

**Learning in complex, multi-component cognitive systems: Different  
learning challenges within the same system**

Bonnie L. Breining<sup>1,2</sup>, Nazbanou Nozari<sup>1,2</sup>, & Brenda Rapp<sup>1</sup>

Johns Hopkins University

<sup>1</sup>Department of Cognitive Science, <sup>2</sup>Department of Neurology

Correspondence concerning this article should be addressed to:

Bonnie Breining, Department of Neurology, Johns Hopkins University School  
of Medicine, 600 N. Wolfe St., Baltimore, MD 21287; Email: breining@jhu.edu;

Telephone: 410-502-6045

### **Abstract**

Using word learning as an example of a complex system, we investigated how differences in the structure of the subcomponents in which learning occurs can have significant consequences for the challenge of integrating new information within such systems. Learning a new word involves integrating information into the two key stages/subcomponents of processing within the word production system. In the first stage, multiple semantic features are mapped onto a single word. Conversely, in the second stage, a single word is mapped onto multiple segmental features. We tested whether the unitary goal of word learning leads to different local outcomes in these two stages because of their reversed mapping patterns. Neurotypical individuals (N=17) learned names and semantic features for pictures of unfamiliar objects presented in semantically-related, segmentally-related and unrelated blocks. Both similarity types interfered with word learning. However, feature learning was differentially affected within the two main

subcomponents of word production. Semantic similarity facilitated learning *distinctive* semantic features (i.e., features unique to each item), whereas segmental similarity facilitated learning *shared* segmental features (i.e., features common to several items in a block). These results are compatible with a model of incremental learning in which learning not only strengthens certain associations but also weakens others according to the local goals of each subcomponent. More generally, they demonstrate that the same overall learning goal can lead to opposite learning outcomes in the subcomponents of a complex system. The general principles uncovered here can be extended beyond word learning to other complex systems with multiple subcomponents.

**Keywords:** word learning; incremental learning; blocked cyclic naming; semantic similarity; segmental similarity

Learning in complex, multi-component cognitive systems: Different learning challenges within the same system

### **Introduction**

Learning within complex cognitive systems often requires changes in multiple subcomponents that all face the challenge of integrating new information with what is already known. This challenge can take different forms depending on the structure of the system in which learning is happening.

One example of a complex cognitive operation involving multiple subcomponents is word production. Theories of word production posit two key stages: Word Selection and Segmental Encoding (Dell, 1986; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000). Based on the meaning we wish to communicate ([furry], [meows], [pet]), we select the word to be produced (CAT) and the segments that make up the word (*c-a-t*) (Figure 1). While learning new words often initially includes perception (i.e., hearing the

new word), ultimately the word can only be produced if it is integrated into the production system. This means that its relevant set of semantic features (accessible from sensory input or generated internally) must be connected to its abstract word form, and that word form must, in turn, be connected to its segments (i.e., phonemes or graphemes). Although the learned representations may be accessible from other systems (e.g., perception and memory), they reflect the dynamics of the production system into which they have been integrated. This paper investigates the influence of these dynamics within the two stages of production that entail opposite mapping patterns: Word Selection involves *many-to-one mapping* (*many* meaning features map onto *one* word) while Segmental Encoding involves *one-to-many mapping* (*one* word maps onto *many* segments).

We use similarity in semantic and segmental features to study the dynamics of learning new labels for novel objects at the first and second stages of production, respectively. We show that, while both types of similarity produce general difficulties for word learning, they have very different and

specific consequences for *feature learning* at the two stages of word production, precisely because of the unique learning problems at the two stages. While this research specifically involves word learning, the findings are relevant more broadly to learning in any complex multi-component system.

### **Similarity and learning in word production**

Although word production is a complex system that requires balancing the competing forces of interference and facilitation due to a variety of factors including lexical competition, repetition priming, and strategy use (e.g., Belke, Shao, & Meyer, 2017), there is substantial evidence that the naming of multiple items that share semantic features (e.g., CAT, DOG, HORSE are all animals and have four legs) predominantly creates interference in word production (e.g., Belke, Meyer, & Damian, 2005; Damian, Vigliocco, & Levelt, 2001). This interference survives the insertion of several unrelated pictures between related items (see Schnur, 2014). For this reason, the interference is proposed to result from long-lasting incremental learning as opposed to

short-term activation-based mechanisms that are subject to rapid decay

(Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, &

Schwartz, 2010).

According to Oppenheim and colleagues' 2010 interactive activation model of word production, each naming attempt is actually a learning experience in that the connections between the target word and its semantic features are strengthened. This strengthening facilitates future production of the same item. However, in this model, learning does not solely strengthen connections: it also weakens the connections between competing words and the features they share with the just-produced word. For example, upon naming a picture of a cat, the connection between CAT and [furry] is strengthened, but the connection between DOG and the same feature is weakened. If cat is to be named again, this will facilitate its selection amongst related competitors. However, if dog becomes the next target, it will be harder to access than it would have been if the previous target had been an unrelated word. Consequently, across a set of related items, the net effect is

typically interference. We recently showed that naming pictures of multiple items that share segmental features (e.g., CAT, COT, MAT) also creates similar patterns of interference during picture naming (Breining, Nozari, & Rapp, 2016; Nozari, Freund, Breining, Rapp, & Gordon, 2016).

### **Consequences of multi-stage learning**

Integrating a new word into one's production system requires connecting two sets of features, meaning (semantic) to form (segmental) features, via an intermediary (word) representation<sup>1</sup>. Here we consider –for the first time– the predictions of an incremental learning mechanism for learning these two types of features, as part of learning a new label for a novel object. A *shared* feature is one that is common to multiple items in the learning set (see Figure 1; e.g., the concept [furry] in a set containing both CAT and DOG or the letter *a* in a set with both CAT and MAT). A *distinctive* feature, on the

---

<sup>1</sup> Note that all models of word production contain an intermediate layer between semantics and segments, even if this layer does not explicitly consist of word units. The dynamics discussed in this paper depend on the existence of some intermediate layer and can, therefore, be expected in all models of production with feedback between segments and the intermediate layer.



other hand, is unique to a single item within the learning set (e.g., the concept [meows] and the letter *c* which only pertain to CAT).

As a mechanism of incremental learning, we adopt the principles proposed by Oppenheim and colleagues (2010), because the model has been successful at explaining a range of findings regarding semantic interference (the computational simulations have been reported in detail in the original paper for the interested reader). An alternative model (Howard et al., 2006) also predicts interference during the production of semantically-related words, but unlike Oppenheim and colleagues (2010), this interference arises from a combination of strengthening the connections between the target words and semantic features (similar to Oppenheim et al., 2010) and lateral inhibition between words (different from Oppenheim et al). Importantly, unlike Oppenheim and colleagues' model it has no mechanism for weakening the connections between the competitor and the shared feature. We later discuss the difference between these two learning mechanisms in light of their consequences.

The scope of both of the above models of incremental learning, however, is limited to the first stage of production. To extend the learning principles of Oppenheim and colleagues (2010) to the second stage of processing, we use the framework of the two-step interactive model (e.g., Foygel & Dell, 2000; Rapp & Goldrick, 2000). Unlike the alternative feedforward models (e.g., Levelt et al., 1999; Roelofs, 1997), the interactive models assume feedback from segments to the word layer. We return to this issue in the General Discussion and compare our findings with the predictions of a feedforward framework.

Figure 2 illustrates the learning of two semantically-related items, each with one shared and one distinctive feature. What is shown here is precisely what is predicted –and simulated-- by Oppenheim et al.'s (2010) model. On trial 1, the first item is introduced (A) and the connections between its semantic features and label are formed (B). When item 2 is presented on trial 2, its semantic features are activated along with its label. However, item 1's label is also partially activated via the shared semantic feature (C). At the

trial's end, connections that supported target item 2's activation are strengthened while those which supported competitor item 1's activation are weakened (D). When item 1 is presented again on trial 3, it is at a disadvantage because of its weakened connection to the shared feature. However, similar processes of activation of new target item 1 and partial activation of competitor item 2 take place (E). At the end of this trial, connections are adjusted to support the current target (item 1) and make item 2's selection less likely in the future (F). Over time, this learning process consistently increases the strength of the connections between each word and its distinctive feature. However, the connection between each word and its shared features is strengthened or weakened depending on its status as target or competitor on each trial (G). This dynamic learning process predicts that distinctive features of semantically-related items should be learned better than their shared features<sup>2</sup>. That is, *semantic similarity during learning should facilitate the retrieval of distinctive vs. shared features.*

---

<sup>2</sup> While in this paper we describe the system as though there is no feedback between the

Figure 3 illustrates the learning of two segmentally-related items, each with one shared and one distinctive segmental feature. What is shown here is our extension of the incremental learning mechanism proposed by Oppenheim et al. (2010) to the second stage of production in an interactive system. After exposure to the first item (A), the connections between the word and its segmental features are formed (B). Presentation of item 2 on trial 2 activates its features. However, item 1 is also partially activated by feedback through the shared segmental feature and in turn it activates its own distinctive segment (C). At the end of trial 2, connections that supported the activation of current target segments are strengthened, while those that supported the activation of the non-target segment are weakened (D). When

---

word and semantic levels, we note that such feedback is possible. However, this would not change the effects of semantic similarity. If there is word-semantic feedback leading to activation of non-target, semantically-related items, the distinctive feature of the non-target item will undergoes slight weakening. This weakening will always be smaller than the shared feature, because weight changes are proportional to the activation of the features, and the shared feature is always more activated than the distinctive feature of any non-target items that were activated through feedback only. So the final outcome is still an advantage for the distinctive over the shared feature. For simplicity (and consistency with the implemented model of Oppenheim et al., 2010), the architecture we describe does not include such feedback.

item 1 is presented again on trial 3, it is at a disadvantage because of its weakened connection to its distinctive segment. However, again similar processes of activation of the new target item 1 and partial activation of competitor item 2 take place (E). At the trial's end, connections are adjusted to support the activation of item 1's segments and make the selection of item 2's segments less likely in the future (F). Over time, this learning process consistently increases the strength of the connections between each word and the shared segmental feature. On the other hand, a word's connection to its distinctive feature is strengthened and weakened depending on the word's status as target or competitor on each trial (G). This learning process predicts that shared features of segmentally-related items should be learned better than their distinctive features. In other words, *segmental similarity during learning should facilitate the retrieval of shared vs. distinctive features.*

Thus, a key prediction is thus that the demands of information integration within different subcomponents of the word production system should lead to different consequences for the learning of semantic vs.

segmental information about new words. In sum, the extension of incremental learning principles to new word learning predicts: (1) overall interference (i.e., reduced accuracy and/or longer response latencies) for naming words trained in semantically- or segmentally-related blocks compared to those trained in unrelated blocks; (2) an advantage for learning distinctive vs. shared semantic features for words trained in semantically-related blocks; and (3) the reverse pattern with an advantage for the learning of shared vs. distinctive segments for words trained in segmentally-related blocks. To be clear, the mechanisms we have proposed for the integration of new words into the production system predict greater difficulty for naming the words learned in both semantically-related and segmentally-related conditions. However, the finding of an interference effect in naming cannot shed light on the precise nature of this difficulty. To investigate the stage-specific consequences of incremental learning discussed earlier in this paper, we propose to test the differential effects of shared vs. distinct features in semantically- and segmentally-related conditions.

We tested these predictions in a 4-session training study in which neurotypical participants were taught names and features for pictures of unfamiliar objects in semantically-related, segmentally-related, and unrelated training blocks. We have previously shown that semantic and segmental similarity have similar effects for speaking and writing words (Breining et al., 2016; Breining & Rapp, 2017). For this study, we chose the written modality because it is more straightforward to conceptualize and manipulate feature overlap for written forms since it removes the factors of accent, co-articulation, and intra-syllable position that complicate determining spoken feature overlap.

## **Methods**

### **Participants**

Seventeen right-handed, neurotypical adult participants aged 18-25 years (mean age 20.4 years, 13 females) were recruited from the Johns Hopkins community. Each participant gave informed written consent

according to the policies of the local institutional review board and received \$60 upon completion of the final experimental session.

## **Stimuli**

Participants were trained on a total of 24 novel items. Figure 4 shows an example stimulus; the complete set is included in Supplemental Material. For each participant, there were six blocks (i.e., lists) each consisting of four items (two semantically-related, two segmentally-related, and two unrelated blocks). Across participants, each item appeared in only one type of block and consisted of a pseudoword label paired with a 4 by 4 inch black and white line drawing of a very unusual object taken from the Ancient Farming Equipment stimuli (e.g., Laine & Salmelin, 2010), the NOUN database (Horst & Hout, 2016), clip art directories, and images freely available online.

With regard to semantic features, for each item, two sensory features and two functional features were provided which were not extractable from the visual properties of the object. With regard to segmental properties, each



item name was a monosyllabic 4-letter pseudoword that was orthotactically and phonotactically plausible in English. All six blocks were matched on length in letters and phonemes of the pseudowords, and on phonological and orthographic neighborhood density from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). The same picture was paired with the same semantic features and segments across all participants.

**Semantic blocks.** In the blocks made up of semantically-related items, semantic feature overlap was distributed unpredictably across the four items such that each item had two features that were shared with at least one other item in the block and two features that were unique to the item (i.e., each item had two shared and two distinctive features). The same feature was never shared by all four items in a block so that participants could not infer the presence of a certain feature from block identity. Within each semantic block, there was no segmental overlap, meaning each segment appeared only once in any position (i.e., within a semantic block all 16 letters were unique).

However, because there is a limited set of segments in English, the same segments were necessarily used in other blocks.

**Segmental blocks.** In the blocks made up of segmentally-related items, distribution of segmental feature overlap was similar to the semantic blocks such that each item had two segments that were shared with at least one other item in the block in the same position in the word and two segments that were unique to the item (i.e., each item had two shared and two distinctive segments). Each segment in a segmental block appeared three times across the semantic and unrelated blocks. Application of these constraints resulted in one segment serving as a shared segment in both segmental blocks. To encourage lexical processing, the relationship between orthography and phonology was not always consistent (e.g., [ɪ] corresponded to *y* in *chys* but to *i* in *lisk*). Semantic features were not repeated across segmental blocks.

**Unrelated blocks.** Within each unrelated block, there was no repetition of semantic features or segments. As in the semantic blocks, there was

repetition of segments across blocks due to the limited set of segments in English. As in the segmental blocks, there was no repetition of semantic features across blocks.

### **Procedure**

The experiment was run using E-Prime 2 Professional (Psychology Software Tools, Pittsburgh, PA) on a Dell Latitude E6500 laptop with a 13-inch by 8-inch screen. Participants attended four training sessions within 10 days. The general procedures for familiarization and training were based on previous research using the Ancient Farming Equipment paradigm, modified to compare different blocking contexts.

First, participants were familiarized with the twenty-four pictures. For each item, the picture was presented along with a list of its four semantic features printed in Arial size 18 font to the right of the picture. The name appeared underneath in Arial size 24 font and was also auditorily presented.

Participants had a maximum of 15 seconds to process the information about each item.

After the familiarization phase, participants were instructed on the use of the Wacom Bamboo tablet for writing responses, with pen strokes appearing on the computer monitor. Response time (RT) was recorded when the writing surface was first touched with the pen. Participants were instructed to write the letters they knew and draw blanks for unknown letters (e.g., for a 4-letter name that started with c and ended in s, they could write c\_ \_ s). They were to write as legibly as possible, but were not restricted to a particular handwriting style (they were free to write in upper or lower case, print or cursive). After writing a response, participants returned the pen to the starting point and pressed a button with their non-dominant hand to advance to the next trial. When this button was pressed, a screen shot of the completed response was saved for scoring accuracy.

Training followed a test-study-test format consisting of five parts (Figure 5), providing for multiple retrieval attempts, known to enhance

learning (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The first phase tested retrieval from memory of the word form and semantic features, followed by a study phase for strengthening this information and then another test phase. All five parts of training were performed sequentially for each item before moving on to the following item.

Part 1 tested memory of the word form, requiring participants to attempt retrieval. Following a 500-millisecond fixation and a 500-millisecond "Prepare to Name" screen, the picture appeared along with the "Name" instruction, and remained there until participants started to write down their response. Once finished, they pressed a button continue. Part 2 tested memory of semantic features. After viewing "Prepare to Verify" for 500 milliseconds, they saw the picture with one semantic feature printed underneath. Participants then pressed a YES or NO button to indicate whether or not the feature belonged to the item and move onto the next part. Part 3 allowed study of information regarding the word form and semantic features. After a 500-millisecond "Study" screen, the picture, its four semantic features,

and its name all appeared on the screen and participants were instructed to copy the name underneath the printed label. Unlike the "Name" and "Verify" parts where participants were encouraged to respond as quickly as possible, there was no time pressure during this part. After that, participants again completed "Name" and "Verify" parts (with a different feature) to further test their memory. Then the trial was complete.

After five practice trials with familiar objects and labels during the first session, participants completed 72 training trials during each of the four separate training sessions. Each day, items were presented in segmental, semantic, and unrelated blocks in a new pseudorandom order, so that each block type was not repeated before the other two block types were presented. The six blocks (2 semantic, 2 segmental, and 2 unrelated) were separated by short breaks. Within each block, all four items were presented over three cycles in random order for a total of 12 training trials per block (72 trials total across the six blocks). For the verify portions of the trials, all incorrect features (50% of feature verification trials across the four sessions)

consisted of features from other items in the same block. Each shared feature appeared more times than each distinctive feature since shared features were correct for multiple items.

At the end of each training session, two probe tasks were administered to assess the speed and accuracy of semantic and segmental feature retrieval for shared and distinctive features. The semantic feature probe task was administered first. This type of task has been used previously to investigate the organization of semantic knowledge (e.g., Cree, McNorgan, & McRae, 2006). The task began with a 500-millisecond fixation cross, followed by the presentation of the picture along with one printed semantic feature.

Participants had 2000 milliseconds to indicate whether the feature belonged to the item or not with YES or NO buttons. A 500-millisecond inter-trial interval separated trials. There were 192 trials total, with four YES trials for each item that paired the picture with a correct feature and four NO trials for each item that paired the picture with an incorrect feature from the same block.

A segment probe task was also administered. This type of task has been used previously to evaluate the activation of orthographic representations (e.g., Rapp & Lipka, 2011). The set-up of the segment probe task was the same as the semantic probe task, except that a single upper-case letter was presented in Arial size 18 font in the place of the semantic feature, and participants were to decide whether or not the name of the item contained that letter. As in the semantic task, there were 192 trials, half YES and half NO.

## Results

All analyses were performed using multilevel mixed models with random effects in R version 3.2.4 with the lme4 and lmerTest packages. Accuracy data were analyzed using logistic regression since they are binary, while log-transformed response time data were analyzed using linear regression. Since the predictions consider the impact of a specific type of blocking (semantic or segmental) relative to an unrelated context, separate



models were constructed to compare items trained in semantic vs. unrelated blocks and in segmental vs. unrelated blocks, while items trained in semantic and segmental blocks were not directly compared. We first report the results of the naming trials obtained during training. Recall that we predict interference will result from both semantic and segmental similarity. Next, we report the results of the probe tasks, which allow us to test the stage-specific challenges of integration.

### **Naming Results**

Analyses considered only the accuracy and response time of the first naming attempt made in each training trial. Participants performed at ceiling on the other naming portions of the trials, with 99.8% whole response accuracy on written copy and 99.6% accuracy on second naming attempt. Overall, across all training sessions, participants correctly produced the whole response on 77.5% of first naming attempts. The majority of errors (66.8%) were omissions in which no segments were produced. There were a few

within-block substitutions (5.1%) and across-block substitutions (3.7%)

whereby another name from the experiment was produced instead of the target. Remaining errors consisted of additions, deletions, and substitutions of segments (24.3%). Accuracy analyses examined whole response accuracy, not segment accuracy.

Naming accuracy models included the following fixed effects: block type (semantic vs. unrelated in the semantic model and segmental vs. unrelated in the segmental model), training attempt within session (1-3), training session (1-4), two- and three-way interactions between those, and the control variables of days since last training session and number of training trials since the target was last trained. Continuous variables were centered and scaled. A full random structure was implemented in each model, with random intercepts for subjects and items, a full random slope structure matching the fixed effect structure over subjects, and the same random slope structure over items with the exception that block type and its interactions were excluded since each item was trained in only one context.

Response times (RTs) entered into analyses were log-transformed and excluded incorrect responses and outliers more extreme than 2.5 standard deviations from each participant's raw mean RT regardless of accuracy (23.0% of total trials). The model architecture was the same as in the accuracy analysis. Figure 6 shows the results of the naming responses from the training task, and Tables 1 and 2 show the results of the models of naming accuracy and response time data for semantic and segmental blocks, respectively, compared to unrelated blocks.

Both models revealed robust evidence of learning: Participants' accuracy increased over sessions ( $z=7.43$ ,  $p<.001$  for the semantic model;  $z=8.00$ ,  $p<.001$  for the segmental model) and their RTs decreased as they completed more sessions ( $t=-16.24$ ,  $p<.001$  for the semantic model;  $t=-15.94$ ,  $p<.001$  for the segmental model). Furthermore, there were also consistent main effects of training attempt within session such that accuracy increased ( $z=6.11$ ,  $p<.001$  for the semantic model;  $z=5.10$ ,  $p<.001$  for the segmental model) and RT decreased ( $t=-6.53$ ,  $p<.001$  for the semantic model;  $t=-5.91$ ,

$p < .001$  for the segmental model) as participants practiced naming the same item multiple times within a session, again demonstrating learning.

Critically, several pieces of evidence also indicated interference generated by similarity during training. In the semantic model, participants were significantly less accurate in the semantic vs. unrelated blocks ( $z = -2.08$ ,  $p = .038$ ). They also had smaller increases in accuracy for items trained in semantic vs. unrelated blocks across training attempts within session ( $z = -2.21$ ,  $p = .027$ ) and across sessions ( $z = -1.96$ ,  $p = .050$ ). Although the RT effects for semantic vs. unrelated blocks did not reach significance, their pattern was generally consistent with the interference observed in the accuracy data. In the segmental model, participants had increasingly longer RTs across training attempts within session in the segmental vs. unrelated blocks ( $t = 3.20$ ,  $p = .003$ ). In addition, they were marginally less accurate in the segmental vs. unrelated blocks overall ( $z = -1.81$ ,  $p = .070$ ). See Tables 1 and 2 for the complete list of effects.

In summary, similarity contexts had negative effects during learning. While participants did learn the names of the items over the course of the experiment, naming was less accurate and improvement was slower within sessions in both semantically- and segmentally-related contexts.

### **Probe Task Results**

On the semantic probe task, over all administrations, participants correctly accepted features on 84.1% of trials (low of 77.0% on session 1 to high of 88.5% on session 4) and correctly rejected features on 83.6% of trials (low of 81.4% on session 1 to high of 86.7% on session 4). On the segmental letter probe task, they correctly accepted segments on 71.5% of trials (low of 55.6% on session 1 to high of 84.7% on session 4) and correctly rejected segments on 84.8% of trials (low of 71.3% on session 1 to high of 92.5% on session 4).

Since the probe tasks evoked a binary (yes/no) response, the first step was to ensure good discriminability at the participant level. To this end we

calculated  $d'$ 's for all participants separately for the semantic and segmental probe tasks. All but one participant had  $d' > 0.9$  (mean  $d' = 1.48$ , standard deviation = 0.56) in the semantic probe and  $d' > 1.0$  (mean  $d' = 1.97$ , standard deviation = 0.81) in the segmental probe tasks, which indicate good discriminability. The participant with  $d' = 0.45$  in the semantic probe condition was excluded from further analyses.

Semantic probe and segment probe data were considered separately. For each probe task, responses to shared features were compared to responses to distinctive features within the same context. Only responses to correct features were considered (i.e., YES and NO responses to probes in which the correct response was YES) because there were clear predictions about the effects of context on the shared and distinctive features. It is less clear if responses to incorrect features should follow the same pattern, especially because some rejections could be made on the basis of general knowledge before any training even occurred (e.g., a type of tree is not going to swim). As in the analysis of the naming data, accuracy and response time

data were analyzed separately. Each model included feature type (shared or distinctive), session (1-4), and the two-way interaction between them as well as days since the last session as fixed effects. Following the recommendations of Barr and colleagues (2013) to “keep it maximal”, we attempted to implement a full random effects structure, but due to failures of convergence, random slopes over items were not included. The resulting random effects structure included random intercepts for subjects and items as well as random slopes over subjects for feature type, training session, their interaction, and days since the last session. Response times entered into analyses were log-transformed and excluded incorrect responses and outliers more extreme than 2.5 standard deviations from each participant’s mean (17.4% of semantic probe trials; 28.3% of segment probe trials).

Figure 7 shows the results of the semantic and segment probe tasks over sessions. Tables 3 and 4 summarize the outputs of the semantic and segment probe models, respectively.

Across the four sessions, participants became increasingly faster ( $t=-4.92$ ,  $p<.001$  in the semantic model;  $t=-2.45$ ,  $p=.014$  in the segmental model) and more accurate (a marginal effect  $z=1.71$ ,  $p=.088$  in the semantic model; a significant effect  $z=6.63$ ,  $p<.001$  in the segmental model) in verifying features, showing that they indeed learned both semantic and segmental features.

Importantly, in the model comparing shared and distinctive semantic features trained in semantic blocks, participants were faster to verify *distinctive* features than shared features ( $t=-3.72$ ,  $p=.001$ ). There was also a marginally significant negative interaction between feature type and session for response time ( $t=-1.83$ ,  $p=.072$ ), suggesting that participants tended to have greater increases in speed for the verification of *distinctive* features vs. shared features.

In contrast, in the models comparing shared and distinctive segments trained in segmental blocks participants were significantly faster ( $t=2.14$ ,  $p=.033$ ) and marginally more accurate ( $z=-1.90$ ,  $p=.057$ ) to verify *shared* segments relative to distinctive ones.



To summarize, semantic and segmental similarity led to opposite results in the feature probe tasks: there was an advantage for verification of distinctive as opposed to shared features in the semantic probe task, while there was an advantage for verification of shared as opposed to distinctive features in the segment probe task<sup>3</sup>.

### **General Discussion**

We examined specific learning challenges within different subcomponents of the word production system, focusing on the consequences of semantic and segmental similarity on word learning. As predicted, analysis of the naming data indicate that both types of similarity

---

<sup>3</sup> Note that we also found consistent results when comparing shared features trained in the critical contexts to distinctive features trained in other contexts (Breining, 2016). That is, the advantage for distinctive semantic features was present not only when directly comparing shared and distinctive semantic features trained in semantic blocks, but also when comparing shared semantic features trained in semantic blocks to the by-definition distinctive semantic features trained in segmental and unrelated blocks. Likewise, the advantage for shared segments was present not only when directly comparing shared and distinctive segments trained in segmental blocks, but also when comparing shared segments trained in segmental blocks to the by-definition distinctive segments trained in semantic and unrelated blocks. These analyses show that the effects discussed are not limited to relative differences between shared and distinctive features trained within the same context, but that these effects are also observed for other contexts as well.

led to overall interference during learning. These findings align well with the findings of Oppenheim (2018) who reported immediate semantic interference when new words were integrated into the lexicon, and extend these findings to segmental overlap. Importantly, we found that semantic and segmental similarity had contrasting effects on the learning of features at the two stages of word production. Semantic feature similarity during Word Selection, which requires the mapping of multiple semantic features to one word, facilitated learning of *distinctive* features. On the other hand, segmental feature similarity during Segmental Encoding, which requires the mapping of one word to multiple segments, facilitated learning of *shared* segments. These results support the central premise of this study: although there is a common challenge of integrating new and known information during learning, the specific form of the challenge is shaped by the structure of the subcomponents. This could mean opposite effects in local learning dynamics within the subcomponents in order to satisfy a single learning goal for the whole system. We expect that this principle is not confined to novel word

learning in the healthy adult language system: Similar dynamics are likely to exist throughout development (e.g., as children acquire representations for homonymous and synonymous words); in rehabilitation contexts (e.g., as individuals with aphasia undergo treatment for anomia); and beyond language in other complex multi-component cognitive systems as well (e.g., episodic memory, visual object recognition, etc.).

The findings have implications for both the type of learning mechanism, and the underlying production architecture on which such a mechanism operates. With respect to the learning mechanism, the asymmetry in learning shared vs. distinctive features stems directly from the differential weakening of the connections between word representations and semantic vs. segmental features during the integration of a new word into the production system. This finding favors models with both positive and negative changes to connection weights (e.g., Oppenheim et al., 2010) over those with only positive weight changes (e.g., Howard et al., 2006) which would not predict differential learning of shared vs. distinctive features or a reversed pattern of

feature learning as a function of similarity type. With regard to the underlying production architecture, the findings are only compatible with models that allow for feedback between segments and words. In the absence of such feedback, segmentally-similar competitors would not even be activated during production. There would thus be no activation of competing segments, and none of the dynamics explained in Figure 3 would be expected. We have previously argued that the interference induced by segmental similarity during production is evidence for feedback between segments and words (Breining et al., 2016; Nozari et al, 2016). The current findings further support this claim.

One might be concerned that the probe tasks involve perceptual and memory processes, but not word production processes. One important implication of these findings is that representations that are learned with a certain goal (e.g., production) reflect the influence of learning dynamics within that system even when accessed via other cognitive systems. Naming a picture repeatedly in order to solidify learning of a new label causes lasting changes to the connections between semantic features, segmental features,

that learned word form, and the word forms for similar words, and these changes are reflected in any task that taps into these connections.

A more serious concern might be that the observed differences in the effects of semantic and segmental similarity have nothing to do with the dynamics of learning but may simply reflect differences in representational sparsity: the set of segments in English is limited, whereas the set of semantic features is virtually unlimited. However, it is unclear why these differences in sparsity would result in the observed pattern, and not any other pattern. Thus, while we cannot refute this possibility with certainty, we can claim that the account we propose provides a parsimonious theoretical framework that predicts precisely the observed effects.

Finally, these findings may have implications for education and rehabilitation contexts. The finding that different subcomponents face different challenges may be helpful in structuring and interpreting learning. Furthermore, while the interference effects we report might suggest that similarity should be avoided, there is solid evidence that at least some kinds

of difficulty during learning have long term positive effects, a concept known as desirable difficulty (e.g., Bjork, 1994). If the interference observed here is indeed a desirable difficulty, one might expect benefits from training items in related sets (see also contextual priming in anomia studies, e.g. Martin & Laine, 2000).

Overall, the word production system must be optimally tuned to manage the common situation of producing similar words in quick succession. The work reported here shows that this may be achieved through a complex process of incremental strengthening and weakening of connections in the different subcomponents of the word production system.

### References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture

naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*, 667–692.

doi:10.1080/02724980443000142

Belke, E., Shao, Z., & Meyer, A. S. (2017). Strategic origins of early semantic facilitation in the blocked-cyclic naming paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*, 1659–1668. doi:10.1037/xlm0000399

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Breining, B.L. (2016). *Effects of semantic and segmental similarity on the production and learning of spoken and written words*. (Unpublished doctoral dissertation). Johns Hopkins University, Baltimore, MD, USA.

Breining, B. L., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*, 500–506.

doi:10.3758/s13423-015-0900-x

Breining, B. L., & Rapp, B. (2017). Investigating the mechanisms of written

word production: insights from the written blocked cyclic naming

paradigm. *Reading and Writing*, 1–30. doi:10.1007/s11145-017-9742-4

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed

practice in verbal recall tasks: A review and quantitative synthesis.

*Psychological Bulletin*, 132, 354–80. doi:10.1037/0033-2909.132.3.354

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a

privileged status in the computation of word meaning: Implications for

theories of semantic memory. *Journal of Experimental Psychology*.

*Learning, Memory, and Cognition*, 32, 643–658. doi:10.1037/0278-

7393.32.4.643

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic

context in the naming of pictures and words. *Cognition*, 81, B77-86.

doi:10.1016/S0010-0277(01)00135-4

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence



production. *Psychological Review*, *93*, 283–321. doi:10.1037/0033-295X.93.3.283

Foygel, D., & Dell, G. S. (2000). Models of Impaired Lexical Access in Speech Production. *Journal of Memory and Language*, *43*, 182–216. doi:10.1006/jmla.2000.2716

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*, 1393–1409. doi:10.3758/s13428-015-0647-3

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*, 464–82. doi:10.1016/j.cognition.2005.02.006

Laine, M., & Salmelin, R. (2010). Neurocognition of New Word Learning in the Native Tongue: Lessons From the Ancient Farming Equipment Paradigm. *Language Learning*, *60*, 25–44. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9922.2010.00599.x/full>

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

doi:10.1017/S0140525X99001776

Martin, N., & Laine, M. (2000). Effects of contextual priming on impaired word retrieval. *Aphasiology*, *14*, 53–70. doi:10.1080/026870300401595

Nozari, N., Freund, M., Breining, B. L., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*, 886–903.

doi:10.1080/23273798.2016.1157194

Oppenheim, G. M. (2018). The paca that roared: Immediate cumulative semantic interference among newly acquired words. *Cognition*, *177*, 21–

29. doi:10.1016/j.cognition.2018.02.014

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–52.

doi:10.1016/j.cognition.2009.09.007

Rapp, B., & Goldrick, M. (2000). Discreteness and Interactivity in Spoken Word

Production. *Psychological Review*, *107*, 460–499. doi:10.1037/TO33-295X.

Rapp, B., & Lipka, K. (2011). The literate brain: the relationship between

spelling and reading. *Journal of Cognitive Neuroscience*, *23*, 1180–1197.

doi:10.1162/jocn.2010.21507

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: the ARC

Nonword Database. *The Quarterly Journal of Experimental Psychology. A,*

*Human Experimental Psychology*, *55*, 1339–62.

doi:10.1080/02724980244000099

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech

production. *Cognition*, *64*, 249–284. doi:10.1016/S0010-0277(97)00027-9

Schnur, T. T. (2014). The persistence of cumulative semantic interference

during naming. *Journal of Memory and Language*, *75*, 27–44.

doi:10.1016/j.jml.2014.04.006

### Tables

**Table 1.** Results of the analysis of the semantic model of the training data, including both accuracy and response time.

Fixed effects	Accuracy				RT			
	Coefficient	SE	<i>z</i>	<i>p</i>	Coefficient	SE	<i>t</i>	<i>p</i>
Intercept	3.35	0.49	6.88	<.001	7.02	0.04	198.07	<.001
block type (semantic vs. unrelated)	-0.69	0.33	-2.08	.038	0.02	0.02	0.93	.365
training attempt within session	1.94	0.32	6.11	<.001	-0.09	0.01	-6.53	<.001
session	2.26	0.30	7.43	<.001	-0.23	0.01	-16.24	<.001



Random effects	Variance	Variance
subject intercept	2.1779	0.0120
block type (semantic vs. unrelated) subject slope	0.2059	0.0005
training attempt within session subject slope	0.2392	0.0013
session subject slope	0.6689	0.0025
days since last session subject slope	0.1202	0.0011
training trials since last trained subject	0.0692	0.0004

slope

block type (semantic vs. unrelated) *	0.2369	0.0003
---------------------------------------	--------	--------

training attempt within session|subject

slope

block type (semantic vs. unrelated) *	0.0662	0.0009
---------------------------------------	--------	--------

session|subject slope

training attempt within session *	0.1289	0.0002
-----------------------------------	--------	--------

session|subject slope

block type (semantic vs. unrelated) *	0.1154	0.0002
---------------------------------------	--------	--------

training attempt within session \*

session subject slope		
item intercept	0.6547	0.0082
training attempt within session item slope	0.0684	0.0004
session item slope	0.1061	0.0003
days since last session  item slope	0.0182	0.0009
training trials since last trained item slope	0.0279	0.0001
training attempt within session *	0.0808	0.0001
session item slope		
Residual		0.0869

---



**Table 2.** Results of the analysis of the segmental model of the training data, including both accuracy and response time.

Fixed effects	Accuracy				RT			
	Coefficient	SE	<i>z</i>	<i>p</i>	Coefficient	SE	<i>t</i>	<i>p</i>
Intercept	3.66	0.53	6.92	<.001	6.99	0.04	191.36	<.001
block type (segmental vs. unrelated)	-0.60	0.33	-1.81	.070	-0.01	0.03	-0.36	.721
training attempt within session	2.13	0.42	5.10	<.001	-0.07	0.01	-5.91	<.001
session	2.48	0.31	8.00	<.001	-0.21	0.01	-15.94	<.001
days since last session	-0.13	0.14	-0.88	.378	0.02	0.01	1.03	.319
training trials since last trained	0.07	0.12	0.61	.545	-0.02	0.02	-1.24	.232

block type (segmental vs. unrelated) *	-0.42	0.28	-1.54	.125	0.02	0.01	3.20	.003
training attempt within session								
block type (segmental vs. unrelated) *	-0.30	0.23	-1.31	.190	0.01	0.01	0.91	.373
session								
training attempt within session * session	-0.03	0.33	-0.10	.919	0.01	0.01	1.30	.211
block type (segmental vs. unrelated) *	-0.23	0.24	-0.99	.321	-0.02	0.01	-2.57	.016
training attempt within session * session								

---

Random effects	Variance	Variance
subject intercept	2.6845	0.0102

---

block type (segmental vs. unrelated) subject	0.0796	0.0004
slope		
training attempt within session subject	0.9401	0.0010
slope		
session subject slope	0.5335	0.0011
days since last session subject slope	0.0041	0.0011
training trials since last trained subject slope	0.0143	0.0010
block type (segmental vs. unrelated) *	0.0338	0.0001
training attempt within session subject		
slope		

block type (segmental vs. unrelated) *	0.0343	0.0006
session subject slope		
training attempt within session *	0.5874	0.0007
session subject slope		
block type (segmental vs. unrelated) *	0.0089	0.0001
training attempt within session *		
session subject slope		
item intercept	0.4328	0.0111
training attempt within session item slope	0.1303	0.0002
session item slope	0.0486	0.0009

days since last session  item slope	0.0057	0.0006
training trials since last trained item slope	0.0142	0.0008
training attempt within session *	0.1853	0.0004
session item slope		
Residual		0.0969

---

**Table 3.** Results of the analysis of the model of the semantic probe data comparing shared and distinctive features trained in semantic blocks, including both accuracy and response time.

Fixed effects	Accuracy				RT			
	Coefficient	SE	<i>z</i>	<i>p</i>	Coefficient	SE	<i>t</i>	<i>p</i>
Intercept	1.41	0.14	10.21	<.001	6.83	0.03	245.43	<.001
feature type (distinctive vs. shared)	0.09	0.07	1.37	.171	-0.02	0.01	-3.72	.001
Session	0.14	0.08	1.71	.088	-0.04	0.01	-4.92	<.001
days since last session	0.06	0.07	0.82	.411	0.01	0.01	0.99	.360
feature type (distinctive vs. shared) *								
session	0.06	0.07	0.88	.380	-0.01	0.01	-1.83	.072

---

Random effects	Variance	Variance
subject intercept	0.1339	0.0114
feature type (distinctive vs. shared) subject		
slope	0.0206	0.0002
session subject slope	0.0343	0.0007
days since last session subject slope	0.0109	0.0003
feature type (distinctive vs. shared) *		
session subject slope	0.0293	<0.0001
item intercept	0.0575	0.0003

---

Residual

0.0445

---



**Table 4.** Results of the analysis of the model of the segment probe data comparing shared and distinctive segments trained in segmental blocks, including both accuracy and response time.

Fixed effects	Accuracy				RT			
	Coefficient	SE	<i>z</i>	<i>p</i>	Coefficient	SE	<i>t</i>	<i>p</i> <sup>a</sup>
Intercept	1.41	0.26	5.52	<.001	6.77	0.03	207.71	<.001
feature type (distinctive vs. shared)	-0.14	0.07	-1.90	.057	0.01	0.01	2.14	.033
Session	0.77	0.12	6.63	<.001	-0.05	0.02	-2.45	.014
days since last session	0.06	0.08	0.71	.480	<0.01	0.01	-0.13	.897
feature type (distinctive vs. shared) *								
session	-0.13	0.07	-1.91	.056	-0.01	0.01	-1.87	.062

Random effects	Variance	Variance
subject intercept	0.5156	0.0127
feature type (distinctive vs. shared) subject		
slope	0.0147	<0.0001
session subject slope	0.0903	0.0053
days since last session subject slope	0.0003	0.0012
feature type (distinctive vs. shared) *		
session subject slope	0.0083	0.0001
item intercept	0.2236	0.0018

Residual	0.0533
----------	--------

---

<sup>a</sup> The p-values reported for this model were calculated using the approximation of the normal distribution instead of the Satterthwaite approximation implemented in lmerTest. This is because the model used here gave a warning about convergence failure that prevented application of lmerTest. Following the recommendations of Bates et al. (2018) in the lme4 documentation, we tried all available optimizers, which converged to practically equivalent values, meaning it is reasonable to treat the convergence warning as a false positive and report the results of the model.

### Figure Captions

**Figure 1.** A schematic of the word production system.

**Figure 2.** Learning of two semantically-related items, focusing on the Word

Selection stage of production prior to Segmental Encoding. w=word

representation; sem=semantic feature. Orange (lighter gray) represents

activation. Thicker lines represent increased connection strength and/or

increased activation.

**Figure 3.** Learning of two segmentally-related items, focusing on the

Segmental Encoding stage of production after Word Selection. w=word

representation; seg=segment. Orange (lighter gray) represents activation.

Thicker lines represent increased connection strength and/or increased

activation.

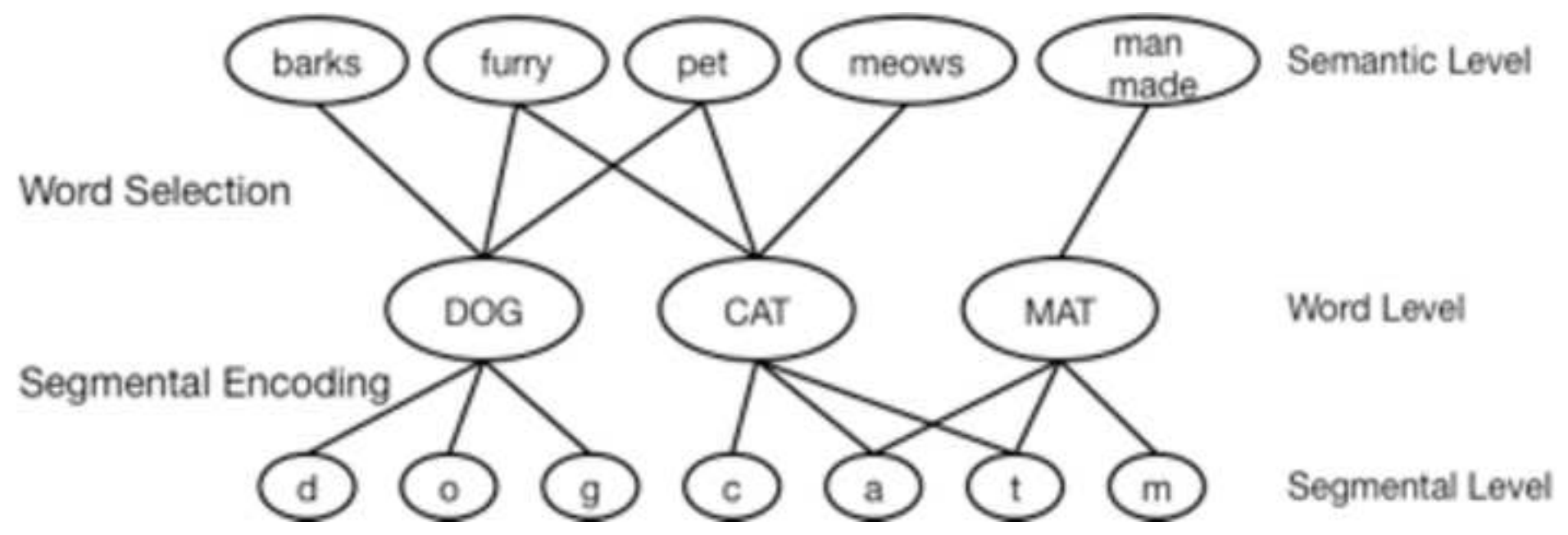
**Figure 4.** Example of an item in the training set.

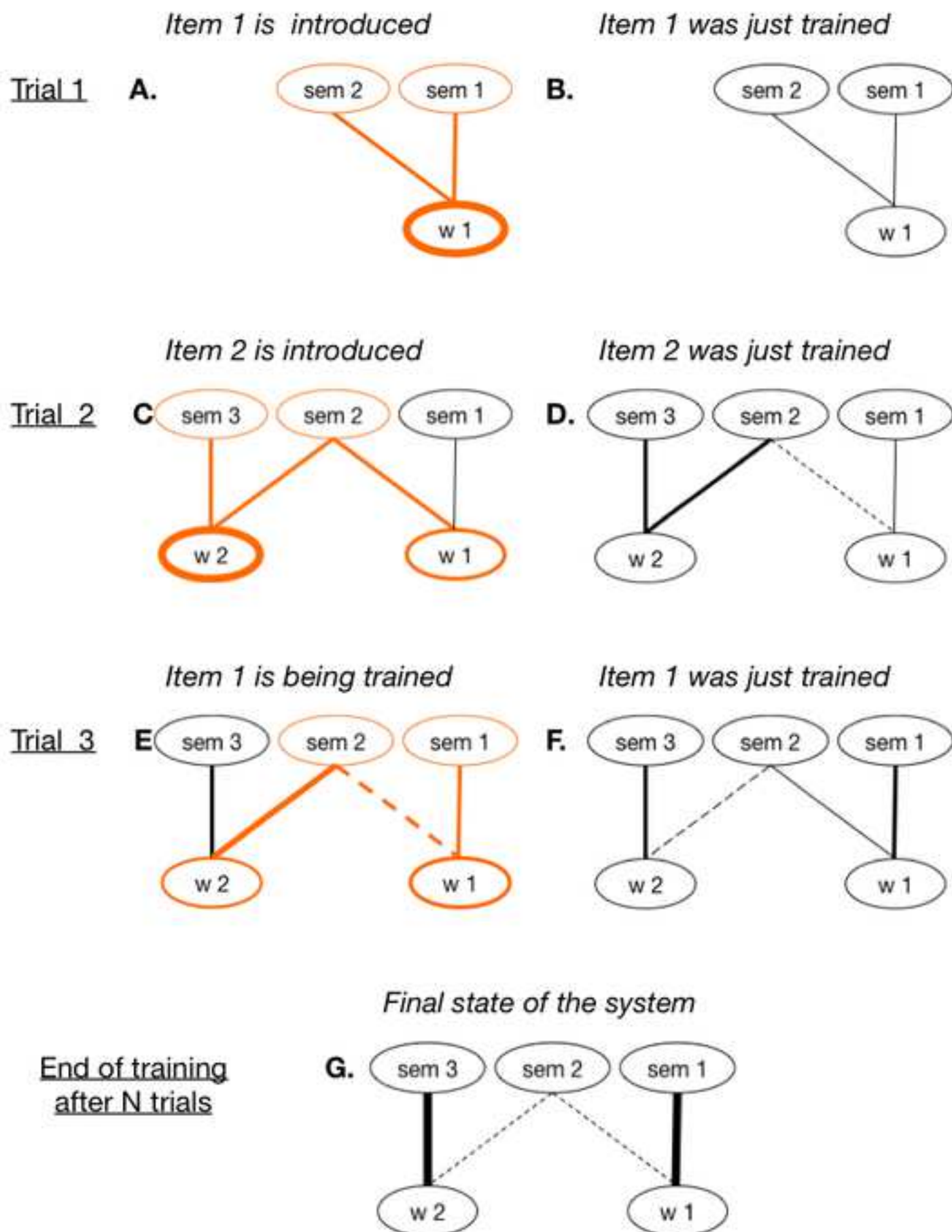
**Figure 5.** Structure of trial during training.

**Figure 6.** Results of the training task. Panel *a* shows accuracy across the three training attempts within each session, collapsed across all four training sessions for the three training contexts. Panel *b* shows accuracy across the three training attempts within each session for each of the four training sessions for the three training contexts. Panel *c* shows response time across the three training attempts within each session, collapsed across all four training sessions for the three training contexts. Panel *d* shows response time across the three training attempts within each session for each of the four training sessions for the three training contexts. All panels depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

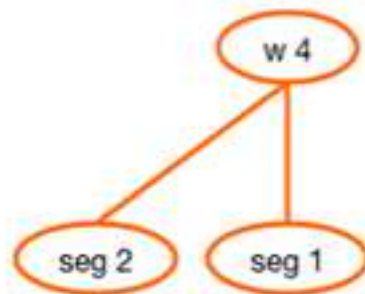
**Figure 7.** Results of the semantic and segment probe tasks over sessions.

Panel *a* shows accuracy for verification of semantic features across sessions, comparing shared and distinctive features trained in semantic blocks. Panel *b* shows response time for verification of semantic features across sessions, comparing shared and distinctive features trained in semantic blocks. Panel *c* shows accuracy for verification of segments across sessions, comparing shared and distinctive segments trained in segmental blocks. Panel *d* shows response time for verification of segments across sessions, comparing shared and distinctive segments trained in segmental blocks. All panels depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

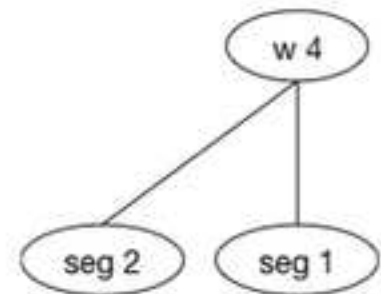
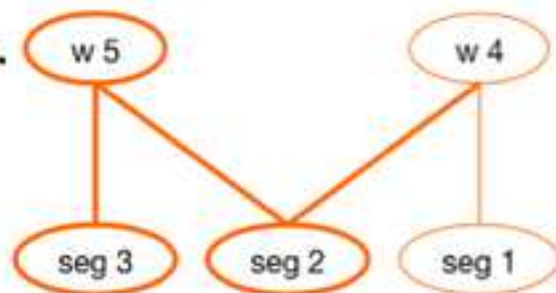




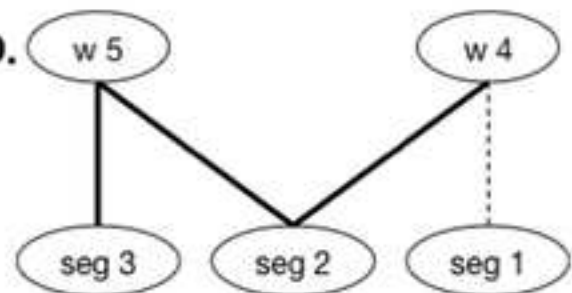
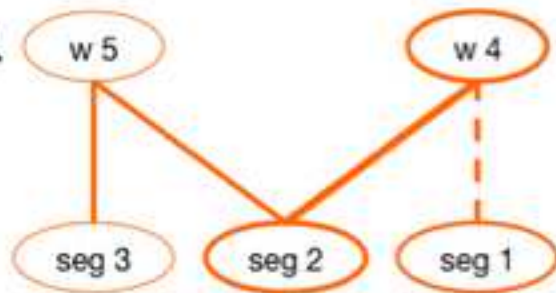


*Item 1 is introduced**Item 1 was just trained*Trial 1 A.

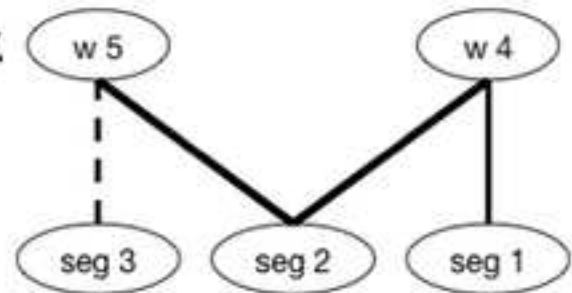
B.

*Item 2 is introduced**Item 2 was just trained*Trial 2 C.

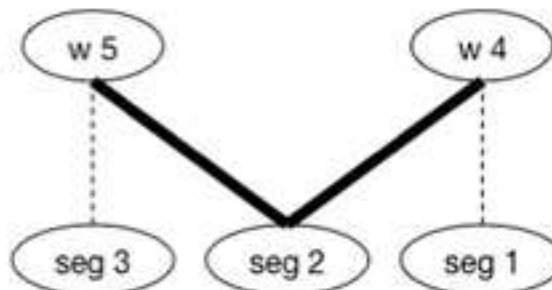
D.

*Item 1 is being trained**Item 1 was just trained*Trial 3 E.


F.

*Final state of the system*End of training  
after N trials

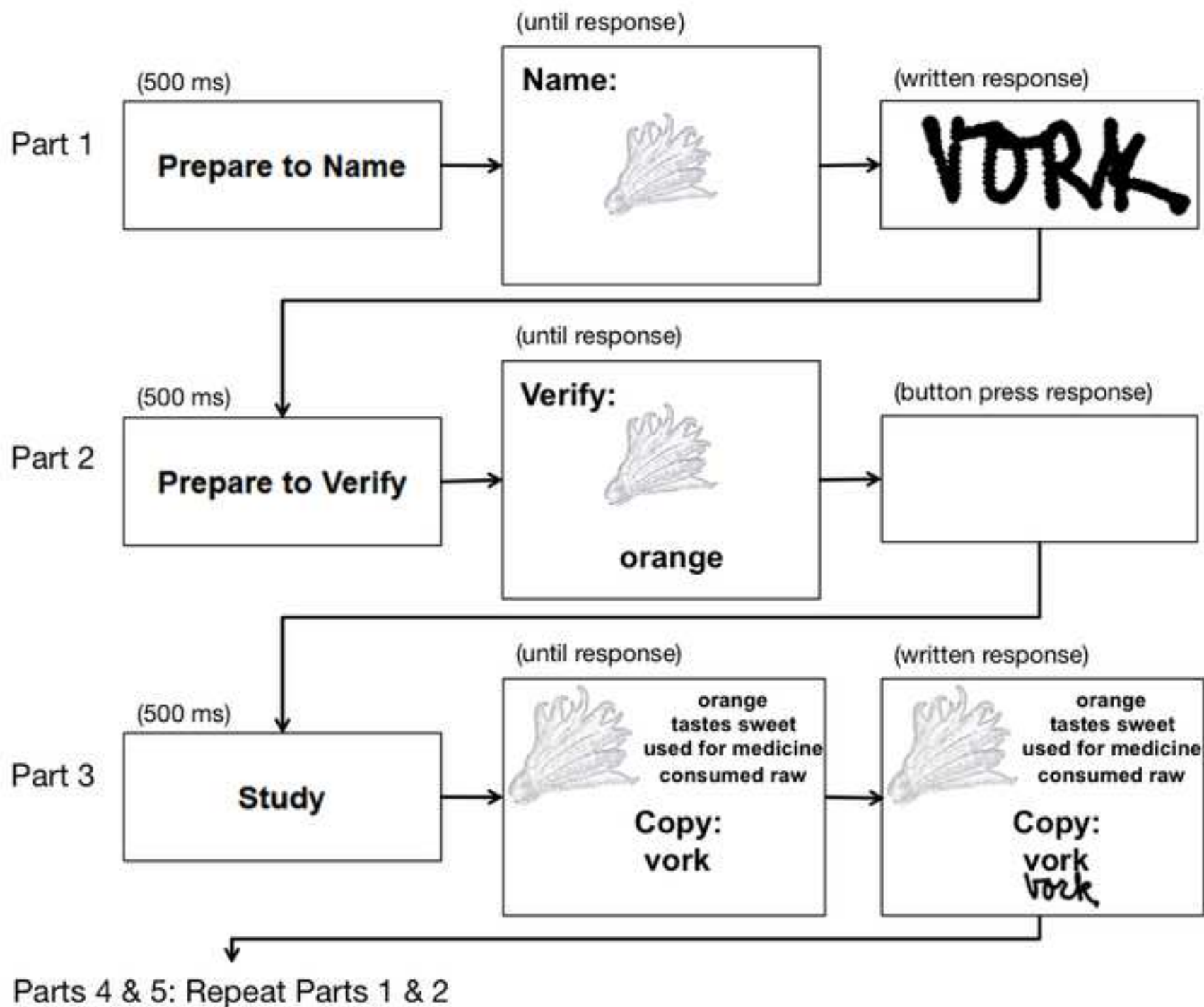
G.

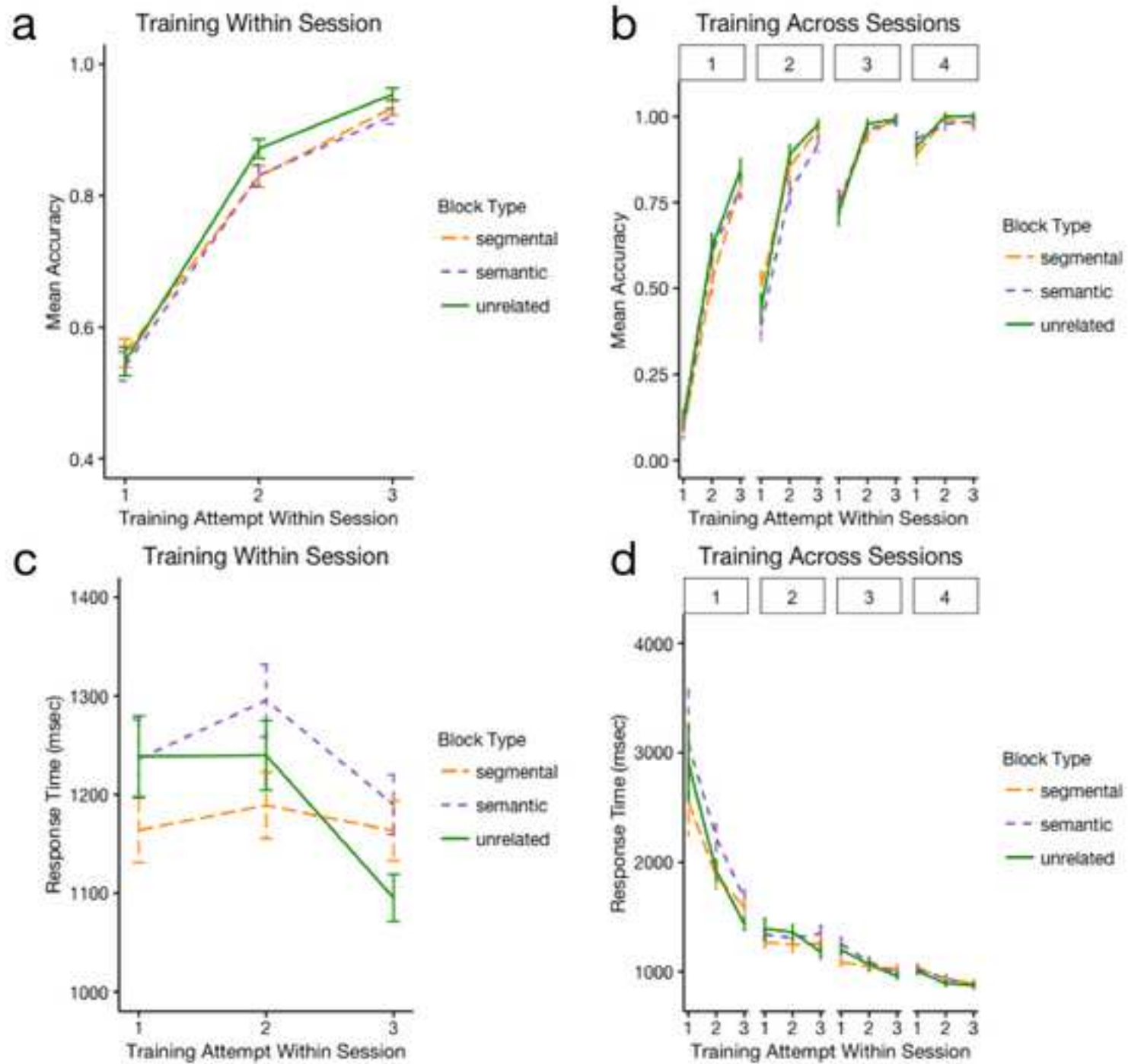


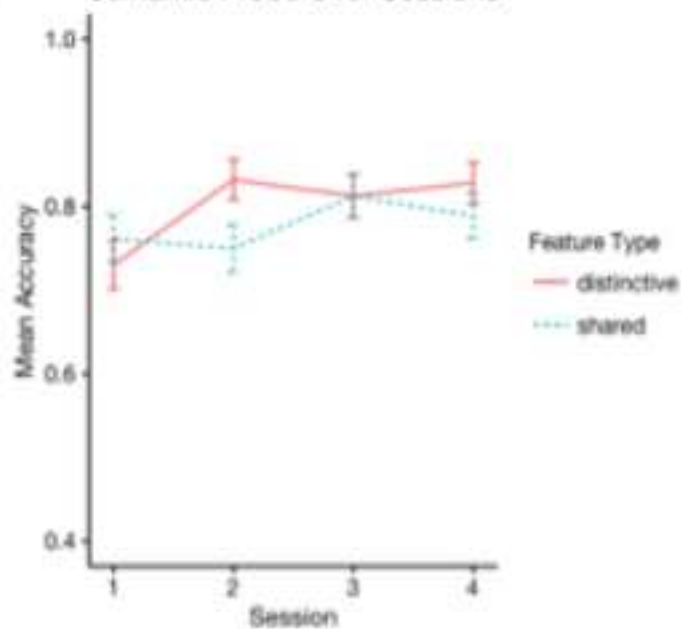
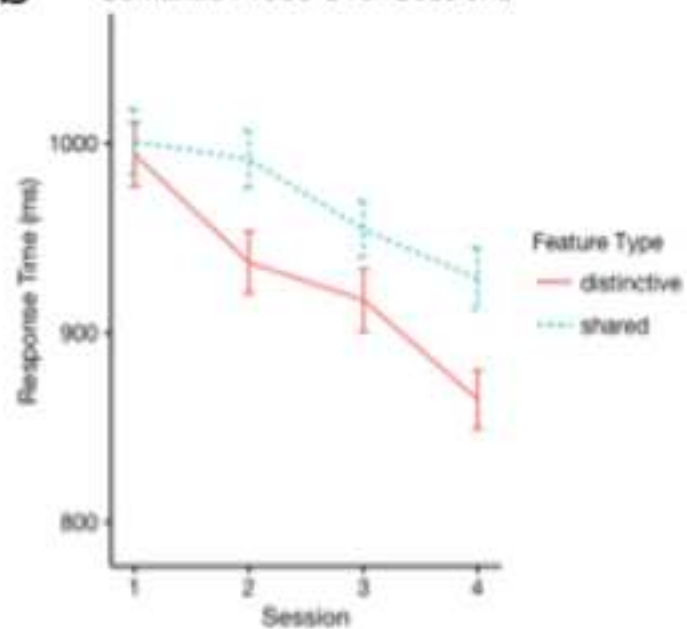
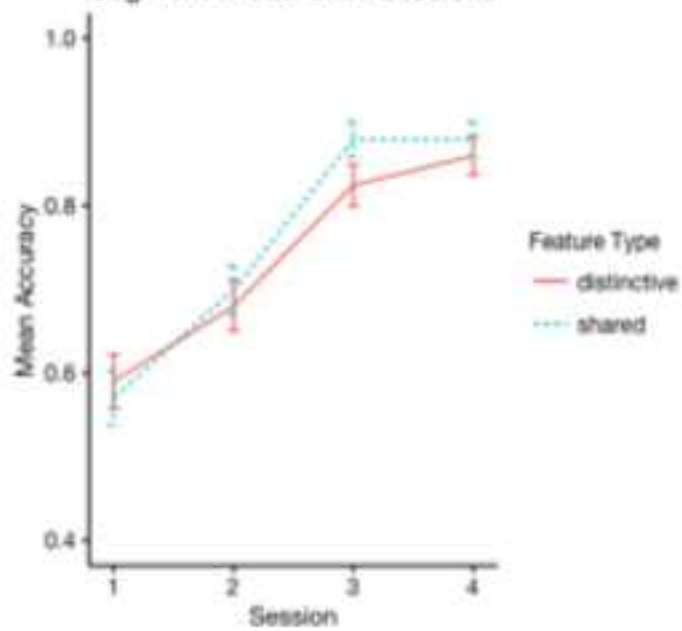
---

Picture	Spelling	Pronunciation	Features
	vork	vʊər̩k	Orange Tastes sweet Used for medicine Consumed raw

---





**a** Semantic Probe Over Sessions**b** Semantic Probe Over Sessions**c** Segment Probe Over Sessions**d** Segment Probe Over Sessions