The best-laid plans of mice and men: Competition between top-down and preceding-item cues in plan execution

Zara Harmon (zharmon@umd.edu)
Department of Linguistics and University of Maryland Institute for Advanced Computer Studies (UMIACS), University of Maryland, 8125 Paint Branch Dr, College Park, MD, 20740

Vsevolod Kapatsinski (vkapatsi@uoregon.edu)
Department of Linguistics, 1290 University of Oregon, Eugene, OR, 97403

Abstract

There is evidence that the process of executing a planned utterance involves the use of both preceding-context and top-down cues. Utterance-initial words are cued only by the top-down plan. In contrast, non-initial words are cued both by top-down cues and preceding-context cues. Co-existence of both cue types raises the question of how they interact during learning. We argue that this interaction is competitive: items that tend to be preceded by predictive preceding-context cues are harder to activate from the plan without this predictive context. A novel computational model of this competition is developed. The model is tested on a corpus of repetition disfluencies and shown to account for the influences on patterns of restarts during production. In particular, this model predicts a novel Initiation Effect: following an interruption, speakers re-initiate production from words that tend to occur in utterance-initial position, even when they are not initial in the interrupted utterance.

Keywords: Serial order; language production; repetition; initiation; retrieval; planning; HiTCH

Introduction

The problem of producing actions in the right serial order is a classic one. Following Lashley (1951), an early period of chain models, in which each action activated its successor, gave way to a dominance of hierarchical models, in which each action was cued by top-down cues specifying its position in a hierarchical plan. Language became the paradigm example of hierarchical planning, with words or morphemes serving as nodes in the hierarchy. Yet, early on, Osgood (1963) argued that preceding-context and top-down cues can both activate upcoming items in processing. Our work develops this idea, arguing that the two types of cue both activate the future in processing, and compete for activating the future during learning.

Empirical evidence for the existence of preceding-item cues in production comes from the existence of 'habit slips' (e.g., Bannard et al., 2019; Reason, 1980; Wickelgren, 1966). Such an error is well exemplified by Benjamin Netanyahu's recent slip of calling the British Prime Minister *Boris Yeltsin* instead of *Boris Johnson*. *Boris* is an exceptionally good cue to *Yeltzin* because it was almost always followed by *Yeltzin* before *Johnson*'s career took off. According to the habit slip account of such errors, *Yeltzin* is produced because it is activated by *Boris*, the predictive preceding item. Having habitually produced *Yeltzin* after *Boris*, Netanyahu slipped into a minor international scandal. Habit slips also occur outside of language (Reason, 1980). A likely familiar example is automatically turning off the lights on someone when exiting a room.

While preceding-item cues can generate habit slips, they are generally useful in constraining the set of possible continuations. For example, Rubin (1977) shows that providing speakers with function words helps them recall upcoming content words in a familiar text (*all … are … and … by their … with … such as … and the …*).

While preceding-context cues are helpful, they are insufficient. Top-down cues are necessary to navigate 'switch points', where the preceding context is consistent with more than one continuation, and the less common continuation needs to be chosen. Such switch points are known to be particularly challenging to navigate and are targeted for extensive practice (e.g., Chaffin & Imreh, 2002, in piano performance). The use of top-down cues in such contexts is necessary for selecting the right continuation. For example, a writer can only know whether to continue *I love my ca* into *cat* or *car* because s/he knows which word s/he intends to produce.

The need for both types of cues raises the question of how preceding-context and top-down cues interact during learning. We argue that this interaction is competitive. In particular, according to the HiTCH (Hierarchy To Chain) model we develop, top-down cues that routinely compete with strong preceding-context cues grow relatively ineffective as a result. This makes it difficult to produce an item that has always been preceded by a predictive cue outside of that predictive context. In this way, a predictive context grows increasingly sufficient and increasingly necessary to activate an upcoming item. For example, if one only makes the tongue motion producing the sound at the end of *hang* after a vowel, it becomes increasingly difficult to make the same motion in other contexts (e.g., to produce the beginning of the Vietnamese name *Nguyen*).

The HiTCH Model

We implemented the idea of competition between preceding-item and top-down cues in a simple computational model, which we call HiTCH, or Hierarchy to Chain, to indicate that practice with a predictive sequence increasingly weakens top-down cues to non-initial items in that sequence. This model is intended as the simplest

possible framework for investigating cue competition between preceding-item and top-down cues during learning. It is related to simple recurrent networks that include 'plan' or 'goal' input units alongside a representation of preceding context (Dell, Juliano & Govindjee, 1993; Cooper, Ruh & Mareschal, 2014) and the classic Widrow–Hoff/Rescorla–Wagner learning rule (RW; Rescorla & Wagner, 1972).

HiTCH claims that top-down cues can *weaken* when they co-occur with other cues, even if they are much more predictive than their competitors. In language, abdication of top-down control is necessary to explain the existence of collocations. To implement this claim, we modified the Rescorla-Wagner learning rule to allow preceding-context cues to interfere with top-down cues, but not vice-versa.

HiTCH is intended to model the effects of experience on an important aspect of plan execution. Namely, it is well known that planning at the semantic/conceptual/lemma level outpaces planning at the form/phonological/lexeme level (Meyer, 1996). The task of HiTCH is to select the right wordform for execution given a planned semantic/conceptual future and the preceding context.

Whenever HiTCH encounters a word for the first time, it initializes a top-down connection from the corresponding semantic node to the wordform at an initial connection weight ($V_{sem,\mathrm{j}}$). These connections are intended to represent the activation that a planned wordform receives from the production plan. All associations between corresponding semantics and forms are initialized as having equal weights, which means that a novel word is approximately equally accessible in all contexts. These initial weights are relatively high because the claim of the model is that novel plans are hierarchical. That is, a novel plan is executed entirely via top-down control. Here, we report simulation results for the two values at the limits of the plausible range, 1 and 0.5 (see Figures 1 and 2).

Top-down connections between noncorresponding semantics and words are initialized at −1 to ensure a low rate of occurrence of speech errors (Dell et al., 1993). These weights never change, providing a stable ability to produce the intended word most of the time. The greater stability of inhibitory connections is inspired by Liberti et al.'s (2016), "principle of motor stability: spatiotemporal patterns of inhibition can maintain a stable scaffold for motor dynamics while the population of principal neurons that directly drive behavior shift from one day to the next" (p. 1665).

The model is exposed to a corpus word by word predicting each word from its semantics and the preceding word, if any. The model distinguishes between two types of word tokens. When a $word_{\mathrm{j}}$ occurs at the beginning of an utterance (i.e., $j = 1$), it is activated by top-down input alone ($V_{sem,\mathrm{j}}$). When it is not utterance-initial ($j > 1$), it is cued both by the top-down input and the preceding word ($V_{\mathrm{i,j}}$). Whenever $word_{\mathrm{j}}$ is utterance-initial, its top-down cue ($V_{sem,\mathrm{j}}$) is incremented; and when it is not initial, its top-down cue is decremented in proportion to the strength of the preceding-word cue ($V_{\mathrm{i,j}}$) occurring in that utterance. This implements the main claim of the model that top-down cues weaken because of competition from preceding-context cues.

The preceding-word cue ($V_{\mathrm{i,j}}$) is incremented whenever $word_{\mathrm{j}}$ occurs after $word_{\mathrm{i}}$ and decremented when one of the words occurs without the other, using the following rule:

$$\Delta V_{\mathrm{i,j}} = \alpha\beta\left(\lambda - V_{\mathrm{i,j}}\right)$$

As evident from the equation above, three sets of parameters influence the magnitude of change in the association weights of the preceding-word cue. Values of λ represent asymptotic levels of associative strength between items. For updating associations between adjacent words λ is set to 1. The $(1 - V)$ term, which represents the difference between the correct activation for a present outcome and its current activation makes the growth of a weight decelerate as it grows, asymptotically approaching 1. For associations between an absent cue and a present outcome or a present cue and an absent outcome, λ is set to 0, so that the weights of associations between items that do not co-occur asymptotically approach 0. Associations between absent cues and absent outcomes are not updated (Rescorla & Wagner, 1972; Bush & Mosteller, 1951).

Values of α and β are restricted a priori to the interval $0 < \alpha, \beta \leq 1$. The parameter α, represents the salience of the cue and β represents the salience of the outcome. The product of α and β represents the learning rate. When both $word_{\mathrm{i}}$ and $word_{\mathrm{j}}$ are present, α is set to 0.1 and β is set to 1. This corresponds to a learning rate of 0.1, a rather high value intended to make sure that conditional probabilities between adjacent words are learned even for the words that are rare in our relatively small training corpus (the Switchboard Corpus of American English conversations; 1.7 million words; Godfrey et al., 1992).

There is evidence that absences are less salient than presences (Tassoni, 1995). They are also less informative than presences because any two words occur apart much more often than they occur together (see McKenzie & Mikkelsen, 2007). For this reason, α and β are reduced for absent cues and outcomes respectively. We set β to 0.1 for absent outcomes and α to 0.00003 for absent cues. Setting β to 0.1 for absent outcomes ensures that item-to-item association weights are affected equally by $log(C(word_{\mathrm{i}},word_{\mathrm{j}}))$ and $-log(C(word_{\mathrm{i}}))$ exactly like surprisal conditional on the preceding word. This means that, like surprisal (Levy, 2008), $V_{\mathrm{i,j}}$ tracks how much information $word_{\mathrm{i}}$ provides about the following $word_{\mathrm{j}}$.

Following Rescorla and Wagner (1972), we assumed absent cues preceding present outcomes to be less salient than absent outcomes following present cues. This is justified by the idea that an absent outcome can be *unexpectedly* absent when anticipated on the basis of a present cue, whereas the absence of a cue is not surprising (Rescorla & Wagner, 1972). While the classic RW model sets α for absent cues at 0, we set it at a near-zero value of 0.00003, selected as the mean probability of occurrence across the words in Switchboard. The low but non-zero

setting indicates that one learns a little about absent cues (as observed in the literature; Tassoni, 1995).

The top-down cue weight to *word*$_j$ from its semantics is updated as follows:

$$\Delta V_{sem,j} = \begin{cases} \alpha\beta\left(1 - V_{sem,j}\right), & j = 1 \\ \alpha\beta\gamma V_{i,j}\left(0 - V_{sem,j}\right), & j > 1 \end{cases}$$

Whenever a *word*$_j$ is utterance-initial, that is, not preceded by another word within the same utterance ($j = 1$), the association from *sem*$_j$ to *word*$_j$ is incremented by $\alpha\beta\left(1 - V_{sem,j}\right)$. That is, when a word occurs without a preceding word, it becomes more accessible from top-down input. To the extent that a word needs to be produced context-independently, its context-independent, top-down accessibility strengthens. Here, α is set to 0.1 and β set to 1 because both the cue and the outcome are present

The second equation above implements cue competition. Whenever a *word*$_j$ is not utterance-initial ($j > 1$), the weight of the association from *sem*$_j$ to *word*$_j$ is decremented by $\alpha\beta\gamma V_{i,j}\left(0 - V_{sem,j}\right)$, where the magnitude of change in top-down associations is determined not only by α and β but also by the weight of the item-to-item association ($V_{i,j}$). According to this equation, the top-down association from semantics to the corresponding word decreases to the extent that the word is encountered in predictive preceding contexts ($V_{i,j} > 0$). This is an important difference between HiTCH and the RW rule, which would not decrease these cue weights. In RW, being the most predictive cues to upcoming words, top-down cues face negligible cue competition. Because $\alpha\beta V_{i,j}$ is always less than 1, top-down accessibility of a word recovers on initial trials, where top-down access is necessary, more quickly than it diminishes on trials when it is *un*necessary.

Multiplication by the weight of the preceding-context cue ($V_{i,j}$) is motivated by results in habit formation. In particular, Ouelette and Wood (1998) found that, in predicting the likelihood of continued performance for actions that are performed frequently in fixed contexts, the frequency of prior performance of a behavior is more important than stated desire to perform the behavior, while the opposite is true for actions performed rarely or in variable contexts.

Here, the parameter γ ($0 < \gamma \leq 1$) determines the *extent* to which the strengthening preceding-item cues interfere with top-down cues. The predictions tested below hold for any value of γ above 0. However, we set this parameter fairly low, at 0.05 in the following simulations because if γ is set too close to 1, then any increase in the weight of a preceding-item cue is offset by a corresponding decrease in the weight of the top-down cue. In contrast, when γ is low, preceding-item cues can strengthen without reducing the top-down weights (too) much. We believe the latter is the right behavior, because experience makes words (and other actions) more accessible in the kinds of contexts in which they have been experienced.

Simulation Results

Top-down association weights in HiTCH are influenced by three lexical characteristics, Initialness proportion, Predictiveness of preceding-item cues and Word frequency. The Initialness proportion refers to the proportion of all occurrences of a word that are utterance-initial. Words with a low Initialness proportion tend to be poorer initiators, having a lower top-down cue weight (Figure 1). The prediction that initiating production from these words is more difficult is the Initiation Effect. The extreme examples are words like *know* and *ago*, which almost never occur utterance-initially.

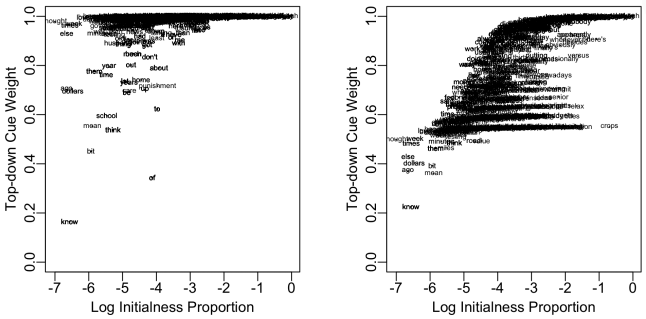

Figure 1: Top-down cue weight to a word as a function of the proportion of times the word occurs in the utterance-initial position. Top-down cues are initialized at 1, the maximum possible weight, in the left panel, and at 0.5, the mean possible weight, in the right panel.

The strength of the top-down cue weight to a word also depends on how predictive the preceding-item cues to the word are on the occasions that the word is not utterance-initial. Here, again, the cases of *know* and *ago* are instructive. *Know* tends to occur after *you* or *don't*, which are highly predictive of its occurrence. *Ago* always occurs after some unit of time (*years*, *weeks*, *months*, *summers* or *Wednesdays* in this corpus), and these words are predictive of *ago*'s occurrence: they greatly raise the likelihood of *ago* occurring next. The influence of cue predictiveness means that, other things being equal, a word that often occurs unexpectedly should be more accessible in new contexts than a word that tends to be expected when it occurs (as found by Adelman, Brown & Quesada, 2006, and Yan, Mollica & Tanenhaus, 2018).

Word frequency influences how far the top-down cue weight for a word can move away from its initial value. As Figure 2 shows, word frequency is negatively correlated with top-down association weight to the extent that words start out with high top-down weights. At the extreme, if top-down weights have nowhere to go but down, frequent words have more opportunities to reduce in top-down cue weight but no words have an opportunity to exceed the initial weight of a top-down cue. As a result, only frequent words develop weak top-down weights. This is why all infrequent words have strong top-down weights in the left panel of Figure 2. However, not all frequent words have weak top-down weights. The top-down cues to words like *well* or

yeah, which tend to occur utterance-initially, retain their strength.

When top-down cues start out with lower-than-maximal weights (right panel of Figure 2), infrequent words have intermediate top-down cue weights while frequent words have more extreme weights, which make them either excellent or poor initiators (*yeah*, *well, like* vs. *know*, *ago*). We believe this to be closer to the truth because it is frequent words like *well* and *yeah* that serve as placeholders and discourse markers, intervening in many different contexts to buy time.

Frequent words that occur in predictive contexts still become poor initiators when the weights have as much room to grow as to shrink. Thus, words like *know* end up with lower top-down weights than they start with even in the right panel of Figure 2. A word like *know* within a sequence like *you know* is therefore expected to be produced in large part because of its strong association with the preceding word rather than top-down input. Sequences like *you know* are more chain-like, less hierarchical, and less subject to top-down control than novel sequences. This is the sense in which hierarchical plans become chainlike in HiTCH.

The reduction of top-down weights does not affect all words in all contexts. Indeed, depending on initial weights of top-down cues and the amount of interference that the strengthening item-to-item associations generate, most sequences may become more reliant on top-down control. This is the case in the right panel of Figure 2 because the effective learning rate for top-down weight increases for initial word tokens was set to be 20 times larger than the rate for weight decreases for non-initial tokens. As a speaker's experience with the language grows, s/he learns where control is needed, and where it can be abdicated in favor of habit.

Figure 3 shows that item-to-item cue weights in HiTCH largely track log probability of an item conditioned on the item that precedes it in the processing sequence, with the caveat that many items' weights are at floor. The correlation is stronger at fast learning rates: a fast learning rate allows all cue weights to approximate conditional probabilities, whereas a slow learning rate prevents the weights of rare cue from moving far away from their initial weights. This appears to be appropriate, as one knows less about rare cues. Accordingly, the strongest associations are between items comprising highly cohesive sequences like *kind of*, *going to*, *used to*, *lot of*, *grew up* and *junior high*.

Controlling for conditional probability of an outcome given a cue, the cue-outcome association is weaker when the outcome is frequent in the absence of the cue (Figure 4). This constitutes the Inverse Base Rate Effect: controlling for the probability of an outcome given a cue, the cue activates a frequent outcome less than it activates a rare outcome.

Repetition Disfluencies: A Test Case

Disfluencies are thought to arise primarily when the speaker is in a temporary tip-of-the-tongue state. That is, the speaker knows what they want to say next but is unable to retrieve the next word (Hieke, 1981). For this reason, they are especially likely to occur primarily before hard-to-access content words that are low in predictability given the preceding context, and have many semantically similar competitors (Harmon & Kapatsinski, 2015; Schnadt, 2009). In a repetition disfluency, one or more words preceding a difficult content word are repeated. Repetitions are common in preposing languages like English, where hard-to-retrieve content words are preceded by highly predictable related function words. For example, in *I work for the speech, for the speech group*, *for the speech* is repeated as the speaker is trying to come up with *group*.

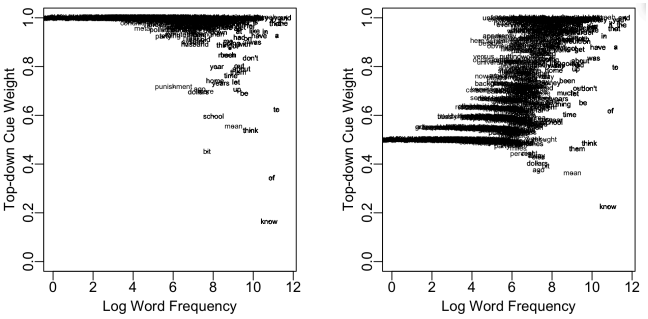

Figure 2: Frequent words have the chance to depart from their initial weights, whether those weights are at maximum (left) or in the middle of the range (right).

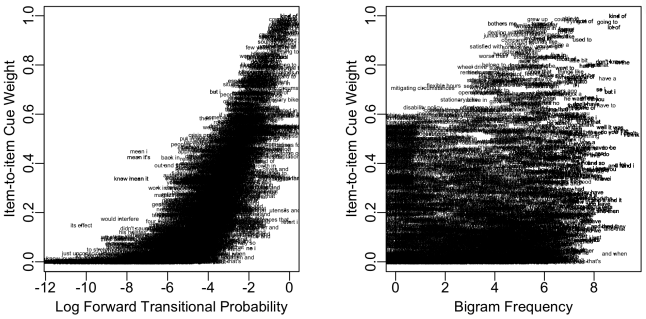

Figure 3: Item-to-item cue weights track forward conditional probabilities (left) rather than bigram frequency (right)

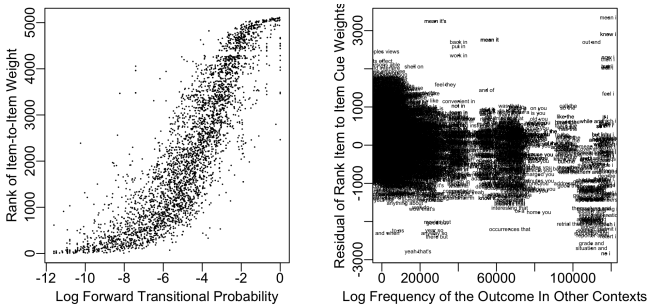

Figure 4: Item-to-item cue weights are strongly influenced by forward conditional probability (left) and weakly weakened by the high frequency of the outcome item after other cues (right). Right panel shows residuals of cue weights from a generalized additive model fit to the left panel as a function of log frequency in other contexts.

Interestingly, repetition disfluencies are rare to unattested in postposing languages, where function words are not related to, and therefore less predictive of the content words that follow (Fox, Hayashi & Jasperson, 1996). We believe repetition disfluencies occur in preposing languages because they allow the speaker to repeatedly cue the upcoming word—the word they are having trouble retrieving—with a somewhat predictive preceding-item cue. Existence of repetition disfluencies in preposing but not postposing languages therefore provides support for the hypothesis that words cue their successors.

Predictions

Initiation: A novel prediction of cue competition A repetition disfluency involves interrupting and reinitiating speech production. According to HiTCH, initiating speech production from a word involves activating this word using top-down input alone, without activation from a preceding-item cue. Initiation of speech production from a word should therefore be difficult when the top-down cue to the word is weak. Because speakers avoid attempting to produce difficult words (Schwartz & Leonard, 1982), they should avoid restarting speech production from words with weak top-down cues. While this effect is predicted by HiTCH, it has not previously been tested.

Retrievability of the past We assume that words need to be deactivated upon execution so that the speaker does not continue to repeat the word they have just selected forever (Dell, Burger & Svec, 1997). However, if the future is unavailable, repetition of the current word is the optimal behavior as it allows the speaker to buy time without accessing any word other than the word currently active. Repetition disfluencies occur when the future is unavailable, and therefore the present is not yet deactivated.

About 25% of the time the speaker repeats more than one word. Words that have already been produced by the time the speaker encounters a difficulty in accessing the upcoming word – e.g., *for* and *the* in the example *for the speech, for the speech group* – must have already been deactivated before they are repeated. We propose that speakers must retrieve deactivated words from working memory by using the words that follow them as cues.

The ability to retrieve a word therefore depends at least in part on how strongly it is activated by its successor. We propose that backward associations from following words to preceding words are learned by retrodiction of recently encountered words from the words that follow, i.e., using $word_j$ to predict and retrieve $word_i$. Just like forward associations are predicted to track log forward conditional probability, backward associations are predicted to track log backward conditional probability, or $p(word_i|word_j)$. In either case, these probabilities reflect the surprisal of the outcome conditioned on the form cue that is used to retrieve it. Human learners are known to be sensitive to backward conditional probabilities, presumably because retrodiction allows the listener to fill in words they might have missed (Lieberman, 1963; Connine et al., 1991; Gwilliams et al., 2018), but the position of disfluencies other than repetitions is not predicted well by backward conditional probabilities (Schneider, 2018).

Inverse Base Rate Effect Because in HiTCH, an association from a cue to an outcome weakens (slightly) when the outcome occurs without the cue, we expect an Inverse Base Rate Effect: words that are frequent in contexts other than the current one should be less accessible given the current context as a cue (Ellis, 2006; Schneider, 2018). For this reason, frequent words should be less likely to be repeated when their probability in context is controlled.

Accessibility of the Future The word following the disfluency may also affect how many words are repeated. Since repetition is assumed to occur whenever the future is unavailable, future words that are easier to access are expected to favor shorter repetitions. Accessibility of a future word in HiTCH is increased by its log probability given the preceding word—i.e., forward conditional probability or $p(word_j|word_i)$—and decreased by high frequency in other contexts. Repetitions are therefore expected to be shorter before probable words, especially low-frequency ones.

Data and Analysis

One- and two-word repetitions were retrieved from Switchboard. We excluded all repetitions in which the interruption was fewer than three words into the utterance. The result is a sample of 8128 one-word repetitions and 1852 two-word repetitions. Note that this exclusion means that none of the repeated words are utterance-initial. This is crucial for testing the Initiation Effect because it rules out the possibility that speakers restart from good initiators simply because they restart from the beginning of the utterance. Restarting from a good initiator means restarting from something that *tends to* begin utterances in the speaker's experience but does not in the particular utterance being constructed. It therefore requires storing some indication that a word is a good initiator with that word's lexical representation.

Data were analyzed using generalized linear (logistic) mixed-effect models using the lme4 package (Bates et al., 2015) in R (R Core Team, 2019). The regression approach is chosen because HiTCH predicts the directions of the effects of the predictors in the regression model regardless of parameter settings but the relative contributions of these predictors to model weights depend on parameters like how much top-down weights recover when a word occurs utterance-initially. Because there is much investigator freedom in determining model-internal weights, while predictor values from corpora can be objectively measured, we use variables such as forward and backward conditional probability and initialness proportion as predictors of behavior rather than model-internal predictors. The empirical results then provide support for the model if the

measurable predictors that the model expects to account for the behavior do account for it, and if their effects are in the expected direction.

We included two important control predictors. One predictor implements the assumption that speakers restart production from the nearest syntactic constituent boundary (Kapatsinski, 2005). The location of the nearest major boundary in each repetition token (before $word_{-1}$, $word_{-2}$, or $word_{-3}$) was coded by a trained linguist blind to how many words were repeated following the Simpler Syntax framework (Culicover & Jackendoff, 2006).

The second important control predictor is word duration. We reasoned that, given the overall tendency for repetitions to be short (mostly a single word), speakers may be less likely to repeat long words compared to short words. Since duration correlates with frequency and predictability, even when segmental content is controlled (Seyfarth, 2014), we consider it an important predictor to include in the model. Mean durations of all $word_{-2}$'s (the word that the speaker may or may not repeat) were extracted from Switchboard.

Finally, random intercepts for the three words preceding the interruption of speech and the word that followed were also included. This is the most complex random-effects structure that allows the model to converge. These random effects are intended to account for other differences between individual words that might affect their accessibility or their likelihood of being repeated.

Results

While controlling for syntax and word duration, we observe the effects of the probability of the future (FCP) and the past (BCP) conditioned on the present, in the expected directions. Speakers tend to repeat the past when it is accessible from the present while the future is inaccessible. As expected, in both cases, non-directional, joint probabilities (bigram frequency) performed more poorly in model comparisons. The IBRE effect is evident from the finding that log frequency of an outcome has the opposite effect to that of its conditional probability (FCP or BCP). Initialness proportion effects indicate that speakers tend not to restart speech from poor initiators. This novel prediction of cue competition is illustrated in Figure 5, which shows a very strong relationship between top-down cue weights in the HiTCH model and the location of the speech restart.

Conclusion

We proposed a novel model of the effects of experience on execution of word sequences. In this model, item-to-item associations compete with top-down associations during learning. When a word's occurrences tend to be predictable, the model claims that it will become a poor initiator, and speakers would avoid initiating speech with it. This novel prediction—the Initiation Effect—is confirmed in a corpus study of speech reinitiation in repetition disfluencies. We show that speakers tend to reinitiate speech from words that, in their prior experience, frequently initiate utterances. With experience, highly predictive cues become increasingly sufficient and increasingly necessary to activate predictable upcoming items, turning a hierarchical plan into something more akin to an associative chain.

Table 1: Choosing to restart from the second ($word_{-2}$, positive) or first ($word_{-1}$, negative) word preceding the interruption. Logistic GLMM results. Subscripts are exemplified by [$_{-3}$ for_{-3} [$_{-2}$ the_{-2} [$_{-1}$ $speech_{-1}$, *uh, for the speech* $group_1$]]]. Brackets are syntactic boundaries.[1]

	b	*z*	*p*
(Intercept)	–2.13	–8.179	<.0001
Log FCP_1	–0.50	–4.342	<.0001
Log $Frequency_1$	0.32	2.753	.006
Log BCP_{-2}	1.16	8.964	<.0001
Log $Frequency_{-2}$	–0.60	–2.89	.004
Log $Initialness_{-1}$	–0.32	–3.094	.00197
Log $Initialness_{-2}$	0.89	6.078	<.0001
Log $Duration_{-2}$	–0.67	–7.263	<.0001
$Syntax_{-2}$	–1.35	–6.337	<.0001
$Syntax_{-3}$	–0.50	–2.04	.04

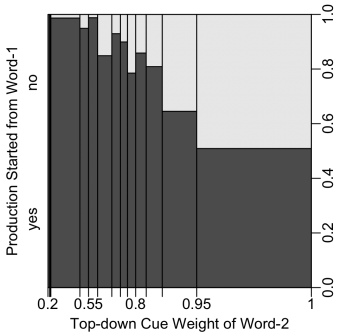

Figure 5: The likelihood of restarting production from the word that immediately precedes the disfluency ($word_{-1}$) vs. the word before it ($word_{-2}$) as a function of whether *word*$_{-2}$ is a good initiator (i.e., has a strong top-down weight). Restarts appear to be attracted to good initiators.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*, 814–823.

Bannard, C., Leriche, M., Bandmann, O., Brown, C. H., Ferracane, E., Sánchez-Ferro, A., ... & Stafford, T.

[1] Because $word_{-1}$ is the 'present' and is always accessible to the speaker, its frequency can matter only insofar as it can influence the top-down cue weight for $word_{-1}$ or accessibility of $word_{-2}$. As Figure 2 illustrates, it does not have a consistent influence on the former across different model parameter settings. We therefore don't have strong expectations about the role of this predictor and do not include it in the model.

(2019). Reduced habit-driven errors in Parkinson's Disease. *Scientific Reports*, *9*, 1–8.

Bates, D. M., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 58(5), 313–323.

Chaffin, R., & Imreh, G. (2002). Practicing perfection: Piano performance as expert memory. *Psychological Science*, *13*, 342-349.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, *30*(1), 234–250.

Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*, 244–74.

Culicover, P. W., & Jackendoff, R. (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences*, *10*(9), 413–18.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123-147.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149–195.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Fox, B. A., Hayashi, M., & Jasperson, R. (1996). Resources and repair. *Studies in Interactional Sociolinguistics*, *13*, 185–237.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 517–520). IEEE.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, *38*(35), 7585-7599.

Harmon, Z., & Kapatsinski, V. (2015). Studying the dynamics of lexical access using disfluencies. In *Disfluencies in Spontaneous Speech 2015* (p. 41–44).

Hieke, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech*, *24*, 147–60.

Kapatsinski, V. (2005). Measuring the relationship of structure to use: Determinants of the extent of recycle in repetition repair. *Berkeley Linguistics Society*, *30*(1), 481–492.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.

Levy, R. (2008). Expectation–based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Liberti III, W. A., Markowitz, J. E., Perkins, L. N., Liberti, D. C., Leman, D. P., Guitchounts, G., ... & Gardner, T. J. (2016). Unstable neurons underlie a stable learned behavior. *Nature Neuroscience*, *19*, 1665–1671.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172–187.

McKenzie, C. R., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*, 33–61.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments.*Journal of Memory & Language*, *35*, 477–96.

Osgood, C. E. (1963). On understanding and creating sentences. *American Psychologist*, *18*, 735–751.

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*, 54–74.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Reason, J. (1990). *Human error*. Cambridge University Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts

Rubin, D. C. (1977). Very long-term memory for prose and verse. *Journal of Verbal Learning and Verbal Behavior*, *16*, 611–621.

Schnadt, M. J. (2009). *Lexical influences on disfluency production*. Ph.D. Dissertation, University of Edinburgh.

Schneider, U. (2018). ΔP as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory* (ahead-of-print).

Schwartz, R. G., & Leonard, L. B. (1982). Do children pick and choose? *Journal of Child Language*, *9*, 319–336.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155.

Wickelgren, W. A. (1966). Associative intrusions in short-term recall. *Journal of Experimental Psychology*, *72*, 853–58.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits* (Technical Report No. TR-1553-1). Stanford University, Stanford Electronics Laboratories.

Yan, S., Mollica, F., & Tanenhaus, M. K. (2018). A context constructivist account of contextual diversity. In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp.1205–1210). Austin TX: Cognitive Science Society.