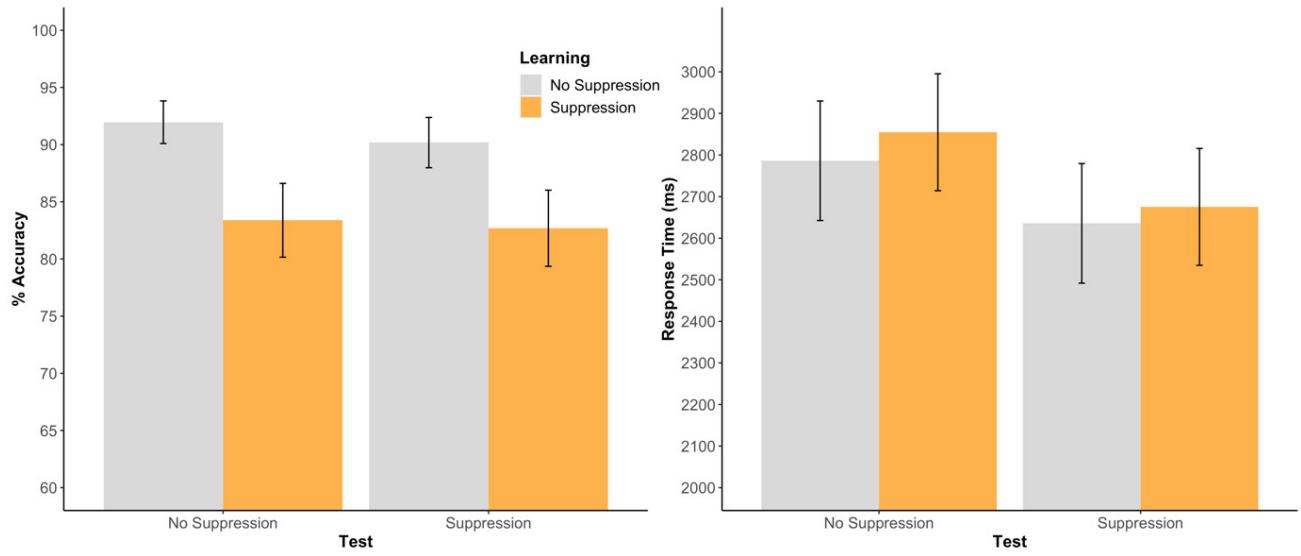

Figure 2. Mean % accuracy and RT per articulatory suppression condition in Experiment 1, calculated as marginal means from the Step 3 models. Error bars represent ± 1 Standard Error.

Table 2. Exploratory analysis of Experiment 1, showing Bayes Factor comparison of each candidate model against the null model (random intercepts only).

DV	Candidate model	BF_{10}
Accuracy	Learning only	0.50
	Test only	0.03
	Learning+Test	0.02
RT	Learning only	0.02
	Test only	27.66
	Learning+Test	0.61

fect was significant in NHST terms, evidence for the learning-only model was very weak in Bayesian terms (i.e., equivocal evidence for the null), and so we treat the effect with caution.

Best-Fitting Model of Response Times. Exploration of random slope structures showed that the best fit emerged from the model without random slopes, as no slope model met the $BF_{10} > 5$ threshold for improving model fit. Therefore, we report models with no slopes. Full statistics can be found in supplemental materials.

In explorations of fixed effects, Bayesian model comparisons showed that, in contrast to the results for accuracy, the best-fitting model was the test-only model, which was BF_{10} = 45.34 times better than the next-best alternative (learning+test model). In the test-only model, articulatory suppression at test had a negative effect on RT (b = -164.89, SE = 43.26, t(2430) = -3.812, *p* =.001), indicating that when language access was suppressed during the testing stage, participants were 165ms *faster* to respond than when they had access to language. Closer examination of RT and accuracy effects suggested that the effect at test was not necessarily due to a speed-accuracy tradeoff: when participants were asked to perform articulatory suppression during object testing, response times were faster, but there was no

accompanying drop in accuracy. We discuss this point more below.

2.3. Discussion

Experiment 1 examined the role of language in remembering real-world object sequences. We found that articulatory suppression during sequence learning weakly impaired accuracy, but not speed of response. This learning effect on accuracy appeared only in NHST coefficient statistics during confirmatory and exploratory analysis, but Bayesian evidence was equivocal in exploratory analysis of fixed effects; we therefore interpret it cautiously as a somewhat weak effect. When language was suppressed at the point of learning, participants' ability to remember objects was impaired to some extent, partially supporting the linguistic bootstrapping hypothesis.

We also found that articulatory suppression during object testing led to *faster* response times in both confirmatory analysis and exploration of best-fitting models, but had no effect on accuracy. It is possible that one of the ways people might use language to support target selection at the test stage is by implicitly naming all the objects presented in the array, which could introduce a processing delay compared to when language access is suppressed and implicit naming cannot take place (see Phillips et al., 1999, for a similar articulatory suppression finding). We therefore conclude that inhibiting language access at the point of object testing does not lead to slower response times (as we originally hypothesised), but instead speeds up response times as participants have less information to process. We return to this point in the general discussion.

Finally, and importantly, we did not find an interaction of learning*test on either accuracy or RT, which was perhaps unsurprising given the unexpected facilitatory effects of articulatory suppression at the test stage. This result did not support our hypothesis that participants would find it most difficult to remember objects when they had language available during sequence learning that was later suppressed at the test stage. Rather, in such circumstances, it appears that participants were still able to identify targets fairly successfully, potentially on the basis of encoded sensorimotor information.

Based on the results of Experiment 1, we calculated that a sequence of 9.9 object concepts on average can be held in mind when relying on sensorimotor information only (suppression/suppression condition), but that this increases to 11.0 objects when linguistic labels are available (no-suppression/no-suppression condition). This finding is in line with the hypothesis that the maximum number of contextually-situated objects that can be simultaneously represented is effectively greater when language is available to act as a placeholder for a full sensorimotor simulation.

The results demonstrate that people's ability to remember sequences of contextually-situated, real-world objects is very good, even when the sequence itself is new, but is ecologically valid and embedded in existing knowledge. For example, being able to remember a shopping list or a recipe of 10-11 items is well within the capabilities of our participants.

Overall, Experiment 1 partially supported our hypotheses regarding the role of linguistic bootstrapping in memory for object concepts. In an ostensibly nonverbal task, suppressing access to linguistic labels via articulatory suppression at the point of learning led participants to remember fewer objects compared to when language was available. This effect did not occur when language was suppressed during object testing, likely due to confounding effects of strategies participants used which led them to respond faster when language was not available. However, the effect at learning was somewhat weak, possibly due to employing a between-participant manipulation of articulatory suppression at the learning stage. Moreover, it could be argued that our findings instead reflect dual task performance (i.e., impaired accuracy due to performing a secondary task), rather than articulatory suppression specifically removing access to the labels that normally enable linguistic bootstrapping. Although a large body of previous research has established that performing articulatory suppression inhibits access to linguistic information (i.e., affecting the ability to remember language stimuli such as numbers, letters, or words) without affecting executive load or nonverbal processing (e.g., Jaroslawska et al., 2018; Murray, 1967), we did not directly control for dual task load in this first experiment. That is, the impairment in accuracy could be due to participants performing articulatory suppression as a secondary task, which takes up some of the available cognitive resources, and not due to the fact that labels cannot be used as placeholders for complex sensorimotor simulations. We address both these possibilities in the next experiment.

3. Experiment 2: Foot-Tapping as Dual-Task Control

In our second experiment (pre-registration, raw data, analysis code, and stimuli are available as supplemental materials at https://osf.io/mwzfh), we had two goals: to replicate and strengthen the effect of articulatory suppression at learning that emerged in Experiment 1 in support of the linguistic bootstrapping hypothesis, and to test whether this effect is specifically due to suppressing access to language rather than due to performing a secondary task. We therefore compared articulatory suppression during learning to foot tapping, a secondary task of comparable difficulty (see Van 't Wout & Jarrold, 2020) that is unrelated to language use but comparable on a number of other characteristics (i.e., it is a rhythmic motor task, does not involve visual perception or hand action that could interfere with stimulus presentation, and it can be sustained throughout the trial without undue fatigue: see Gaillard et al., 2012). This time, we manipulated the secondary task only at the point of learning the object sequences, and did so *within* participants, so that all participants learned the sequences in all three conditions (i.e., with no secondary task, while performing articulatory suppression, and while performing foot tapping). This design allowed us to evaluate the effects of articulatory suppression on memory for objects without the confounding effects of participant strategies present in Experiment 1 at the testing stage (e.g., not labeling objects

when they anticipated articulatory suppression at test). Critically, the comparison between articulatory suppression and foot tapping allowed us to determine whether the effects in Experiment 1 were specifically due to suppressing access to linguistic labels, rather than simply due to performing a concurrent distractor task.

As in the previous experiment, we predicted that accuracy and latency of performance would be impaired when access to linguistic labels is inhibited via articulatory suppression compared to when language is available (i.e., no secondary articulatory suppression task). In addition, we hypothesised that this impairment would not be solely due to performing a secondary task, and that inhibiting access to language via articulatory suppression would lead to greater impairment of object memory than a secondary task of foot-tapping that does not affect access to language.

3.1. Method

3.1.1. Participants

Eighteen native speakers of English (18 female; mean age = 18.6 years, *SD* = 0.8 years) were recruited from Lancaster University, and received course credit or a payment of £3.50 for participation. Data from two originally-recruited participants were replaced due to not being native speakers of English. As pre-registered, we used Bayesian sequential hypothesis testing to determine sample size, and stopped at the minimum sample size *N*min = 18 when our Step 3 models for both RT and accuracy cleared the specified threshold of evidence $BF_{10} > 5$ or its reciprocal $BF_{01} < 0.2$ for three successive participants (see Design and Analysis section for model details; full statistics are reported in the Results section).

3.1.2. Materials

We used the same materials as in Experiment 1 with the following changes. We created 3 additional sequences of 12 objects apiece, bringing the total number of target objects to 108, divided into nine sequences; this number allowed for three sequences to be tested per secondary task condition. As per the existing sequences, each new sequence represented an ecologically-valid order of objects that would be plausibly used in a real-world setting, and were labelled with a brief description that provided a naturalistic, situated context. We also altered 3 target items and 1 distractor item in the existing sequences, to avoid duplicating targets in the new sequences. The order of objects for new sequences was determined by 8 volunteers who did not take part in any of the studies, and was established based on their mean rank as in Experiment 1.

Photographic images of new target and distractor objects were sourced and edited as per Experiment 1. In total, the present experiment utilised 756 object images: 108 target objects presented for sequence learning, 108 target objects presented for testing (i.e., different images to the learning stage), and 540 distractor objects presented at test.

3.1.3. Procedure

As before, participants were tested individually and consented to publicly share their anonymised data. They sat in front of a computer and were informed that they would be asked to perform a secondary task at some point during the experiment. The experimenter then presented them with three symbols which would be used to signal what they should do for each task condition, and then explained what the tasks involved. A picture of a mouth (the same as in Experiment 1) was used to indicate that participants should repeat the word "the"; a picture of a foot indicated that participants should tap their foot continuously, and a picture of an X was used to indicate that they should stop or not perform either of the tasks (i.e., to indicate the control condition of no secondary task). The experimenter demonstrated both secondary tasks, where articulations of "the" and foot taps were repeated at approximately the same rhythmic rate, and asked the participant to practice them. Once the participant confirmed that they understood and could perform the tasks correctly to the experimenter's satisfaction, they provided demographic information and read the instructions onscreen.

The secondary task was manipulated within-participants at the learning stage; that is, participants took part in each of the three secondary task conditions, where the order of conditions was rotated in a latin-square design. The object sequences were divided into three sets of three sequences apiece, and the assignment of each set to a secondary task condition was counterbalanced across participants. Within each condition, sequences were presented in a randomised order, and each sequence appeared in each condition an equal number of times. Participants did not perform any secondary task at the test stage.

As before, participants were instructed that 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 belonged to the sequence they had been asked to remember. Participants first commenced a practice sequence without any secondary task, which provided time-related feedback to habituate them to responding within a time limit ("good job" was displayed on screen if the response was given on time, and "too slow" if they failed to respond within 6000 ms). When the participant confirmed that they understood the task and were happy to continue, they commenced the experimental trials. After a brief description of the sequence, the image indicating the secondary task condition was displayed for 3000ms, and then the 12 objects were presented one by one. The break between learning the sequence and being tested on it was the same as in Experiment 1 (clicking on 4 dots on the screen), after which participants were asked to stop the secondary task. The test stage then proceeded as per the no-suppression condition in Experiment 1. There was no feedback on experimental trials, but the trial timed out and was marked as incorrect if participants failed to respond within 6000ms.

3.1.4. Design and Analysis

We analysed accuracy (incorrect = 0, correct = 1) in a mixed-effects hierarchical logistic regression (binomial, logit link), and response times (RT) for correct responses in a mixed-effects hierarchical linear regression. For both accuracy and RT analysis, participants and items (nested within sequence) were included as crossed random effects. Fixed effects were dummy coded using articulatory suppression as the reference level, which allowed us to test each critical hypothesis (i.e., that articulatory suppression would be worse than no task *and* worse than foot tapping) with a distinct parameter. That is, we included two learning stage variables as fixed effects: no task (representing no-task vs. a secondary task at learning: 1 = no task, 0 = foot tapping or articulatory suppression) and foot tapping (distinguishing foot tapping as a secondary task: 1 = foot tapping, 0 = no task control or articulatory suppression).

In regressions of both accuracy and RT, Step 1 entered random intercepts, Step 2 added no-task as a fixed effect, and Step 3 added foot-tapping as a fixed effect. We ran Bayesian model comparisons between steps, with Bayes Factors (BF) calculated via BIC as in the previous experiment. We also report null hypothesis significance testing (NHST) statistics for parameter coefficients in the Step 3 model. Specifically, the no-task coefficient in Step 3 allowed us to test the hypothesis that articulatory suppression produced a greater impairment than no secondary task (i.e., replicating Experiment 1), and the Step 2-3 model comparison along with the foot-tapping coefficient in Step 3 allowed us to test whether articulatory suppression produced a greater impairment than foot tapping (i.e., the critical new hypothesis of the present experiment).

We used the Step 3 coefficients to calculate the marginal mean accuracy and RT per secondary task condition, and used the mean accuracy to obtain an estimate of the maximum number of objects which can be held in mind.

3.2. Results

No trials were excluded for the accuracy analysis. For analysis of correct RTs, 8 trials (0.56% of data) were removed as motor errors or for being more than 3 SDs above the individual participant's mean. All reported results relate to confirmatory analysis.[1]

3.2.1. Accuracy

Bayesian model comparison showed very strong evidence for Step 2 over Step 1, indicating the data were BF_{10} = 2321.57 times more likely under a model that separated a no-task control from some form of secondary task. As predicted, there was also very strong evidence for Step 3 over Step 2, BF_{10} = 1556.20, meaning that the data favoured a model that distinguished between articulatory suppression and foot tapping as secondary tasks.

We then used the coefficients in the Step 3 model (Table 3) to estimate the mean marginal accuracy for each secondary task (see Figure 3). Critically, and as predicted, performance was worst in the articulatory suppression (reference) condition, with participants correctly remembering an average of 8.1 ± 0.6 (*M* ± *SE*) out of 12 objects per sequence. Performance was better during foot tapping, where participants were 89% more likely to respond correctly compared to articulatory suppression, and correctly recognized an average of 9.6 ± 0.5 (*M* ± *SE*) objects. Finally, participants performed best when there was no secondary task, and, with an average of 10.0 ± 0.4 (*M* ± *SE*) objects recognized per sequence, were 146% more likely (i.e., more than twice as likely) to respond correctly compared to when performing articulatory suppression. Overall, as expected, memory performance was impaired when access to language was suppressed (replicating Experiment 1 with a stronger effect), and suppressing access to language via articulatory suppression impaired performance more than a secondary task that was unrelated to language (i.e., foot tapping).

3.2.2. Response Times

Bayesian model comparison showed strong evidence *against* the Step 2 model over Step 1 (BF_{10} = 0.03); that is, the RT data were BF_{01} = 33.12 times more likely under a model containing only random intercepts than a model that distinguished secondary tasks from the no-task control condition. Similarly, there was strong evidence at Step 3 *against* distinguishing between foot tapping and articulatory suppression as secondary tasks, BF_{10} = 0.03, whereby the data were BF_{01} = 36.60 times more likely under the Step 2 model than the Step 3 model.

Nonetheless, we used the coefficients of the Step 3 model (Table 3) to estimate the marginal means for each secondary task condition (Figure 3). RT was similar in all conditions, and coefficients indicated no reliable differences. That is, against our predictions but consistent with the previous experiment, participants were equally fast to select the correct object in the object testing stage regardless of which secondary task (if any) was performed when learning the sequences.

3.3. Discussion

Experiment 2 determined whether or not the effects of articulatory suppression at sequence learning that were observed in the earlier experiment could be attributed to a dual-task load rather than to suppressing access to linguistic labels. The results replicated Experiment 1 in showing that suppressing access to language while learning a se-

1 As in Experiment 1, we attempted to explore different random effects structures in order to select the best-fitting model of accuracy and RT. However, none of the candidate structures involving random slopes improved model fit over the random intercepts of confirmatory analysis, and so we do not report them further. Full details are in supplementary materials.

Table 3. Experiment 2 unstandardized regression coefficients, standard errors, and associated statistics from Step 3 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for effects of secondary task at learning with articulatory suppression as the reference level.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
Accuracy	Intercept	0.736	0.234	-	3.151	<.01
	No task (control)	0.900	0.140	-	6.452	<.001
	Foot tapping	0.637	0.135	-	4.732	<.001
					t	
RT	Intercept	2439.76	87.60	29.55	27.850	<.001
	No task (control)	10.40	48.99	1314.24	0.212	.832
	Foot tapping	-17.89	49.57	1312.81	-0.361	.718

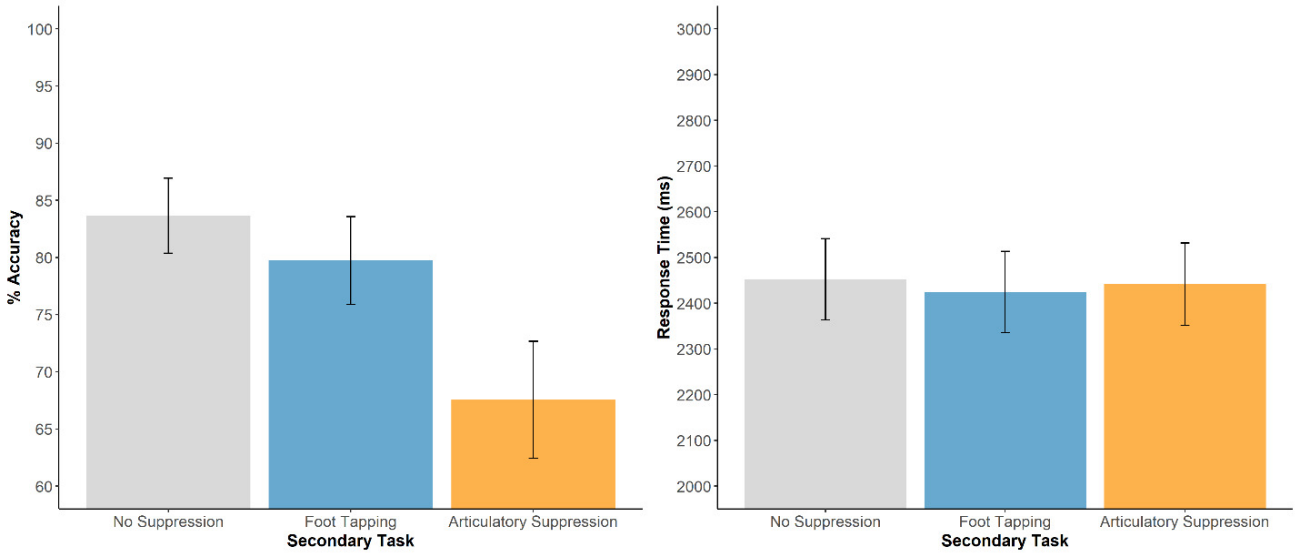

Figure 3. Mean % accuracy and RT per secondary task condition in Experiment 2, calculated as marginal means at Step 3 models. Error bars represent ± 1 Standard Error.

quence of objects impaired accuracy – but not speed – of memory performance. Critically, comparison with a foot-tapping task indicated that this effect was not a mere artifact of a dual-task paradigm. Rather, memory was adversely affected specifically by articulatory suppression, supporting our hypothesis that holding object concepts in mind normally relies on language (i.e., implicitly-activated object labels, used as linguistic placeholders), and that suppressing access to language reduces accuracy of performance and the number of objects which can be represented simultaneously. In other words, access to language enables linguistic bootstrapping, whereby people can use linguistic labels as placeholders for complex sensorimotor representations to increase the maximum number of objects they can hold in mind, and can hence remember a greater number of objects when language is available compared to when it is not.

However, performing a secondary task at the point of learning had no effect on RT measured during testing, and there was also no difference in RT between the two secondary tasks of articulatory suppression and foot tapping. That is, objects which were successfully remembered were identified during testing with the same difficulty regardless of what concurrent task participants performed while learning them, consistent with the findings of Experiment 1. This lack of effects on RT suggests that regardless of whether an object is held in mind via its linguistic label, a sensorimotor simulation, or a combination of both, it is relatively easy to match this representation with a target on-screen.

In this experiment, we calculated that people can successfully remember on average 8.1 ± 0.6 objects when language is suppressed, and 10.0 ± 0.5 objects when language is available. This estimate was slightly smaller than the estimate in Experiment 1 and – given that the items and relevant task conditions were the same in both experiments – is most likely due to inter-participant variability. However, the within-participant manipulation of the learning conditions in the present experiment means that the effect of articulatory suppression on performance has a much more robust grade of evidence than in Experiment 1, and allows us to conclude that the presence of language allows an additional 2 object concepts (approximately 25% more) to be held in mind, which is consistent with the linguistic bootstrapping hypothesis.

Nonetheless, the possibility remains that in both Experiments 1 and 2, performance in the no-task control condition may have been subject to ceiling effects. For example, in the current experiment, 44% of participants reached 100% performance on at least one sequence in the no-task condition, and 67% of the time participants scored 10 or more on a sequence. The same concern did not apply to the articulatory suppression condition, where only 22% of participants reached 100% performance on at least one sequence, and 31% of the time scored 10 or more. It is therefore possible that at least some participants may be capable of remembering more than 12 objects when language is available to support the representation of multiple objects. Experiment 3 therefore addressed this possibility by using a range of sequence lengths to determine the maximum number of objects which can be held in mind when language was fully available.

4. Experiment 3: Maximum Sequence Length with Language Available

In our final study (pre-registration, data, analysis code, and full results are available as supplemental materials at https://osf.io/mwzfh), we wanted to establish the maximum sequence length of familiar, contextually-related object concepts which can be held in mind under optimal conditions; that is, the number of objects that can be remembered when language is fully available and linguistic placeholders may be used to their full extent, which might have been underestimated in Experiment 2 due to ceiling performance. Using a similar paradigm to previous experiments but with no secondary task, we asked participants to remember sequences that varied in length between 8 and 14 objects.

We hypothesised that performance accuracy would start to drop when the number of object concepts in a sequence reached the maximum limit for linguistic information. That is, participants would remember shorter sequences relatively easily because their representations (sensorimotor simulation and/or linguistic labels) fit within the limit of what participants can hold in mind. Even some longer sequences may still fit within this limit via the use of linguistic placeholders, where objects are represented via their labels only rather than via sensorimotor simulation. However, once the length of the sequence exceeds the maximum limit even for linguistic labels, participants will not be able to retain them all and accuracy will suffer. As sequence length increases, the tipping point at which accuracy starts to reliably drop would reflect the maximum number of object concepts which can be simultaneously represented via linguistic labels.

We also predicted that response times would slow down as the available representational resources are strained and linguistic placeholders are increasingly employed, so that participants would be slower to recognise remembered objects when the number of objects in a sequence exceeds the maximum number of object concepts that participants can hold in mind.

4.1. Method

4.1.1. Participants

Twenty native speakers of English took part in the study (17 female; mean age = 18.95; *SD* = 1.07). Data from one participant was replaced due to not being a native speaker of English. As before, the sample size was determined using sequential hypothesis testing with Bayes Factors. We stopped at the sample size $N = 20$ when the Step 4 models for accuracy cleared the specified threshold of evidence $BF_{10} > 5$ for five consecutive participants. (See Design & Analysis section for model details; full statistics are reported in Results).

4.1.2. Materials

Test items comprised 112 target objects, divided into 8 sequences of 14 items each. These sequences were based on materials from Experiment 2, which we extended from 12 to 14 items by adding two extra objects to each sequence. Four of the extra target objects, with their accompanying distractors, were taken from the ninth sequence of Experiment 2 unused in this experiment; two were objects and accompanying distractors used in Experiment 1 that were not used in Experiment 2, while 10 were new objects. Distractor items for the new target objects were selected using the same criteria as in Experiment 1. As before, the order of objects for new sequences was determined by 8 volunteers who did not take part in the study, and was established based on their mean rank. We then created subsets of items within each sequence using the first 8, 10, 12 or all 14 items, so that each sequence could be plausibly represented in different lengths while still being ecologically valid (e.g., the context situation of making a cake still applied regardless of whether strawberries or whipped cream were included in the sequence). This subsetting approach allowed us to compare sequences of different lengths without confounding context situation with sequence length.

Photographic images of new target and distractor objects were selected and edited as per Experiment 1, leading to a total of 784 object images: 112 target objects for presentation at learning, 112 target objects (different images) for presentation at test, and 560 distractor objects for presentation at test.

4.1.3. Procedure

The procedure was the same as in the no-task condition of Experiment 2. After a practice sequence of 8 items, participants completed all eight test sequences in a fixed order of increasing length (i.e., two sequences of 8 objects each, then two sequences of 10 objects each, and so on). We rotated sequences across length conditions so that across the experiment as a whole, each sequence was presented in each of its possible subsets (i.e., 8, 10, 12, and 14 objects) and therefore in different ordinal positions in the procedure.

4.1.4. Design and Analysis

We analysed accuracy (incorrect = 0, correct = 1) in a mixed-effects hierarchical logistic regression (binomial, logit link), and response times (RT) for correct responses in a mixed-effects hierarchical linear regression. In both analyses, we included participants and items (nested within sequence) as crossed random effects. Sequence length was included as a categorical fixed effect, coded using reverse Helmert coding to compare the effect of each sequence length with the mean of the previous (shorter) sequences, which resulted in 3 coded variables (10 vs. 8 objects, 12 vs. 8-10 objects, 14 vs. 8-10-12 objects). This coding method is suited to capturing nonlinear monotonic trends (e.g., a plateau followed by a fall) and allowed us to determine the tipping point at which accuracy dropped (or RT slowed down) due to the sequence length surpassing the maximum number of concepts participants could hold in mind.

Hierarchical regressions comprised the following steps: Step 1 entered random intercepts, Step 2 entered sequence length as 10 vs. 8 objects, Step 3 entered sequence length as 12 vs. 8-10 objects, and Step 4 entered sequence length as 14 vs. 8-10-12 objects. We ran Bayesian model comparisons between steps, with Bayes Factors (BF) calculated via BIC as in earlier experiments. Specifically, the first step to show an improvement in model fit represented the sequence length at which memory performance differed from that of shorter sequences. In addition, the parameter of the first sequence length variable to produce an accuracy effect allowed us to estimate the maximum number of objects in a sequence which could be held in mind when language is fully available to support their representations.

4.2. Results

No trials were excluded for the accuracy analysis. For analysis of correct RTs, 14 trials were removed as outliers more than 3 standard deviations from the individual participant's mean (total 0.96% of data removed). All reported results relate to confirmatory analysis, as the attempt to model random slopes of sequence length on items led to non-convergence in both accuracy and RT analysis (see supplementary materials).

4.2.1. Accuracy

Bayesian model comparison showed evidence *against* any effect of sequence length on accuracy while sequence length remained at 12 objects or fewer: Step 2 did not improve model fit over Step 1 (BF_{10} = 0.02, i.e., data favoured Step 1 at BF_{01} = 41.08), and Step 3 did not improve model fit over Step 2 (BF_{10} = 0.20, i.e., data favoured Step 2 at BF_{01} = 5.01). However, there was strong evidence for Step 4 over Step 3 (BF_{10} = 14.95), meaning that the data favoured a model that distinguished 14-object sequences from shorter sequences.

We then used the coefficients in the Step 4 model (Table 4) to estimate the marginal accuracy for each sequence length parameter (see Figure 4). When sequence length reached 14 objects, accuracy decreased compared to shorter sequences; that is, when participants were asked to remember a sequence of 14 objects, they were 76% more likely to make an error in responding than for sequences of 8-12 objects (where performance was relatively stable), suggesting that 14 objects exceeded the maximum number of objects participants could simultaneously represent. Participants successfully remembered an average of 11.9 ± 0.5 (*M* ± *SE*) out of 14 objects.

We noted that in the Step 4 model, the parameter for sequence length of 12 (vs. 8-10 objects) was also significant in NHST terms, reflecting a small drop in accuracy between 8-10 objects and 12 objects. However, since Bayesian model comparisons found evidence *against* the addition of this parameter in Step 3, it indicates that the data were more likely under a model that ignored the distinction between sequence length 12 and sequence lengths 8-10 than under a model that distinguished them. We therefore treat the NHST effect for sequence length 12 with caution, whereas both Bayesian model comparison and NHST coefficient statistics supported a tipping point at sequences of 14 objects. That is, sequences of 14 objects exceeded the maximum number of objects that could be remembered in a way that sequences of 12 objects did not robustly do, hence the drop in accuracy; we thus estimate the maximum limit to be approximately 11.9 object concepts when linguistic labels are available.

4.2.2. Response Times

Bayesian model comparison showed no effect of sequence length on RT: there was evidence *against* Step 2 over Step 1 (BF_{10} = 0.14, i.e., data favoured Step 1 at BF_{01} = 7.39), against Step 3 over Step 2 (BF_{10} = 0.05, i.e., data favoured Step 2 at BF_{01} = 20.09), and against Step 4 over Step 3 (BF_{10} = 0.03, i.e., data favoured Step 3 at BF_{01} = 33.12). Nonetheless, for the sake of complete reporting, we used the coefficients in the Step 4 model (Table 4) to estimate the marginal mean RT for each sequence length parameter (see Figure 4). RT was similar in all sequence length conditions. That is, against our expectations, participants were equally fast to recognize objects regardless of how many objects were being remembered.

4.3. Discussion

Experiment 3 aimed to establish the maximum sequence length of object concepts which could be successfully represented when language is fully available (i.e., with no articulatory suppression), using longer sequences of objects than previous experiments in this study. We found strong evidence that sequences of 14 objects caused a drop in accuracy relative to sequences of 8-12 objects. On average, participants successfully remembered 11.9 out of 14 objects, meaning that sequences of 8, 10, and even 12 objects could generally be held in mind without a reliable drop in accuracy. Sequences of 14 objects, however, *exceeded* the maximum limit of what people could concurrently represent, and hence accuracy dropped. That is, when language is available, participants can accurately remember a sequence of up to approximately 12 object concepts, which, according

Table 4. Experiment 3 unstandardized regression coefficients, standard errors, and associated statistics from Step 4 models of Accuracy (logistic mixed-effect regression) and RT (linear mixed-effect regression), for Helmert-coded effects of sequence length.

DV	Parameter	Coefficient	SE	*df*	*z*	*p*
Accuracy	Intercept	2.160	0.258	-	8.382	<.001
	10 vs. 8 objects	-0.060	0.241	-	-0.249	.804
	12 vs. 8-10 objects	-0.473	0.179	-	-2.635	.008
	14 vs. 8-10-12 objects	-0.563	0.151	-	-3.725	<.001
					t	
RT	Intercept	2271.09	76.43	24.68	29.715	<.001
	10 vs. 8 objects	100.73	53.23	1346.28	1.892	.059
	12 vs. 8-10 objects	54.51	43.11	1373.06	1.264	.206
	14 vs. 8-10-12 objects	13.82	39.06	1414.61	0.354	.724

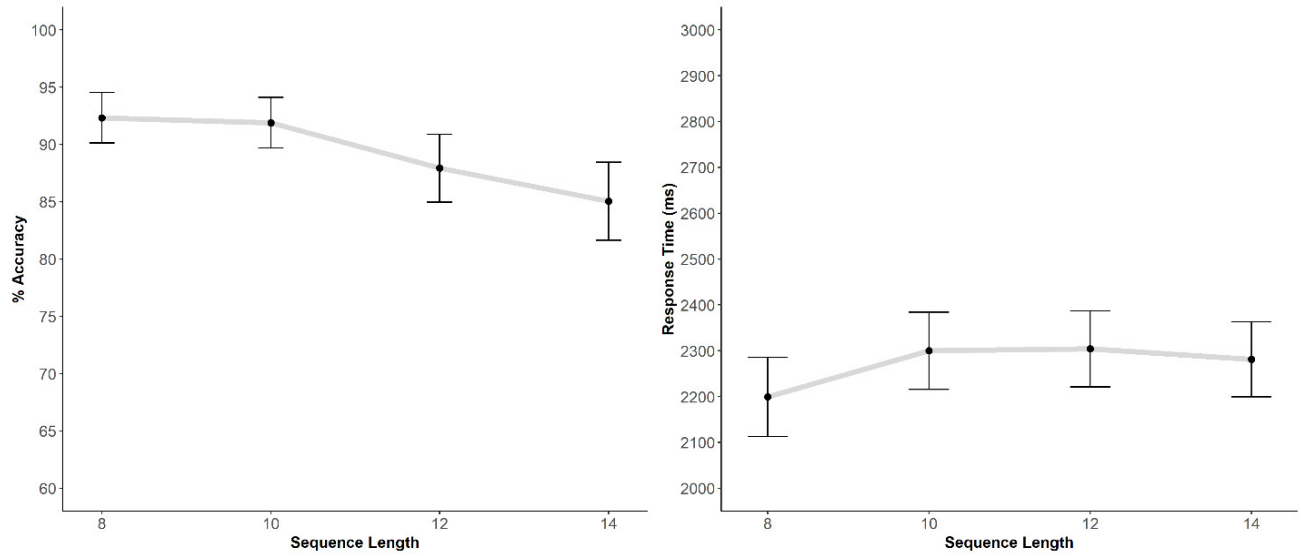

Figure 4. Mean % accuracy and RT for each sequence length condition in Experiment 3, based on marginal means from the Step 4 model. Error bars represent ± 1 Standard Error.

to the linguistic bootstrapping hypothesis, is possible because a linguistic label can serve as a placeholder for a full sensorimotor representation of an object when the available representational resources are under strain.

We also found that sequence length had no effect on RT, suggesting that the time required to match an object representation in memory to a visual stimulus onscreen was not influenced by demands of varying sequence length. The evidence against any effect on RT also suggested that the time required was not influenced by the format of object representation that participants held in mind; that is, participants could identify the onscreen target relatively quickly whether an object was represented via sensorimotor simulation (as would be possible for short sequences of 8 objects) or via linguistic labels (as was most likely for long sequences of 12-14 objects). We discuss the possible processes involved in the general discussion.

5. General Discussion

The present study is the first to examine the linguistic bootstrapping hypothesis; that is, whether word labels can act as placeholders for real-world object concepts when there are insufficient representational resources to maintain a sensorimotor simulation in full (Connell & Lynott, 2014), and thereby allow language to increase the number of contextually-situated, real-world object concepts that can be held in mind. We tested this hypothesis in a series of pre-registered experiments using a nonverbal task that asked participants to learn a naturalistic sequence of contextually-situated pictured objects and then tested their ability to select each previously-presented object from a distractor array. As predicted, we found that suppressing access to language via articulatory suppression when learning sequences resulted in poorer accuracy in performance and a lower limit to the sequence length which could be represented. Participants could remember 8 (Experiment 2) to 10 (Experiment 1) objects concepts when relying on

sensorimotor information only to learn the sequence (i.e., when language was suppressed), but the sequence length increased by 25% – approximately two items (Experiment 2) – to an upper limit of 12 objects (Experiment 3) when linguistic labels were available to act as placeholders and ease the strain on representational resources. Critically, this effect was not an artifact of dual task performance, as suppressing access to language via articulatory suppression impaired accuracy markedly more than an alternative secondary task of foot tapping that left access to language intact (Experiment 2). This pattern of findings overall supports the linguistic bootstrapping hypothesis that, even in nonverbal paradigms, the ability to remember multiple real-world object concepts normally relies on language (i.e., implicitly-retrieved object labels). When studying a long sequence of objects for later testing, people can drop the sensorimotor representations of the objects and allow the linguistic labels to deputise as placeholders, in order to maximise the number of objects that can be held in mind. Suppressing access to language during this process therefore results in fewer objects being remembered.

However, there was no comparable effect on RT, since participants were just as quick to select a target object from a distractor array regardless of whether or not language had been available while learning the object sequence (Experiments 1-2). There are two possibilities regarding what happens to sensorimotor representations of objects when the number of objects exceeds the available representational resources *and* language is unavailable to provide linguistic placeholders (i.e., the articulatory suppression condition). Since sensorimotor simulations are flexible and responsive to task demands (Connell & Lynott, 2014), we had originally expected that all sensorimotor representations held in mind would degrade to some extent (i.e., lose some detail, such as information from less-relevant perceptual modalities or action effectors) in an effort to maintain all objects. This possibility would have led both to greater errors and slower responses to the object arrays (compared to the no-task control that allowed linguistic placeholders), due to the difficulty of matching degraded object representations to target object pictures, but the lack of RT effects makes this possibility unlikely. The second possibility was that some object representations are dropped entirely, but the sensorimotor representations of the remaining objects retain their original quality of detail. This possibility would have led to greater errors during object testing (because some objects are no longer held in mind at all) but would elicit no effect on RT for the objects which were successfully selected (because sensorimotor objects still held in mind can be easily matched to target object pictures). Results in Experiments 1-2 followed this pattern and therefore suggest that, when the maximum limit of sequence length for object concepts is reached, sensorimotor representations of individual objects are lost rather than maintained in some degraded form.

In addition, articulatory suppression during the object testing stage unexpectedly led to *faster* RT but no difference in errors (Experiment 1). As discussed in Experiment 1, this RT facilitation may have resulted from time saved by not implicitly labelling the object array, but other phenomena are of course possible. Regardless, this pattern of findings does not follow our original predictions, and suggests that memory for real-world object concepts is flexible and robust enough that it can survive losing access to language between learning and testing stages. If this supposition is correct, it could also explain the absence of interaction between articulatory suppression at learning and test. When language is suppressed during sequence learning, and an object concept is represented via sensorimotor simulation alone, then during later object testing, participants have two options: they can either directly compare their sensorimotor representation to what they see onscreen, or they can implicitly label the representation and the target stimulus and compare the two labels. On the other hand, when a concept is represented via a linguistic label alone, then in the later testing stage participants have the same two options: they can implicitly label the object onscreen and compare the two labels, or they can retrieve a sensorimotor representation of the label's referent object and then compare that to what they see onscreen. In both cases, suppressing access to language during the object testing process leaves the sensorimotor option available, and so there is no interaction between the format of object representation held in mind and the availability of language during the testing stage. In other words, while the pattern of findings may still be consistent with the idea that learning nonverbal object concepts normally relies on language (i.e., implicitly-retrieved object labels), we conclude that our original conception of the effects of linguistic bootstrapping on the test stage should be updated. Rather than automatically maintaining the format of object representation that was held in mind during sequence learning, linguistic bootstrapping allows people to flexibly adapt the contents of the representation during object testing by accessing linguistic or sensorimotor aspects of a concept's representation according to available resources and demands of the task.

Finally, we estimate that people are able to remember 12 contextually-situated object concepts when language is fully available to bootstrap cognition, and 10 when it is not available (although absolute estimates varied a little between experiments). It is possible that the sequence length limits for concepts other than objects may differ from our current findings. For instance, recent work in our lab has found that the ability to remember sequences of action events is limited to approximately three events (Banks & Connell, 2022), and the limit is even lower when language is suppressed, which suggests that sensorimotor representations of events are larger (i.e., take up more representational "space") than sensorimotor representations of objects. Future work should examine in more detail how the maximum number of concepts which can be held in mind may differ according to the nature of the concept or entity in question. Moreover, given that linguistic bootstrapping should be useful in other circumstances where there are insufficient representational resources to maintain a sensorimotor simulation in full, such as when online processing demands perceptual attention (Banks & Connell, 2022),

future work should also explore its impact on other domains of cognition. In the present paper, we used naturalistic, contextually-situated sequences of real-world object concepts to examine the importance of language in object representation and to determine how many object concepts can be simultaneously held in mind. As predicted by the linguistic bootstrapping hypothesis, we found that language increases the number of familiar objects that people can represent at one time from approximately 10 to 12 object concepts. In other words, by implicitly retrieving object labels and using them as linguistic placeholders (that can be fleshed out again to sensorimotor representations when required), the use of language allows people to remember a list of up to 12 items, for example a shopping list or the ingredients for a recipe. These findings are in line with linguistic-simulation theories of cognition that hold language to be a critical part of how people represent, remember, and use their knowledge about concepts. We hope they contribute to a more comprehensive understanding of the role of language in human cognition.

Conflict of Interest

The authors have no competing interests to declare.

Author Contributions

Contributed to conception and design: LC, AD
Contributed to acquisition of data: AD
Contributed to analysis and interpretation of data: AD, LC, BB
Drafted and revised the article: AD, LC, BB

Funding

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682848) to LC.

Data Accessibility Statement

All images, code, and data used in this article are licensed under a Creative Commons Attribution 4.0 International License (CC-BY), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, so long as you give appropriate credit to the original authors and source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of the license, visit http://creativecommons.org/licenses/by/4.0/. Pre-registration, stimuli, data and analysis code can be found at: https://osf.io/mwzfh

Submitted: January 27, 2022 PST, Accepted: October 31, 2022 PST

References

Allen, R. J., Havelka, J., Falcon, T., Evans, S., & Darling, S. (2015). Modality specificity and integration in working memory: Insights from visuospatial bootstrapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 820–830. https://doi.org/10.1037/xlm0000058

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233–252. https://doi.org/10.1080/14640748408402157

Banks, B., & Connell, L. (2022). *Language enhances working memory for action events* [Manuscript in preparation]. Department of Psychology, Lancaster University.

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic distributional knowledge and sensorimotor grounding both contribute to semantic category production. *Cognitive Science*, *45*(10), e13055. https://doi.org/10.1111/cogs.13055

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/s0140525x99002149

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, & A. C. Graesser (Eds.), *Symbols, Embodiment, and Meaning* (pp. 245–284). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.003.0013

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using**lme4**. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bidet-Ildei, C., Meugnot, A., Beauprez, S.-A., Gimenes, M., & Toussaint, L. (2017). Short-term upper limb immobilization affects action-word understanding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1129–1139. https://doi.org/10.1037/xlm0000373

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*, *29*, 120–153. https://doi.org/10.1016/j.plrev.2018.12.001

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756. https://doi.org/10.1016/j.neuropsychologia.2007.10.007

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 157–165. https://doi.org/10.1037/0278-7393.18.1.157

Connell, L. (2018). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1308–1318. 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), e33321. https://doi.org/10.1371/journal.pone.0033321

Davis, C. P., Joergensen, G. H., Boddy, P., Dowling, C., & Yee, E. (2020). Making it harder to "see" meaning: The more you see something, the more its conceptual representation is susceptible to visual interference. *Psychological Science*, *31*(5), 505–517. https://doi.org/10.1177/0956797620910748

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*, *72*(5), 1112–1118. https://doi.org/10.1177/1747021818785096

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia*, *51*(8), 1510–1517. https://doi.org/10.1016/j.neuropsychologia.2013.04.008

Forsberg, A., Johnson, W., & Logie, R. H. (2020). Cognitive aging and verbal labeling in continuous visual memory. *Memory & Cognition*, *48*(7), 1196–1213. https://doi.org/10.3758/s13421-020-01043-3

Gaillard, V., Destrebecqz, A., & Cleeremans, A. (2012). The influence of articulatory suppression on the control of implicit sequence knowledge. *Frontiers in Human Neuroscience*, *6*(208), 1–9. https://doi.org/10.3389/fnhum.2012.00208

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience*, *26*(18), 4917–4921. https://doi.org/10.1523/jneurosci.5389-05.2006

Goodhew, S. C., McGaw, B., & Kidd, E. (2014). Why is the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review*, *21*(5), 1287–1293. https://doi.org/10.3758/s13423-014-0593-6

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/s0896-6273(03)00838-9

Jaroslawska, A. J., Gathercole, S. E., & Holmes, J. (2018). Following instructions in a dual-task paradigm: Evidence for a temporary motor store in working memory. *Quarterly Journal of Experimental Psychology*, *71*(11), 2439–2449. https://doi.org/10.1177/1747021817743492

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

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. 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., & 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. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. deVega, A. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning: A debate* (pp. 309–326). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.003.0015

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

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*, *3*(54). https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*(1), 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*(3), 263–276. https://doi.org/10.1037/h0082978

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499–508. https://doi.org/10.1037/xhp0000313

Paivio, A. (1971). Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches* (pp. 7–32). Academic Press. https://doi.org/10.1016/b978-0-12-635450-8.50008-x

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*(10), 1–8. https://doi.org/10.3389/neuro.11.010.2008

Phillips, L. H., Wynn, V. E., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory*, *7*(2), 209–231. https://doi.org/10.1080/741944066

Richardson, J. T. E., & Baddeley, A. D. (1975). The effect of articulatory suppression in free recall. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 623–629. https://doi.org/10.1016/s0022-5371(75)80049-1

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345. https://doi.org/10.1111/j.1756-8765.2010.01111.x

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3*(1), 83–119. https://doi.org/10.1515/langcog.2011.004

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, *49*(1), 222–231. https://doi.org/10.1016/j.cortex.2011.10.005

Souza, A. S., Overkott, C., & Matyja, M. (2021). Categorical distinctiveness constrains the labelling benefit in visual working memory. *Journal of Memory and Language*, *119*, 104242. https://doi.org/10.1016/j.jml.2021.104242

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

Van 't Wout, F., & Jarrold, C. (2020). The role of language in novel task learning. *Cognition*, *194*(104036). https://doi.org/10.1016/j.cognition.2019.104036

Vermeulen, N., Chang, B., Mermillod, M., Pleyers, G., & Corneille, O. (2013). Memory for words representing modal concepts. *Experimental Psychology*, *60*(4), 293–301. https://doi.org/10.1027/1618-3169/a000199

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. https://doi.org/10.1515/langcog.2009.011

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). MIT Press. (Original work published 1934)

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/bf03194105

Weywadt, C. R. B., & Butler, K. M. (2013). The role of verbal short-term memory in task selection: How articulatory suppression influences task choice in voluntary task switching. *Psychonomic Bulletin & Review*, *20*(2), 334–340. https://doi.org/10.3758/s13423-012-0349-0

Wingfield, C., & Connell, L. (2022). Understanding the role of linguistic distributional knowledge in cognition. *Language, Cognition and Neuroscience*, *145*, 1–51. https://doi.org/10.1080/23273798.2022.2069278

Wood, E. A., Rovetti, J., & Russo, F. A. (2020). Vocal-motor interference eliminates the memory advantage for vocal melodies. *Brain and Cognition*, *145*, 105622. https://doi.org/10.1016/j.bandc.2020.105622

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). 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