Exploratory investigations of SVO sentence production: evidence for category-specific interference effects

Sam Tilsen
Department of Linguistics
Cornell University

9/3/2020

Introduction

This paper reports several experiments on how quickly speakers can prepare and produce a subject-verb-object (SVO) sentence whose verb and arguments are provided as visual-orthographic stimuli. The key experimental manipulations were the relative timing of the subject (S), verb (V), and object (O) stimuli. The general motivation for investigating S, V, and O stimulus timing in this way is that in everyday settings, our attention to events in the world and to the entities which participate in them is unlikely to be simultaneous or evenly distributed in time; the effects of this nonsimultaneity on utterance planning and production may reveal important information about the mechanisms of syntactic organization.

For example, consider the following scenario, illustrated in Fig. 1A: you and your friends, Al, Bo, and Cam, are playing a game of tag. Al is chasing Bo, and is about to tag him, but you are unaware of this, because you and Cam are hiding around the corner. As you peer around the corner, you see Al (A1), then you see Al reach out his hand to tag someone (A2), then you see that it is Bo who is being tagged (A3). A simple description of the event is *Al tagged Bo*. Now imagine a different scenario (Fig. 1B), where again you and Cam are hiding around the corner. This time you see Bo run into view from around the corner (B1), then you see a hand tag him (B2), followed by Al, the owner of the hand (B3). You might say *Bo was tagged by Al*, but you can still say *Al tagged Bo*.

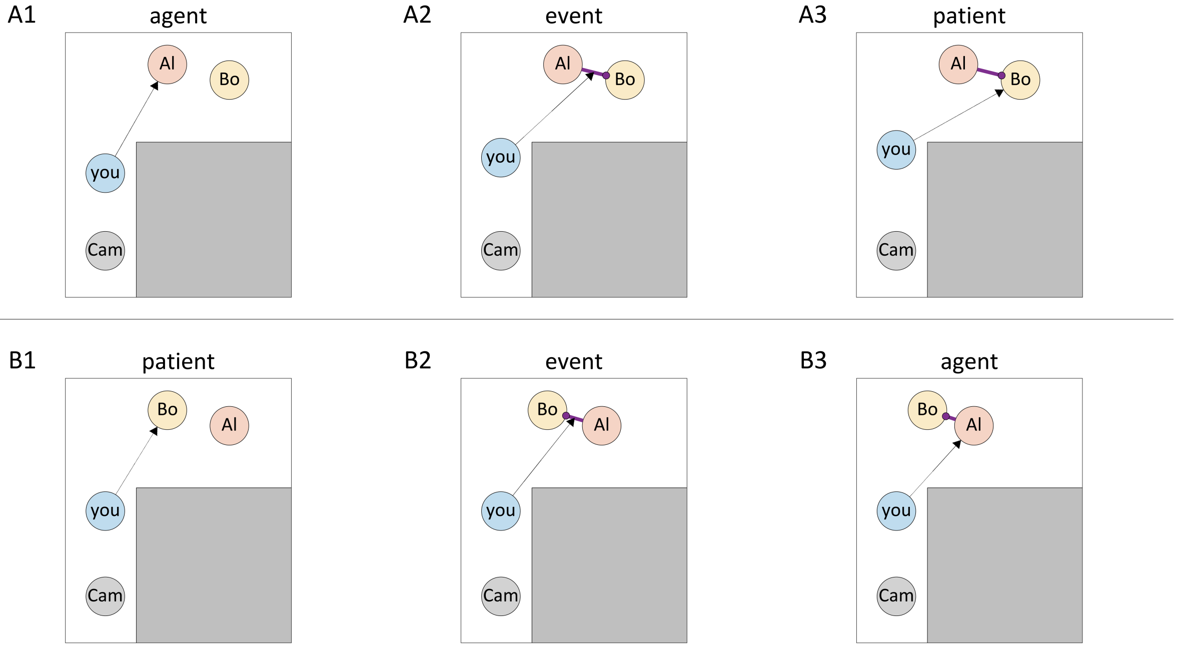


Fig. 1. Schematic illustration of two scenarios in which the timecourse of attention to arguments and an event differ. Panels A1-A3 and B1-B3 depict temporal sequences. Black arrows indicate what you see in

each time step; purple lines are tagging events. Top row: attention to agent precedes attention to patient. Bottom row: attention to patient precedes attention to agent.

What is the difference between these two scenarios? In the first case, your attention was drawn initially to the agent of the event (Al), then to the action, and then to the patient (Bo). In the second case, your attention was initially draw to the patient of the event (Bo), then to the action, and then to the agent (Al). The main question we are interested in here is: do differences in the timecourse of attention such as these have any effect on how quickly you can begin to produce the utterance *Al tagged Bo*?

We do not have to construct special hypothetical scenarios to obtain contexts in which the timecourse of attention to agents, patients, and actions is variable and nonsimultaneous. Even as events unfold in full view, our attention to those events and to the entities which participate in them can be unevenly distributed in space and time. This nonsimultaneity may have consequences for how speakers organize and produce utterances which describe events. By manipulating non-simultaneity of stimuli in a tightly controlled production task, we may learn something about how speakers organize syntactic and conceptual systems in generating sentences. We focus on production of simple subject-verb-object (SVO) sentences here, because SVO is the basic word order of English. The experiments reveal several remarkable phenomena, illustrated in Fig. 2 and listed below.

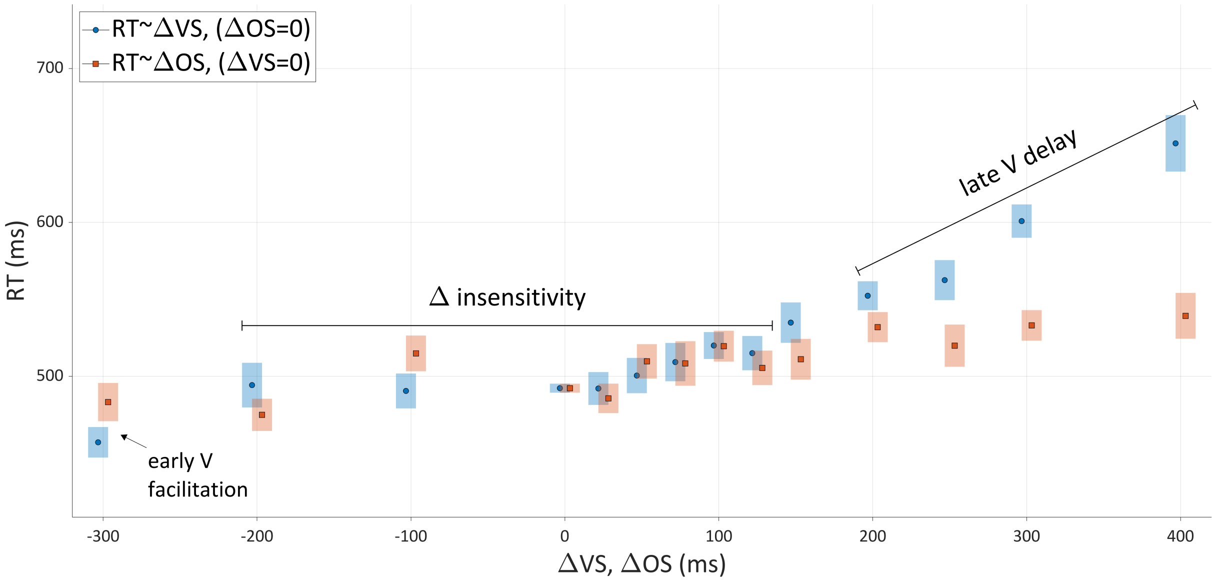


Fig. 2. Overview of experimental findings, showing mean RT as a function of O and V stimulus timing relative to S (ΔVS and ΔOS). Negative values of ΔVS or ΔOS mean that V or O stimuli precede the S stimulus; positive values mean that V or O follow S. Rectangles are 95% confidence intervals; horizontal coordinates of datapoints are slightly offset for visual clarity. RT is measured relative to the time at which the S stimulus appears, for reasons that we elaborate subsequently.

(i) Δ insensitivity: Response initiation is relatively insensitive to the timing of V and O stimuli relative to S stimuli (ΔVS and ΔOS), for asynchronies in the range of ±125 ms.

(ii) Late V delay: Response initiation is delayed substantially when the V stimulus occurs more than 150 ms after S; the initiation delay associated with late O stimuli is much smaller.

(iii) Early V and O facilitation: Response initiation is facilitated when V precedes S by more than 200 ms. To a lesser extent early O facilitates response initiation as well.

How do such effects arise? Model simulations conducted here show that the effects can be generated by combining two mechanisms: (i) activation-based initiation thresholds, and (ii) activation-dependent interference between syntactic systems. Furthermore, the specific form of the activation-dependence which is sufficient for generating empirical RT patterns is one in which a syntactic system that is *more* active has a *weaker* interference effect on other syntactic systems. This property may seem counter-intuitive, but makes sense when we consider what activation represents in the current context. This property of the interaction follows from a theoretical principle proposed in (Tilsen, 2019) whereby coupled syntactic and conceptual systems must reach a stable oscillatory state before they can be selected for production. The stability of this state can be characterized by the phase coherence of the relevant systems. Tilsen (2019) argued that when multiple conceptual systems are excited and begin to form resonant states with multiple syntactic systems, those resonant interactions interfere with one another due to differences in the phases and frequencies of the conceptual systems. Crucially, this interference diminishes over time as the conceptual-syntactic resonances stabilize.

The model presented here adopts a simplified version of this mechanism by reconceptualizing syntactic system activation as an index of the phase-coherence of coupled syntactic and conceptual systems. Hence the interaction property that greater activation is associated with weaker interference follows from the idea that activation of a syntactic system represents the extent to which it has formed a stable resonant state with a conceptual system: higher activation equals greater stability and less interference.

Why are the empirical findings important for syntactic theory? Let us assume that a scientific discipline should strive to obtain a theoretical understanding which can accommodate the widest variety of phenomena. The empirical phenomena reported here are important because most syntactic theories have no way of accounting for them. In that case, we should seek a theory which is more powerful.

Basic description of task and design

Here a brief description of the experimental task and design is provided (see *Methods* for more detail). On each experimental trial, the speaker produces a sentence. Each sentence consists of three words, a subject, a verb, and an object. The specific nouns which constitute the subject and object arguments, as well as the specific verbs, are conveyed via visual-orthographic stimuli. There are only three unique nouns (all monosyllabic proper names), and two unique verbs (both monosyllabic, past tense, experiencer verbs). The stimuli are presented in an inverted triangular arrangement on a computer screen (Fig. 3). Participants are instructed to interpret the stimuli in the upper left, upper right, and bottom vertices of the triangle as the subject (S), object (O), and verb (V), respectively. The S and O are never identical, and the specific names/verbs vary randomly from trial-to-trial, selected from the lexicon shown below.

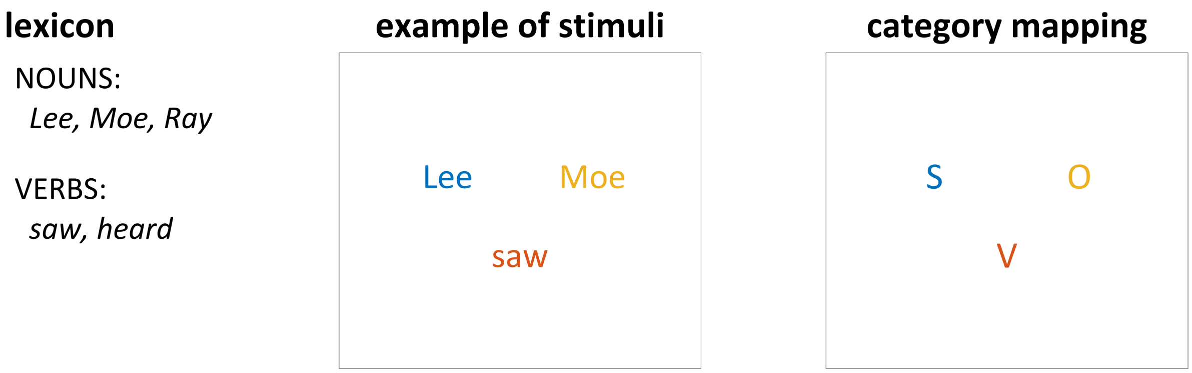


Fig. 3. Lexicon, spatial arrangement of stimuli, and mapping from spatial positions to syntactic categories. Color is added here to indicate mapping of stimuli positions to categories.

The S-V-O relative timing space

The main experimental manipulation was the relative timing of the appearance of S, V, and O stimuli. Note that once a stimulus appeared on the screen, it remained visible until the end of the trial. Consider the space of possible stimulus orderings, where *ordering* refers to temporal precedence relations. We adopt the following convention using curly-brackets, “{}”. For any two stimuli A and B, there are three possible orderings in which those stimuli can occur:

{A}{B}:	A precedes B
{B}{A}:	B precedes A
{AB}:	A and B occur at the same time

The stimulus {AB} is an unordered set of stimuli, and is indistinct from {BA}. For three stimuli—S, O, and V—there are 13 unique orderings. These can be organized in dimensions of pairwise relative ordering as in Fig. 4A, where rows are arranged by the ordering of O relative to S, and columns are arranged by the ordering of V relative to S. (For example, the row δOS = -1 contains orderings in which the stimulus set containing O immediately precedes the set containing S, and the row δOS = -2 is where the set containing O precedes V and V precedes S. Empty cells are impossible orderings.

An alternative representation is shown in Fig. 4B. The central ordering is the one where no members of the set are ordered. The inner circle of orderings are ones which impose partial ordering: two stimuli are simultaneous. The outer circle is sets with members which are totally ordered. Fig. 4C shows the 13 unique orderings as a function of time, aligned to the stimulus set (*stimset*) containing S.

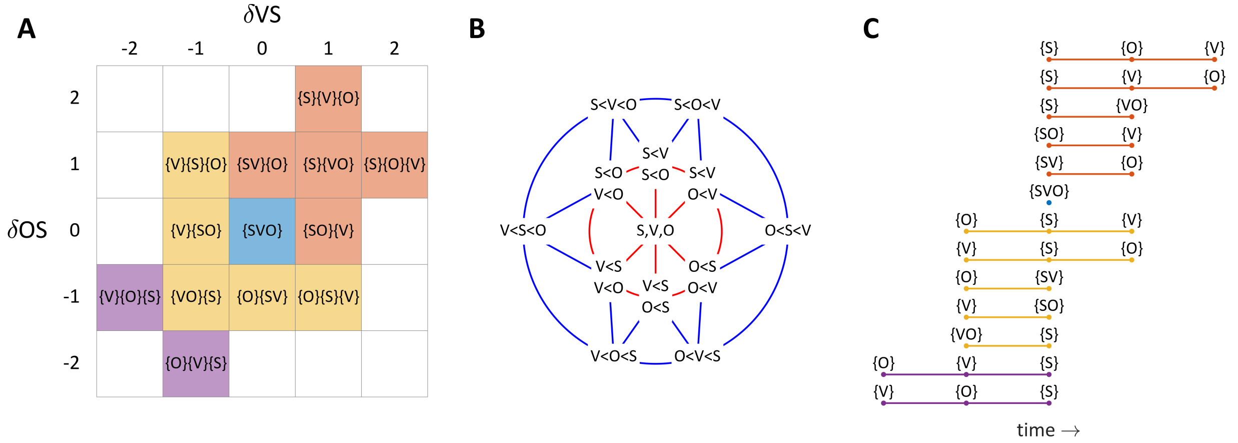


Fig. 4. Visualizations of stimulus ordering space. (A) tabular organization of orderings; (B) circular arrangement of orderings; (C) orderings in a temporal dimension, aligned to the stimulus set containing S. Colors in (A) and (B) indicate whether all stimuli are synchronous (blue) or whether S is in the first (red), second (yellow), or third (purple) stimulus set.

Whereas ordering is an abstract temporal relation, the experiments instantiate ordering with specific temporal intervals. A useful depiction of the relative timing of the stimuli can be provided in a two-dimensional space, which we refer to as a Δ-space, because its dimensions are relative timing measures (Δ-measures), i.e. variables which are differences of stimulus times t_S, t_V, and t_O. In this manuscript the character "Δ" can often be read as "the relative timing of" or "relative timing". There are six Δ-measures: ΔSV, ΔSO, ΔVO, ΔVS, ΔOS, and ΔOV. The measure "ΔVS" for example, can be read as "the timing of the verb stimulus relative to the subject stimulus," and is defined as ΔVS = t_V - t_S. A positive value of ΔVS indicates that the time of the V stimulus comes after the time of the S stimulus. Likewise, a negative value of ΔVS indicates that V comes before S, and a value of 0 indicates that V and S appeared on the screen simultaneously. There are three anti-symmetric pairs of relative timing measures:

$$\Delta\mathrm{VS} = -\Delta\mathrm{SV}$$
$$\Delta\mathrm{OS} = -\Delta\mathrm{SO}$$
$$\Delta\mathrm{VO} = -\Delta\mathrm{OV}$$

We will almost exclusively use ΔVS and ΔOS, where positive values indicate that V or O occurred after S. The main reasons for choosing these particular ones is that they express timing of V and O relative to S, and it turns out that the timing of the S is particularly influential in determining RT. (This is not surprising given that the sentence always begins with an S.) Nonetheless, the choice of any two measures is sufficient to describe the entire space of stimulus relative timing, because the third measure can be calculated from the other two, i.e.:

$$\Delta_{OV} = \Delta_{OS} - \Delta_{VS} = (t_O - t_S) - (t_V - t_S) = t_O - t_V$$
$$\Delta_{OS} = \Delta_{OV} + \Delta_{VS} = (t_O - t_V) + (t_V - t_S) = t_O - t_S$$
$$\Delta_{VS} = \Delta_{VO} + \Delta_{OS} = (t_V - t_O) + (t_O - t_S) = t_V - t_S$$

The two-dimensional SVO relative timing space (i.e. Δ-space) with dimensions ΔVS and ΔOS is shown in Fig. 5. Of the three experiments reported here, Experiment 1 sampled the widest area of Δ-space. Experiment 1 used inter-stimulus-intervals (ISIs) of 100, 200, and 300 ms, and imposed the constraint that if there were three stimulus sets, the ISI between the second and third sets was the same as the ISI between the first and second set. This constraint only comes into play when there are three unique stimulus sets. Observe that the 13 unique orders from Fig. 4 correspond to lines in the Δ-space of Fig. 5, or in the case of {SVO} a single point at the origin. The points are colored according to whether the stimuli were simultaneous (blue), or whether the S stimulus occurred in the first set (orange), second set (yellow), or third set (purple).

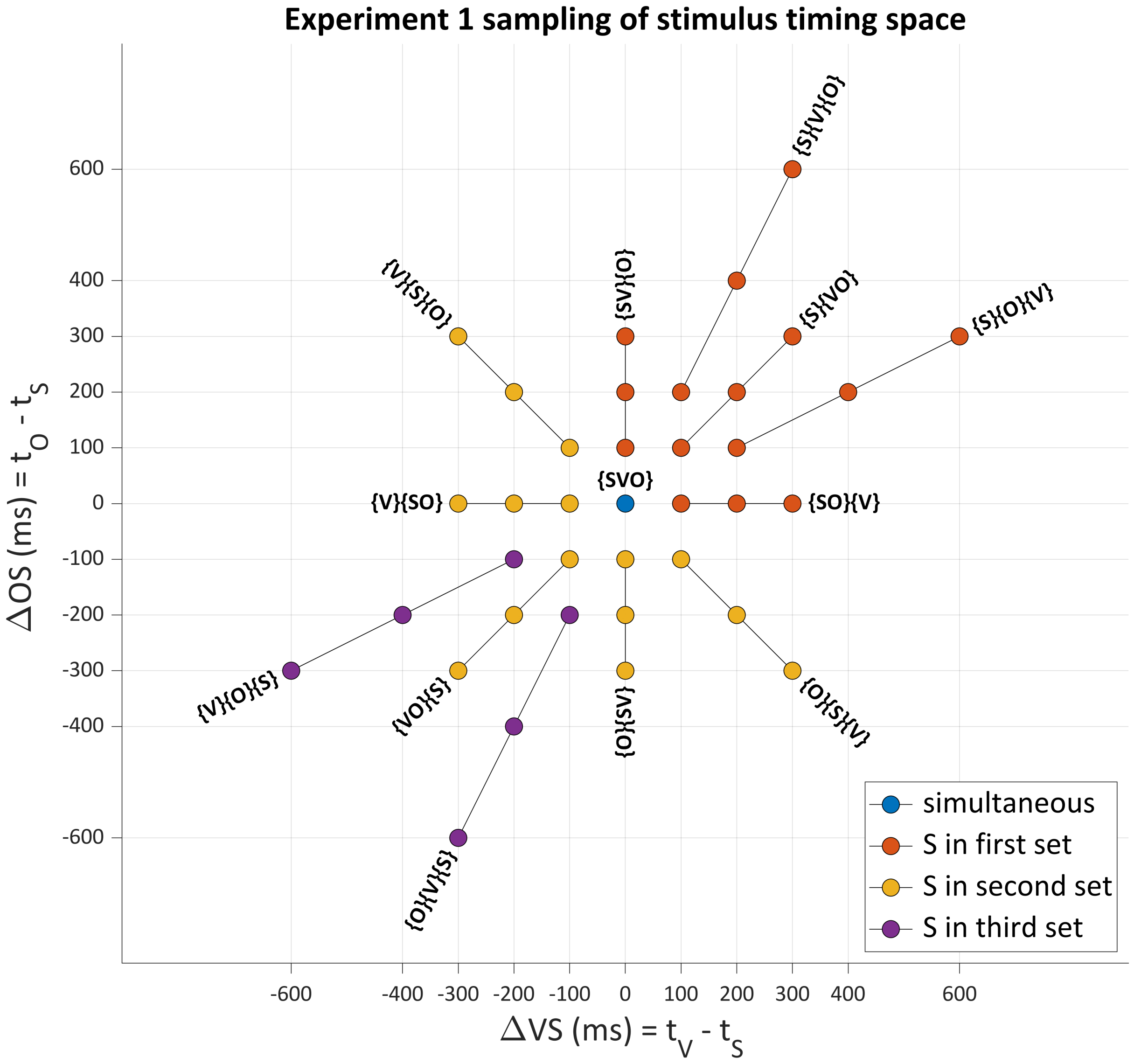


Fig. 5. Sampling of SVO stimulus timing space in Experiment 1. Lines connect timing patterns with the same ordering but different inter-stimulus-intervals (ISIs).

(Non)orthogonality of ΔVS and ΔOS. An important consideration for regression analyses of effects of stimulus timing on RT is that ΔVS and ΔOS (the predictor variables) are not orthogonal across the full Δ-space. Thus linear regressions of RT on the full set of Δ-values may incorrectly estimate the coefficients of Δ terms, because these terms are collinear (Tomaschek et al., 2018). In order to obtain more precise coefficient estimates, a subset of the space in which predictors are orthogonal is analyzed when appropriate. However, our main interest in the current context is not in estimating linear coefficients (i.e. linear statistical inference), but rather, in evaluating generative dynamical models of RT patterns. In that context, the nonorthogonality of the full space of predictors in unproblematic.

Reaction time definition

Consider the timecourse of events associated with an example trial in Fig. 6. In this example, the ordering is {V}{S}{O} and the ISI is 200 ms. Henceforth we refer to timing patterns as a combination of an ordering and an ISI, e.g. {V}{S}{O}/200. The timing pattern {V}{S}{O}/200 has ΔVS=-200 and ΔOS=200 ms. The trial onset (t_{ONS}, the appearance of the fixation cross) always precedes the first stimset by 750 ms. Because this interval is fixed, the timing of the first stimulus is highly predictable. The images at the bottom of the figure represent what the participant sees on the screen.

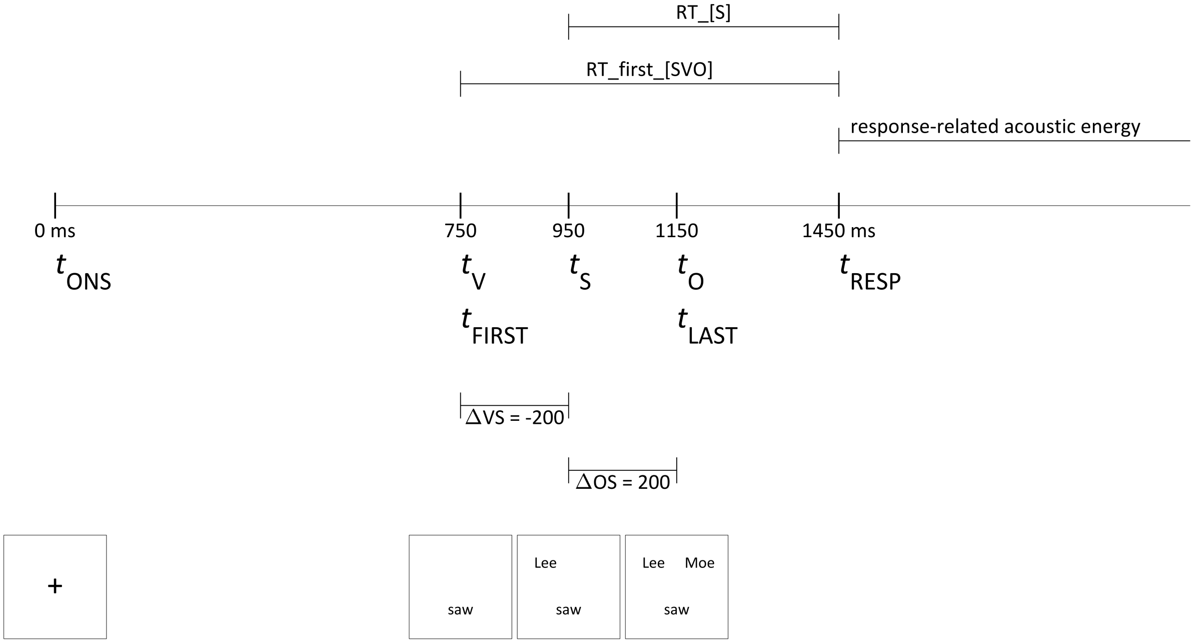


Fig. 6. Stimulus timing pattern {V}{S}{O}/200. Two alternative definitions of RT are shown, RT_[S] and RT_first_[SVO].

There are five events on each trial, which occur at times: t_{ONS}, t_S, t_V, t_O, and t_{RESP}—but not necessarily in that order. The trial onset t_{ONS} is defined as time 0. The event times t_S, t_V, t_O are the stimulus times. The event t_{RESP} is the first time of detection of response-related acoustic energy (i.e. *response detection time*). Our main interest is in constructing a dynamical model of response preparation and initiation which relates the stimulus event times (independent variables) and the response initiation time (a dependent variable). We can also define more abstract “events” such as the time of the first stimset, t_{FIRST} = min(t_S, t_V, t_O), and the time of the last stimset, t_{LAST} = max(t_S, t_V, t_O). On {SVO} trials, t_{FIRST} = t_{LAST}.

An important point to make here is that there is not one unique definition of RT, and so we have to make decisions regarding which RT measure(s) to analyze. All RTs we consider will be defined as t_{RESP} relative to a reference time, i.e. t_{RESP} - T_{REF}, and hence we have to choose a reference time. One reference time which seems obvious is the trial onset, t_{ONS}, in which case RT is the time from the trial onset to the response onset. Indeed, this definition of RT seems to require minimal intervention on our part given that t_{RESP} and all other events have already been defined relative to the trial onset. Instead of using trial onset as a reference time, we could alternatively use the first stimulus, i.e. T_{REF} = t_{FIRST}. This is equivalent to using T_{REF} = t_{ONS} as far as our models are concerned, because there is a constant offset between t_{ONS} and t_{FIRST} (750 ms)—the two variables are perfectly correlated. We refer to this definition of RT as RT_first_[SVO], to indicate that RT is defined relative to whichever comes first, S, V, or O.

However, consider that there is a hard constraint that the participant cannot begin to produce the response until the S stimulus has appeared. This suggests that t_S might be a particularly relevant reference event for the processes which determine t_{RESP}, in which case we might choose to analyze a different RT measure: RT_[S] = $t_{RESP} - t_S$. Indeed, analysis of RT_[S] is consistent with our decision to adopt ΔVS and ΔOS as the coordinates of relative timing space—these Δ-measures and RT_[S] arise from transforming event times t_S, t_V, t_O, and t_{RESP} in an equivalent way on each trial, i.e. by subtracting t_S from each of them.

There are in fact many other ways of defining the reference time that we might consider. Eleven alternative definitions of RT are listed in Table 1, along with some examples of the reference time that is calculated from some specific stimulus event times. These definitions are distinguished by the function used to calculate T_{REF}. Here we consider references times which are defined with a single minimum, maximum, or identity function of a set (or subset) of stimulus times.

Table 1. Alternative definitions of RT

		[t_S, t_V, t_O] =			
name	**T_{REF} =**	[0, 0, 0]	[0, .1, .2]	[0, 0, .1]	[.2, .1, .0]
RT_first_[SVO]	min(S,V,O)	0	0	0	0
RT_first_[SV]	min(S,V)	0	0	0	0.1
RT_first_[SO]	min(S,O)	0	0	0	0
RT_first_[VO]	min(V,O)	0	0.1	0	0
RT_[S]	S	0	0	0	0.2
RT_[V]	V	0	0.1	0	0.1
RT_[O]	O	0	0.2	0.1	0
RT_last_[SVO]	max(S,V,O)	0	0.2	0.1	0.2
RT_last_[SV]	max(S,V)	0	0.1	0	0.2
RT_last_[SO]	max(S,O)	0	0.2	0.1	0.2
RT_last_[VO]	max(V,O)	0	0.2	0.1	0.1

The decision to analyze one measure instead of any other measures may seem arbitrary. Is there a principled way in which we can justify this choice? One rationale for choosing a particular RT measure is to assess which measure has the lowest variance in empirical data. Fig. 7 shows the empirical standard deviations from Exp. 1 of the measures in Table 1, arranged horizontally from lowest to highest variance. Note that standard deviations were calculated after outliers were excluded and after participant means were subtracted (see *Methods* for further information). RT_last_[SV] is the measure with the lowest variance. This RT measure takes t_{REF} as t_S or t_V, whichever comes later. The measures RT_[S] and RT_last_[SVO] are the measures with the next lowest variances. Note that the ranking of measures by variance is the same regardless of whether participant means are factored out.

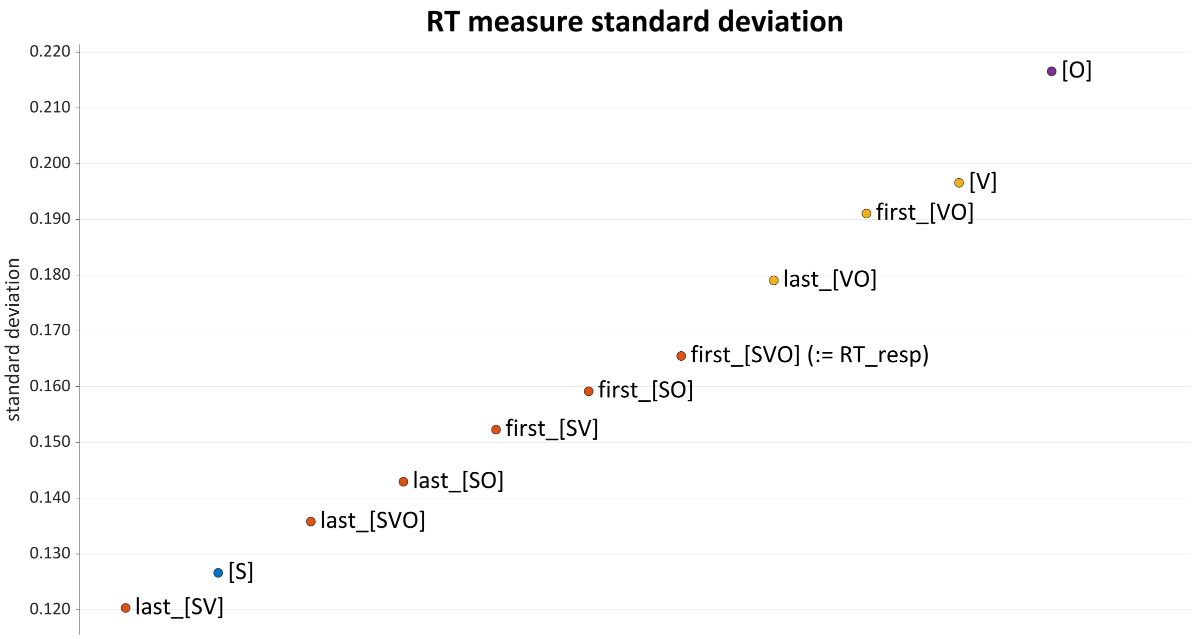


Fig. 7. Standard deviations of RT measures from Experiment 1. The horizontal axis is sorted according to the standard deviations of the measures, with increasing variance from left to right.

The four lowest-variance measures are the only ones in which t_{REF} will be t_S when S is in the last stimset. This may not be surprising, because knowledge of S is required for the participant to initiate the correct response. The time of the stimset that contains S provides a hard constraint on when the response can be initiated. The fact that RT_last_[SV] has the lowest variance is somewhat unexpected. For this measure the reference time is whichever comes later, S or V. The measures with the highest variances are ones in which the S stimulus does not contribute to the reference time.

Although standard deviation is one possible basis for rationalizing a choice of RT measure, its utility is based on the assumption that all RT distributions are approximately Gaussian. This assumption is false, as is evident from the distributions of several different RT measures are shown in Fig. 8. An alternative to standard deviation is entropy, a measure of the uniformity of the distribution (or the amount of information produced by sampling from the distribution). Table 2 shows entropies of the RT distributions and the mutual information of the RT and stimulus time distributions. The entropies of the RT distributions were calculated by counting the empirical RTs in 20 ms bins, converting the counts to probabilities, and applying the formula $\mathrm{H} = -\sum_i p(i)\, ln\, p(i)$. Note that to calculate entropies and joint entropies, we follow the convention that 0 × ln (1/0) = 0 (MacKay, 2003). In a loose sense, the entropies correspond to the amount of disorder in the distributions. RT_[S] is the most ordered distribution, i.e. the distribution associated with the least uncertainty/lowest entropy, despite the fact that RT_last_[SV] has a lower standard deviation.

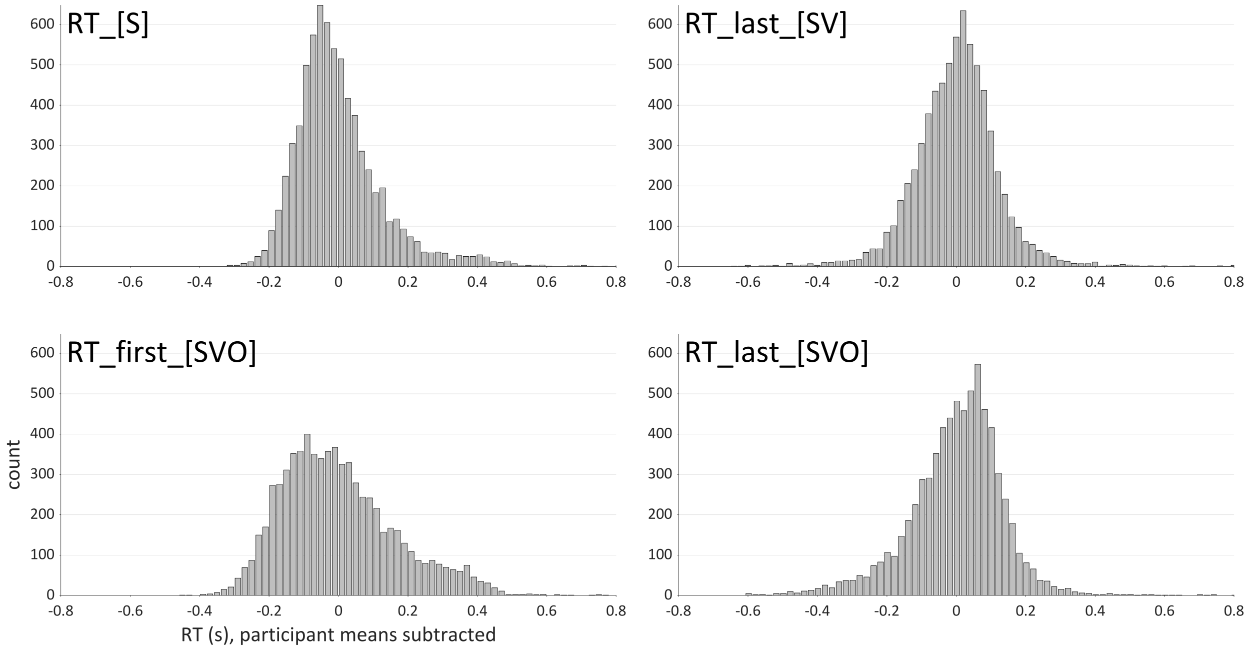


Fig. 8. Selected RT distributions. Participant means were subtracted from RTs, and values were rounded to the nearest multiple of 20 ms.

The mutual information I(RT;STIM) in Table 2 is a measure of dependence between RT and stimulus times. It is the amount of information about RT that is available from knowledge of the stimulus times. The mutual information is calculated as the information in the joint distribution of stimulus times plus the information in the RT distribution minus the information in the joint distribution of RTs and stimulus times, i.e.:

$$\mathrm{I}(RT; t_S, t_V, t_O) = \mathrm{H}(t_S, t_V, t_0) + \mathrm{H}(RT) - \mathrm{H}(RT, t_S, t_V, t_O)$$

where:

$$\mathrm{H}(RT, t_S, t_V, t_O) = \sum_{RT} \sum_{t_S} \sum_{t_V} \sum_{t_O} p(RT, t_S, t_V, t_O)\, ln\, p(RT, t_S, t_V, t_O)$$

Table 2**. Comparison of RT measure distributions**			
RT_meas	**std(RT)**	**H(RT)**	**I(RT;STIM)**
RT_last_[SV]	0.120	3.15	0.39
RT_[S]	0.127	3.13	0.38
RT_last_[SVO]	0.136	3.27	0.51
RT_last_[SO]	0.143	3.32	0.57
RT_first_[SV]	0.152	3.37	0.62
RT_first_[SO]	0.159	3.41	0.66
RT_first_[SVO] (=RT_ons)	0.166	3.47	0.71
RT_last_[VO]	0.179	3.59	0.84
RT_first_[VO]	0.191	3.65	0.90
RT_[V]	0.197	3.67	0.92
RT_[O]	0.217	3.79	1.03

If we were to use standard deviation as the sole criterion for selecting an RT measure for analyses, then we would choose RT_last_[SV] over RT_[S]. However, one problem with this choice is its interpretability: it puts us in a situation where the independent variables ΔVS, ΔOS depend on t_S while the dependent variable depends on max(t_S, t_V). On the other hand, consider that the RT_[S] distribution has the lowest entropy and the lowest mutual information with the stimulus times, which means that RT_[S] is most ordered/least uniform distribution, and that stimulus times contain less information about RT_[S] than they do about any other measure. This second property is desirable in order to maximize the extent to which Δ measures can predict variation in RT.

On the basis of the above, and also taking into account its greater interpretability, RT_[S] compares favorably to the other measures and hence will be our measure of choice for regression analyses. Nonetheless, the reader should keep in mind that this choice is somewhat arbitrary. This should not affect our interpretation of the mechanisms which determine response initiation. Fortunately, the optimization of dynamical models we implement below is not sensitive to the choice of RT measure. The models take stimulus times t_S, t_V, and t_O as input and output t_{RESP}. Both the stimulus times and RT are defined in the temporal coordinate of the dynamical model, where by definition $T_{REF} = t_{ONS} = 0$. Moreover, the entropies of the joint distribution of {ΔVS, ΔOS, RT} or equivalently of {t_S, t_V, t_O, RT}, are equal and independent of the choice of RT measure.

How to analyze the experimental data?

We should not assume that regression models are appropriate analysis instruments. Typically, we might be interested in a question such as “how does the relative timing of the S and V stimuli influence RT?”, where ΔVS is a predictor and RT is a response variable. There are a variety of problems with such an approach.

First, our relative timing variables ΔVS and ΔOS are highly correlated, which induces collinearity in a regression analysis. Second, our choice of how to define RT is somewhat arbitrary. Third, what basis do we have for assuming that there are *linear* functional relations between predictors and response variables, or even particular nonlinear functional relations? Our use of regression is so habitual that we often fail to recognize that such relations are unlikely to obtain for the vast majority of systems that we investigate. To the extent that we *can* (but not necessarily *should*) use regression instruments, we may obtain functional relations between predictor and responses, where coefficients are merely approximations of how certain variables influence the outcomes of complex processes. Indeed, these coefficients can be highly biased even for fairly simple processes.

Instead of focusing on these approximations, we direct most of our focus to hypotheses regarding the dynamical processes themselves, and evaluate the ability of these hypothesized processes to generate data which are similar to our empirical data. This analysis-by-synthesis approach is crucial when we have reason to believe that the relations between dependent and independent variables are not well-described by known analytic functions.

A simplified dynamical model

Here we construct a simplified dynamical model of task behavior. The model is “simplified” relative to the model proposed in Tilsen (2019), in that it does not have an explicit mechanism for associating syntactic categories (S, V, O) to concepts (*Lee*, *Moe*, *Ray*, *heard*, *saw*). Instead, the simplified models assume that categories are correctly associated with stimuli. In all cases, the arguments (S, O) and event (V) are modeled as systems with activation variables that exhibit linear growth when the corresponding stimuli appear. As described above, these activation variables are understood to represent the stability of coupled oscillations of syntactic and conceptual systems (Tilsen, 2019), and when interactions between systems are included in the model, it is assumed that systems with greater activation have *weaker* effects on other systems.

Before optimizing the model to fit empirical data, we examine how well various parameterizations of the model can account for *hypothetical* RT patterns. Consider that it is obviously the case that RT patterns depend on absolute timing of stimuli (i.e. t_S, t_V, t_O). The analysis of hypothetical patterns shows us that if RT patterns also depend on the *relative* timing of stimuli (i.e. ΔVS and ΔOS), then the model should include both category-specific parameters and asymmetric interactions between systems. This finding is highly relevant to interpreting empirical data, which unambiguously exhibit dependence on relative timing of stimuli. We can thus infer that any sufficient model must distinguish S, V, and O systems and must allow them to interact in asymmetric ways.

A key ingredient in all of the models is an *initiation criterion*. The initiation criterion is state which must obtain before a response can be initiated. This criterion is modeled as a function of system activations, and determines RT. We will consider how various definitions of the initiation criterion constrain what types of RT patterns (i.e. relations between independent variables and RT) can occur. Crucially, we show that there is a class of RT patterns which cannot be generated without allowing for S, V, and O systems to interact.

What is the participant doing?

What are participants in the experiment *doing* to accomplish the task? Consider that participants were instructed to respond as quickly as possible, but also to speak clearly. Feedback was provided to promote these goals (see *Methods* for further details). Fig. 9 depicts a hypothetical timecourse of the preparatory processes which occur on a single trial, with the timing pattern {V}{S}{O}/200. Recall that the onset of each trial was cued by a central fixation cross that remained visible for 750 ms. During this period of time the speaker has no knowledge of the particular arguments (S and O) or verb (V) that determine the response—hence we refer to this period as the uninformed preparation phase. Note that the beginning of the uninformed preparation phase may extend as early as the offset of the preceding trial. We will assume that each unique stimset (set of simultaneous stimuli) must be visually processed and its corresponding word forms must be linguistically and motorically organized, before those word forms can be produced, but we do not assume that all stimuli must be processed before the response can be initiated.

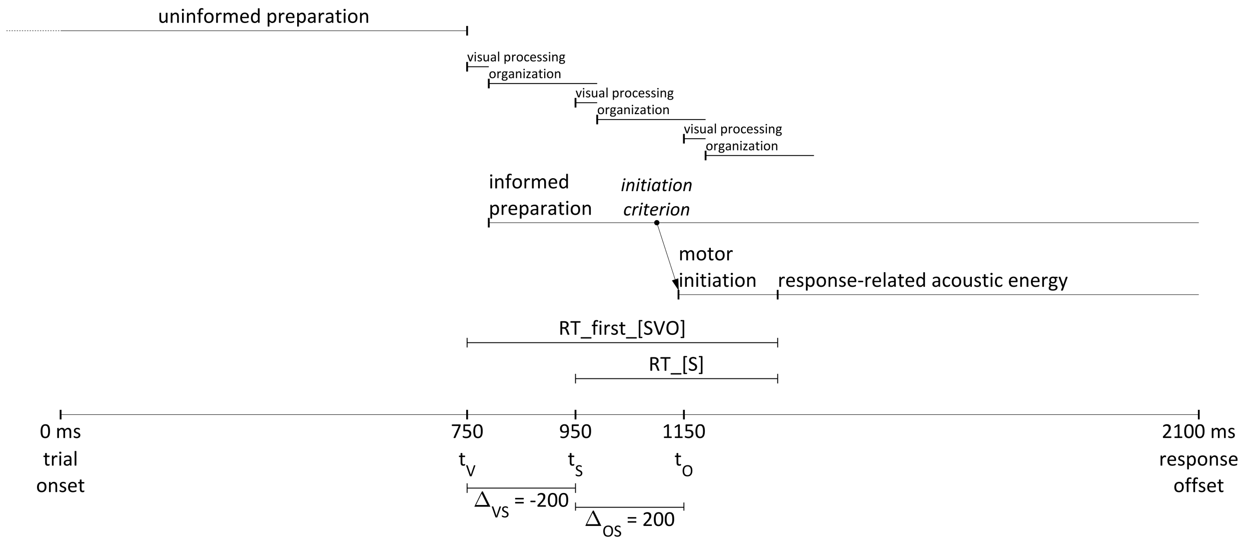


Fig. 9. Schematic illustration of preparatory processes for a single trial with the timing pattern {V}{O}{S}/200. The intervals which correspond to RT_first_[SVO] and RT_[S] are shown.

In order to construct various alternative interpretations of the cognitive processes which influence RT in the task, we adopt a model in which there are syntactic systems |S|, |V|, and |O|. Furthermore, we posit there are processes affecting the states of these systems which must occur before the response can be initiated. Specifically, there is an *initiation criterion* which must obtain before a speaker initiates the motor response associated with the stimuli. The initiation criterion will always be formulated in relation to the *states* of the syntactic systems.

In the simplified models developed here, there is no explicit representation of the conceptual and motoric systems associated with the stimuli. Nonetheless, it is assumed that there are conceptual systems which, when excited, give rise to meaning experiences that we can associate with names like *Lee*, *Moe*, and *Ray*, or with event frames like *saw* and *heard*; furthermore, the stimuli are also associated with sensory and motoric systems involved in the production of the word forms. The conceptual and motoric systems must be at least partly associated with syntactic systems before the response can be initiated. The mechanism of conceptual-syntactic "association" (or conceptual-syntactic binding) is hypothesized to be an entrainment of oscillatory systems through phase-coupling (Tilsen, 2019), and crucially, is extended in time. For example, to produce the sentence *Lee saw Moe*, the conceptual system associated with *Lee* must be associated with the |S| syntactic system.

To represent the timecourse of conceptual-syntactic association processes we employ state variables, one for each syntactic system. For simplicity, these syntactic system state variables are modeled as scalar variables with values in the range [0, 1]. A value of 0 corresponds to a state in which the conceptual-syntactic association process has not begun; a value of 1 corresponds to a state in which the association process has completed. To further flesh out the model, we address the following questions:

(i) How do syntactic system states change in response to stimuli?
(ii) What is the response initiation criterion?
(iii) What are the states of syntactic systems in the pre-stimulus period?
(iv) What is the nature of the interaction between syntactic system states?

(i) *How do system states change in response to stimuli?*
When a stimulus appears, the corresponding syntactic system will begin to associate with the motoric/conceptual systems that correspond to the stimulus, and the syntactic system will be organized in a way that prepares the associated motoric systems for execution. These association and organization processes occur over a period of time. The timecourse of this process is represented with an activation variable, one for each syntactic system, with values in the range [0, 1]. Furthermore, for simplicity, we assume that activation grows at a constant rate, once the stimulus is visible. Note that because there is an invariant mapping between the spatial locations of the stimuli and their syntactic roles, there is no ambiguity in which conceptual and motoric systems should be associated with which syntactic systems.

(ii) *What is the initiation criterion?*
Recall that the *initiation criterion* is a state that must obtain before motor initiation. There are many reasonable initiation criteria, but we will focus on some of the simpler possibilities and the relations between them. All initiation criteria we consider here are based on the idea that motor initiation can begin when all syntactic system activations have reached thresholds (or, a function of activation states reaches a threshold). Specifically, we will consider two types of initiation criteria: one type involves a *uniform threshold*, where the same threshold value applies to all syntactic systems, and the other type involves *category-specific thresholds*, where each syntactic system has a different threshold We will also assume that there is a fixed delay between the time that the initiation criterion is achieved and the time when the first acoustic evidence of motor initiation is observed. To represent the dynamics of response initiation, we construct a *response initiation gate* system with a binary state variable.

(iii) *What are the states of syntactic systems in the pre-stimulus period?*
Because activation of syntactic systems by definition represents stimulus-specific preparatory processes, we impose the constraint that syntactic systems have 0 activation in the pre-stimulus period. This constraint is probably incorrect, as it seems reasonable that syntactic systems will have some non-zero level of activation in anticipation of the stimuli. However, as we will see below, the effects of variation in initial activation can be equivalently modeled via activation growth rate and threshold parameters. (However, this equivalence only strictly holds in the absence of interactions).

(iv) *What is the nature of the interaction between syntactic system states?*
One possibility is that syntactic systems do not interact at all. In contrast, if interactions do occur, there are many reasonable ways to model those interactions. The theory underlying the model (Tilsen, 2019) holds that conceptual-syntactic association processes may interfere with each other. For example, if both |Moe| and |Lee| concept systems have been excited, then the association processes that must occur between these systems and |S| and |O| syntactic systems will interact, such that these processes may delay one another. Moreover, these interference effects are hypothesized be stronger earlier on in the association process, before stable, coherent systems states have been achieved. Thus the magnitude of the interference effect of system A on system B should be related to the activation variable of A, such that it is stronger when the activation of A is lower. To limit the space of possible models, we only consider linear interactions where the influence of A on B is linearly proportional to the activation of A. We often refer to these interactions as *interference*, and we impose the constraint that the interference is always destructive (it has a negative sign).

Model implementation
To facilitate subsequent exposition, we define the following variables and parameters, listed in Table 3. It is useful to distinguish between three categories of variables/parameters. First, there are state variables, which represent the time-varying states of systems in the model. Second, there are optimized parameters,

which are quantities that determine various model processes and outputs. These are the quantities that we may optimize so that the model can generate empirical patterns as closely as possible. We do not necessarily allow of these to vary in a given optimization, however. Third, there are fixed parameters/variables. These are quantities that provide information that is already known—in particular, a vector which represents whether each stimulus is visible at a given point in time. To optimize the models we include t_{RESP}, the empirically observed response time, which is associated with a particular stimulus event time vector T (= [t_S, t_V, t_O]). The cost function is the sum of squared errors between empirical response times and model generated response times.

Table 3. Variables and parameters of the dynamical models

State variables:		
x(t)	activation	a vector of syntactic system activation values: [x_S, x_V, x_O]
R(t)	binary	response initiation gate
Optimized parameters:		
x_0	activation	a vector of syntactic system initial conditions: [x_{S0}, x_{V0}, x_{O0}]
g	activation/second	a vector of syntactic system growth rates: [g_S, g_V, g_O]
τ	activation	a vector of syntactic system initiation criterion thresholds: [τ_S, τ_V, τ_O]
C	dimensionless	a matrix of syntactic system coupling strengths
m	seconds	response gating initiation delay
Fixed parameters/variables:		
T	seconds	a vector of stimulus event times: [t_S, t_V, t_O]
σ(t)	binary	a vector of stimulus states (0=visible, 1=not visible): [σ_S, σ_V, σ_O]
t_{RESP}	seconds	an empirically observed response initiation time

For convenience we use the indices i,j ∈ {S,V,O} to refer to any syntactic system. The general equation describing the time evolution of each syntactic system is:

$$\dot{x_{\iota}} = \sigma_i(t)\left[g_i + \sum_j C_{ij}\nu_j\right]$$

The variable $\nu_j = 1 - x_j$ is the difference between the maximum system activation value and the current value. The equation states that the rate of change of system activation is the product of the time-varying stimulus state $\sigma_{\mathrm{i}}(t)$ and the sum of the system intrinsic growth rate g_i and coupling forces. The stimulus state is either 0 or 1 (not visible or visible), so the equation holds that there is no change in activation of a system from the initial condition until the corresponding stimulus occurs. This ensures that coupling forces have no effect on a system until its corresponding stimulus has appeared. Notice that this equation is non-autonomous—the stimulus states σ(t) are time-dependent. The stimuli act to allow for activation to grow at a constant rate specified by g and to admit interaction forces specified by $\sum_j C_{ij}\,\nu_j$, where $\nu_j = 1 - x_j$. Note also that this equation does not reflect the present of an activation ceiling or floor; these are imposed externally in dynamical simulations. When there are no interactions between systems, i.e. C_{ij} = 0, the solution of the equation for one system is below, where x_{i0} is the initial activation of the system:

$$x_i(t) = x_{i0} + \sigma_i(t)g_i$$

Furthermore, the state of the initiation gate is defined as:

$$\mathrm{R(t)} = \mathrm{x_S(t)} \geq \tau_\mathrm{S} \;\wedge\; x_V(t) \geq \tau_V \;\wedge\; x_O(t) \geq \tau_O$$

This means that the initiation gate is open (R=1) when all three system activation values are at or above their thresholds; otherwise the gate is closed (R=0). For convenience we define δ_i as the first time that $x_i \geq \tau_i$, i.e. the time of threshold acheivement for a given system. The response initiation time is then:

$$t_{RESP} = \max_i(\delta_\mathrm{i}) + \mathrm{m}$$

The parameter m represents the duration of motoric processes that intervene between the time when the initiation gate is first opened and the first detection of acoustic energy associated with the response. Note that the model does explicitly represent the fixed period of time between when a stimulus appears and when the relevant information is available for syntactic organization (i.e. low-level visual processing). When there are no interactions between systems, δ_i can be analytically determined as:

$$\delta_i = t_i + \frac{\tau_i - x_{i0}}{g_i} = t_i + \frac{\epsilon_i}{g_i} \quad \text{where} \quad \epsilon_i = \tau_i - x_{i0}$$

The equation above states that the first time δ_i at which a given system reaches its threshold is the time that its corresponding stimulus appeared t_i plus the ratio of the difference between the threshold and the initial activation, $\tau_i - x_{i0} = \epsilon_i$, to the growth rate, g_i, which is expressed in units of activation per time. This relation only holds when there are no system interactions, and assumes that $x_{\mathrm{i}0} < \tau_i$, i.e. the initial activation is below the threshold, or equivalently $\epsilon_i > 0$. This latter assumption is only necessary for the |S| system, because otherwise the response could be initiated before the S word form is known. Note that as $\epsilon_i \rightarrow 0$, the time of threshold achievement becomes the stimulus time, i.e. $\delta_i \rightarrow t_i$.

As illustrated in Fig. 10, there are three ways to lower δ_i: increase the starting activation, decrease the threshold, or increase the growth rate. Indeed, parameter combinations associated with constant δ_i are planes in x_{i0}, τ_i, g_i space, as shown in on the right of Fig. 10. A consequence of this redundancy is that in optimizing models which lack interactions between systems, we can fix any two of these parameters; moreover, the planes of constant δ_i for a given system do not depend on the activation of the other two systems. However, when interactions are included in the model, planes of constant δ depend on the timing of all three stimuli, often in complicated ways.

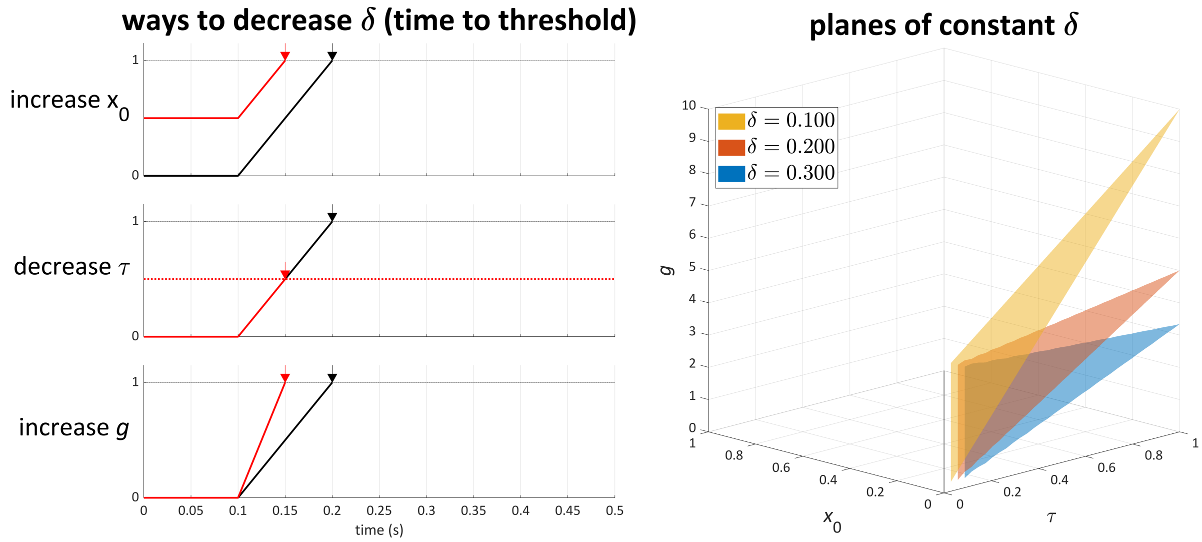


Fig. 10. Parameter combinations associated with constant δ_i are planes in x_{i0}, τ_i, g_i space

Model capabilities

Below we examine which model parameters must be unconstrained in order to generate different types of RT behavior. The models are listed in Table 4 and derive from varying: (1) whether τ parameters have a uniform value across systems or are category-specific; (2) whether g parameters are uniform or category-specific; and (3) whether there are interactions between systems. The motor initiation delay (m) was fixed at 0.050 s in all cases. Even though it is possible in some cases (when there are no interactions) to determine the model outputs analytically, a brute force numerical optimization procedure was used (see *Methods* for further detail).

Table 4. Models examined

models	**num. free params**	**threshold**	**growth rate**	**interactions**
τ1_g1	2	$\tau_S = \tau_V = \tau_O$	$g_S = g_V = g_O$	$C = 0$
τ1_g3	4		g_S, g_V, g_O	
τ3_g1	4	τ_S, τ_V, τ_O	$g_S = g_V = g_O$	
τ3_g3	6		g_S, g_V, g_O	
τ1_g1_C6	8	$\tau_S = \tau_V = \tau_O$	$g_S = g_V = g_O$	$C = \begin{bmatrix} 0 & c_{SV} & c_{SO} \\ c_{VS} & 0 & c_{VO} \\ c_{OS} & c_{OV} & 0 \end{bmatrix}$
τ1_g3_C6	10		g_S, g_V, g_O	
τ3_g1_C6	10	τ_S, τ_V, τ_O	$g_S = g_V = g_O$	
τ3_g3_C6	12		g_S, g_V, g_O	

The RT behaviors we use for assessing the model capabilities are simulated, hypothetical behavioral patterns, listed in Table 5. Note that these behaviors do not necessarily reflet empirical data, but rather are simulated behaviors which correspond to different ways in which stimulus timing might influence RT. The point of beginning our analysis with the stimulated RTs is to help us reason about what sorts of RT behaviors the models can generate, i.e. to test the capabilities of the model under different constraints. The stimulated RTs assume a processing time (μ) of 500 ms for each stimulus, which is close to the empirical mean RT across participants for the {SVO} condition in Experiment 1.

Table 5. Hypothesized RT patterns for testing model capabilities

	RT behaviors	**definition**
(A)	SVO-all	RT = μ + max(S,V,O)
(B)	SV-all	RT = μ + max(S,V)
(C)	S-only	RT = μ + S
(D)	SVO-any	RT = μ + min(S,V,O)
(E)	SV-any	RT = μ + min(S,V)
(F)	S_ΔVS	RT = μ + S + aΔVS

Behaviors (A)-(E) are ones in which RT depends on a min, max, or identity function of some set of stimulus times. Behavior (F) is different: RT depends on the timing of S and the *relative timing* of V and S, i.e. ΔVS. Here the parameter a = 0.25. Table 6 shows the mean absolute error (MAE) between optimized model predictions and the simulated RTs from each of hypothesized RT behaviors, for all of the Exp. 1 timing patterns. The cells with "*" indicate that the model generates the behavior with zero error (or more precisely, within the tolerance of the optimization, 0.001 s).

Table 6. MAE of models for simulated RT behaviors

		no interactions				**interactions**			
		uniform threshold		**specific thresholds**		**uniform threshold**		**specific thresholds**	
	behavior	**τ1_g1**	**τ1_g3**	**τ3_g1**	**τ3_g3**	**τ1_g1_C6**	**τ1_g3_C6**	**τ3_g1_C6**	**τ3_g3_C6**
(A)	SVO-all	*	*	*	*	*	*	*	*
(B)	SV-all	0.049	0.001	*	*	0.023	0.002	*	*
(C)	S-only	0.132	0.010	*	*	0.057	0.009	*	*
(D)	SVO-any	0.124	0.106	0.097	0.097	0.055	0.051	0.038	0.032
(E)	SV-any	0.135	0.126	0.124	0.124	0.074	0.074	0.054	0.073
(F)	S_ΔVS	0.099	0.045	0.042	0.042	0.041	0.021	0.013	0.013

First, the results of the optimizations show that category-specific thresholds are necessary to generate RT behaviors in which RT is determined by the maximum of a subset of stimulus times (SV-all) or a single stimulus (S-only). This is evident from the fact that the only the models with category-specific τ can generate behaviors (B) and (C) with minimal error.

Second, the optimizations show that no models can generate behaviors in which RT is determined by the minimum of a set of stimulus times. This is evident from the fact that no models obtained zero error for SVO-any (D) or SV-any (E). This deficiency can be viewed as a desirable feature of the models, since we do not expect such behavior to occur experimentally: for example, this would amount to participants initiating the sentence based upon whether either the S or V stimulus has appeared.

Third, the optimizations show that interactions and category-specific thresholds are useful for generating behavior (F), where the relative timing of V and S stimuli influences RT. No models were able to generate (F) with zero error. Nonetheless, the models with both specific τ and interference interactions provided substantially lower error than other models. This observation justifies our use of such models as tools to interpret empirical RT patterns.

Overview of experiments

The timing patterns examined for all three experiments are shown in Fig. 11. Exp. 1 investigated inter-stimulus-intervals (ISIs) of 100, 200, and 300 ms for all 13 orderings, and imposed the constraint that if there were three stimsets, the ISI between the second and third stimsets was the same as the ISI between the first and second stimsets. Note that this constraint only comes into play when there are three unique stimsets. The 13 orderings of stimuli correspond to lines in Δ-space, or in the case of {SVO}, a single point at the origin. The points are colored according to whether the stimuli were simultaneous (blue), or whether the S stimulus occurred in the first stimset (orange), second stimset (yellow), or third stimset (purple).

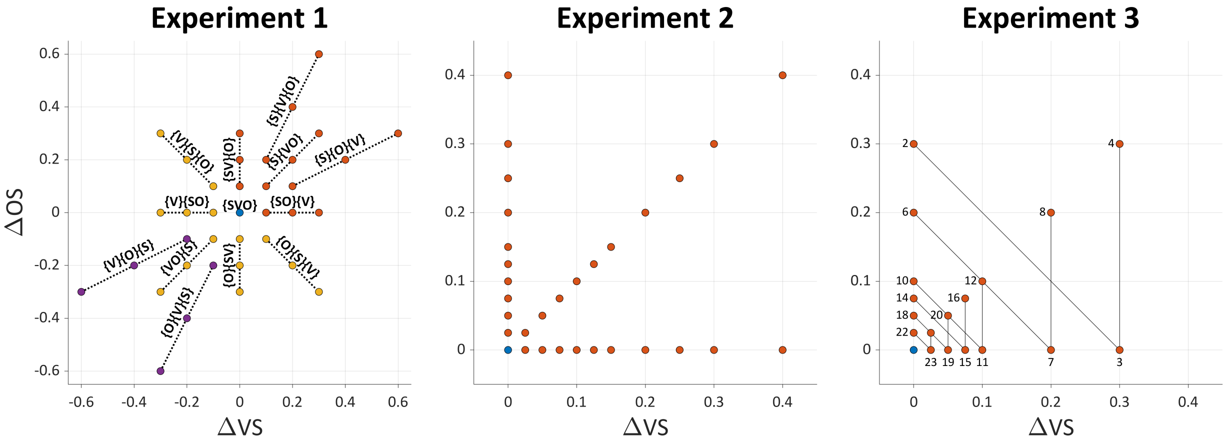


Fig. 11. Sampling of Δ-spaces in Experiments 1-3. The points are colored according to whether the stimuli were simultaneous (blue), or whether the S stimulus occurred in the first stimset (orange), second stimset (yellow), or third stimset (purple). In Exps. 1 and 2, stimulus timing patterns were sampled randomly; in Exp. 3, timing patterns were blocked and ordered by decreasing ISI.

Exps. 2 and 3 focused on orderings in which S is in the first stimset. In Exps. 1 and 2, the timing pattern for each trial was sampled randomly from the set of all timing patterns tested in the experiment. More precisely, trials were sampled randomly without replacement in blocks which contained the set of all possible timing patterns, obtained by crossing all orderings with all ISIs (see *Methods* for further detail). Participants were unaware of this blocking structure. In contrast, in Exp. 3, participants could easily infer the blocking of timing patterns: blocks were ordered by decreasing ISI and by stimulus orderings. There were 24 total blocks in Exp. 3 (the sequential order of non-simultaneous blocks is shown in the figure), and these were grouped in subsets of four, where each subset began with a simultaneous block {SVO}, and then had {SV}{O}, {SO}{V}, and {S}{VO} blocks, in that order (see Fig. 11). Hence the first block was {SVO}, the second was {SV}{O} with an ISI of 300 ms, and so on. The aim of the Exp. 3 design was to eliminate uncertainty regarding the timing and spatial location of information necessary to produce the sentence.

Experiment 1 Reaction times

Predictions

The RT patterns that we expect to see at the sampled points in Fig. 11 depend not only our model of task behavior but also on the RT measure that we choose. As motivated above, we focus primarily on RT_[S]. In some cases we also illustrate predicted and empirical patterns with RT_first_[SVO], which when compared with RT_[S] helps highlight the importance of the S stimulus. Fig. 12 shows heatmaps of predicted RT patterns for two different dynamical models, the parameters of which are detailed in Table 7. The S-only and SV-all models differ in that S-only has a non-zero threshold only for τ_S, while SV-all has non-zero thresholds for τ_S and τ_V. Consider that $\tau_i = 0$ entails that the initiation criterion is not sensitive to stimulus time t_i. Thus RT predictions of the S-only model are not directly sensitive to t_V or t_O (and accordingly, are not sensitive to ΔVS or ΔOS); this represents a null hypothesis that response initiation does not depend on relative timing of stimuli, only on the absolute timing of S.

Table 7. Model parameters and predictions.

response	**initial**			**Δ-space RT variation by measure**	
model	**activation**	**threshold**	**growth rate**	**RT_first_[SVO]**	**RT_[S]**
S-only	$x_S, x_V, x_O > 0$	$\tau_O = \tau_V = 0$ $\tau_S > 0$	$g_S, g_V, g_O = g$	minimum for ΔVS > 0, ΔOS > 0 (upper right quadrant)	constant for all ΔVS, ΔOS
SV-all		$\tau_O = 0$ $\tau_S, \tau_V > 0$		minimum for ΔVS = 0, ΔOS > 0 (positive ΔOS axis)	minimum area for ΔVS < 0 (left half-plane)

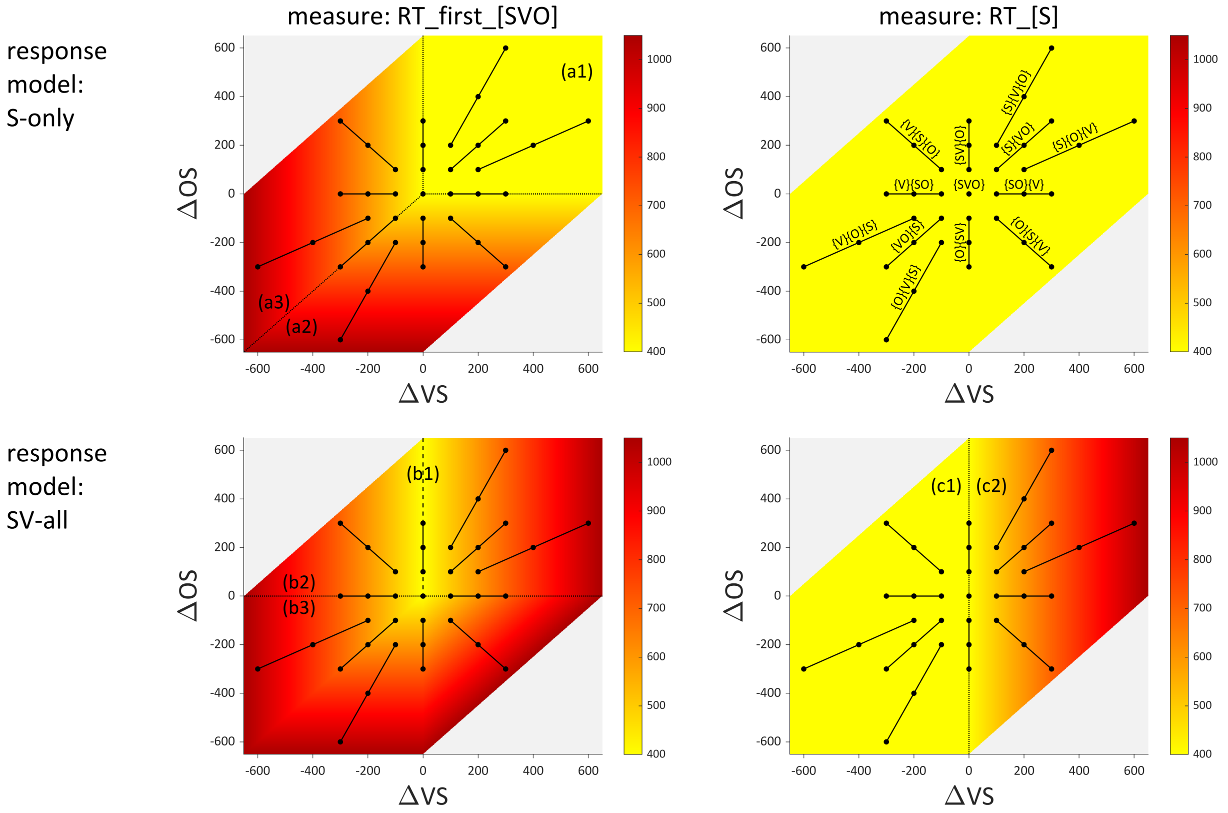


Fig. 12. Predicted RT patterns under S-only and SV-all response models, shown for the RT measures RT_first_[SVO] and RT_[S].

For the S-only model, where response initiation only requires processing of S, we can see that the prediction is that RT_first_[SVO] is minimal and constant for (a1) ΔVS>0, ΔOS>0 (the upper right quadrant), decreases linearly with ΔOS in (a2), and decreases linearly with ΔVS in (a3). The reason that RT_first_[SVO] varies in regions (a2) and (a3) is that the S stimulus is not in the first stimulus set in these regions. In contrast, the S-only model predicts constant RT_[S] over Δ-space: response initiation is a constant offset from the time of the S stimulus in this model, and the dependent measure RT_[S] is defined as the reaction time relative to the S stimulus.

For the SV-all response model, where response initiation requires processing of both S and V, we see that RT_first_[SVO] has a minimum value for orderings {SVO} and {SV}{O}, in which S and V are members of the first stimset. This corresponds to the positive ΔOS axis, labeled (b1). In the upper half plane (b2), the model-predicted RT_first_[SVO] increases linearly with the absolute value of ΔVS; in the lower half (b3) it increases in a more complicated way, depending on the distance and direction from the origin. In contrast, the SV-both model predicts constant RT_[S] in region (c1), the left half plane (ΔVS<0), and a linear increase with ΔVS in region (c2), the right half plane. Note that the predicted patterns under both models are somewhat simpler for the measure RT_[S] than for RT_first_[SVO].

Another way to illustrate the predicted effects of S-only and SV-all models is to plot them as surfaces, as in Fig. 13 below. These more clearly show the fjord-like structure of SV-both model predictions for RT_first_[SVO].

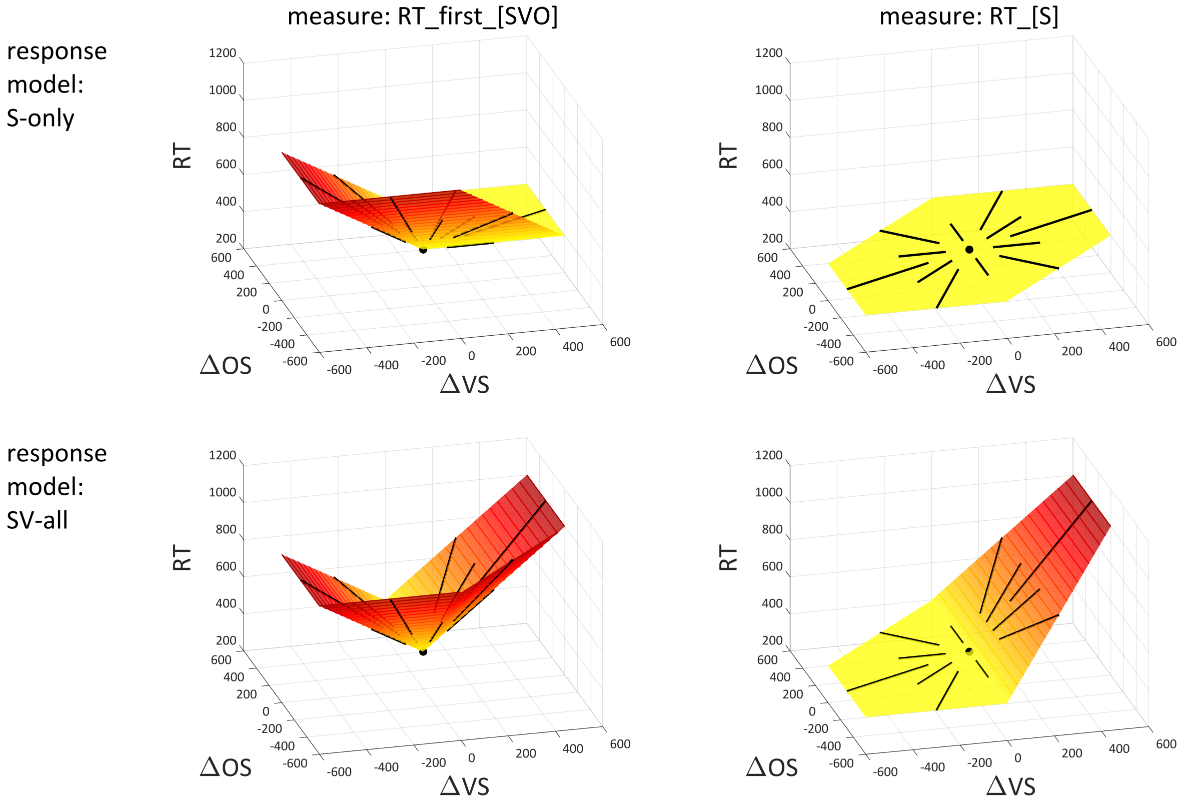


Fig. 13. Predicted RT surfaces under S-only and SV-all response models, shown for the RT measures RT_first_[SVO] and RT_[S].

Results

The results of Exp. 1 are more consistent with the SV-both model than the S-only model. Although RT depends strongly on S, it also depends on ΔVS and to a lesser extent on ΔOS. Mean RT_first_[SVO] and RT_[S] for each timing pattern are shown in Fig. 14. The colored intervals are ±2 standard error intervals. For purposes of exposition, argument orders are grouped as follows:

$\{S\}^0$: simultaneous (blue)
$\{S\}^1$: S in first stimset, excluding simultaneous (orange)
$\{S\}^2$: S in second stimset (yellow)
$\{S\}^3$: S in third stimset (purple)

When comparing RT_first_[SVO] and RT_[S], the groupings illustrate how the timing of S has a strong influence on RT. By definition, RT_first_[SVO] and RT_[S] are identical for the $\{S\}^0$ and $\{S\}^1$ groups. This is because the reference times for RT_first_[SVO] and RT_[S] are the same for all of the stimulus orderings in these groups. The key contrast is evident when comparing $\{S\}^2$ and $\{S\}^3$ groups: the spread of mean values of RT_first_[SVO] is much higher for these groups than it is for values of RT_[S]. This is not surprising, given the importance of S for response initiation. It also reinforces why choosing a low variance, low entropy RT measure (such as RT_[S]) is useful: we can more clearly see the effects of the relative timing of stimuli.

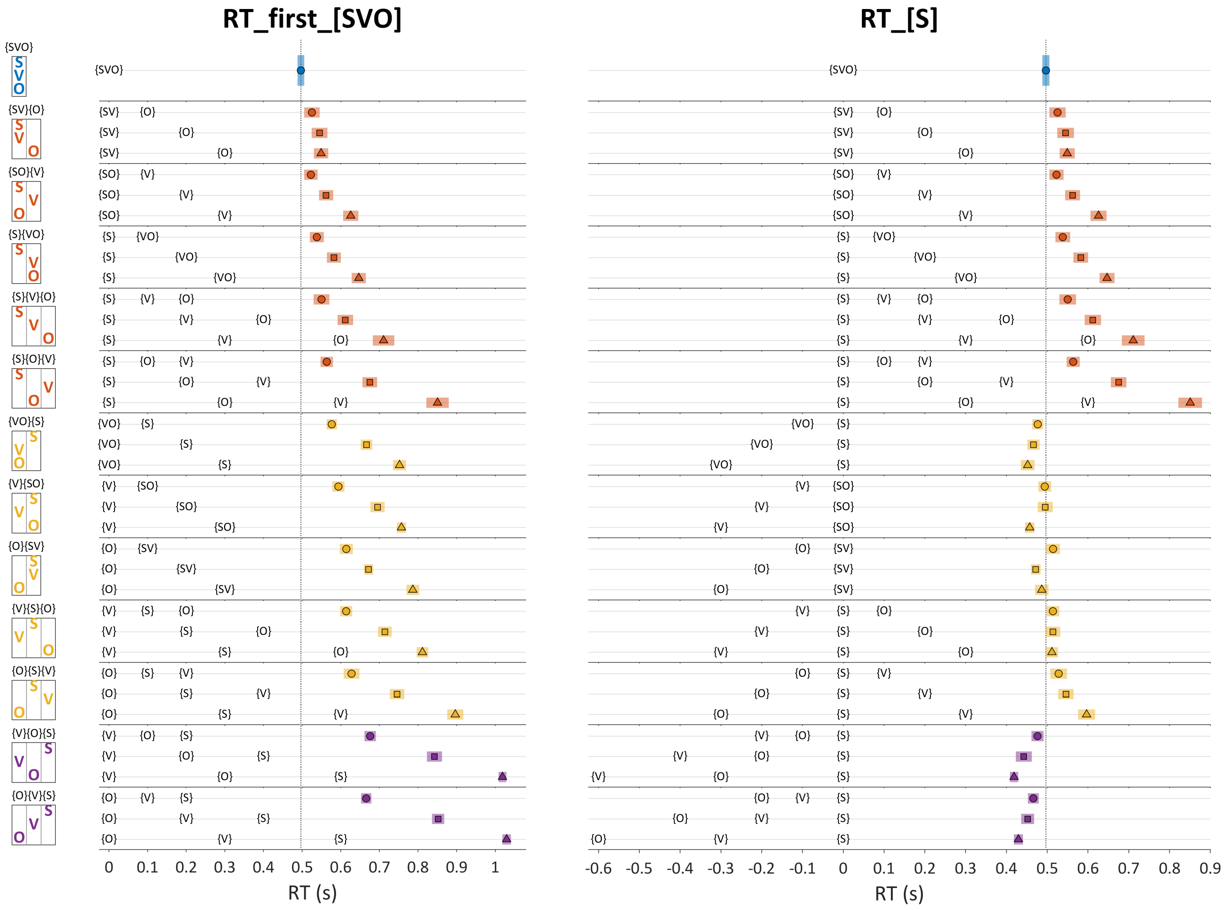


Fig. 14. RT measures by stimuli order and ISI. Left: RT_first_[SVO], 0 is time of first stimulus. Right: RT_[S], 0 is time of S. Vertical line is RT in simultaneous condition {SVO}. Stimulus orders are arranged vertically. Within each order, ISIs 100, 200, and 300 ms (circles, squares, triangles) are arranged from top to bottom.

Despite the strong dependence of response initiation on S timing, the response times are not consistent with an S-only model of initiation: delays of O or V relative to S delay mean response initiation, and early V and O facilitate response initiation. The patterns are also not consistent with a model in which S, V, and O are equally important for response initiation. We pursue further elaborations of these observations below.

Response initiation stimulus contingency

Is there evidence that any stimuli are absolutely required for response initiation? The S stimulus is indeed absolutely required: as we show below, no responses precede the S, and no responses occur before a reasonable estimate of when the speaker can be aware of S. Here we assume that a category-specific awareness of a visual stimulus is possible no sooner than 150 ms after the stimulus (Vanrullen & Thorpe, 2001), and so if a response is initiated earlier than 150 ms after a stimulus, it entails that it was initiated without category-specific processing of that stimulus. Also consider that the early left anterior negativity (ELAN)—an EEG/MEG signal that may reflect word-category related processing—peaks around 200 ms after a stimulus. Indeed, the shortest RT_S in Experiment 1 was 240 ms (see Fig. 15), which suggests a lower bound on the timecourse of processes that occur between the S stimulus and generation of acoustic energy associated with the S.

Regarding response initiation contingency on V and O, the data are more equivocal: on one hand, there are no timing patterns in which the mean response time precedes V or O stimuli, and delays of V and O relative to S tend to result in delays of mean response initiation times. On the other hand, there are timing patterns in which some proportion of responses were initiated before the participant could obtain a category-specific processing of the identity of V or O (i.e. pre-V and pre-O responses).

To illustrate the above, Fig. 15 shows distributions of RT_[S], RT_[V], and RT_[O]. The overall percentages of pre-V and pre-O responses were 2.0% and 4.2%, respectively. There were no pre-S responses. Given that the response initiation involves an unknown motor initiation delay, and that 150 ms is merely the minimum time required for stimulus awareness—(and does not necessarily include association/organization processes)—it is reasonable to infer that the percentages of trials initiated without fully organized V or O systems is higher than the percentages based on the distributions.

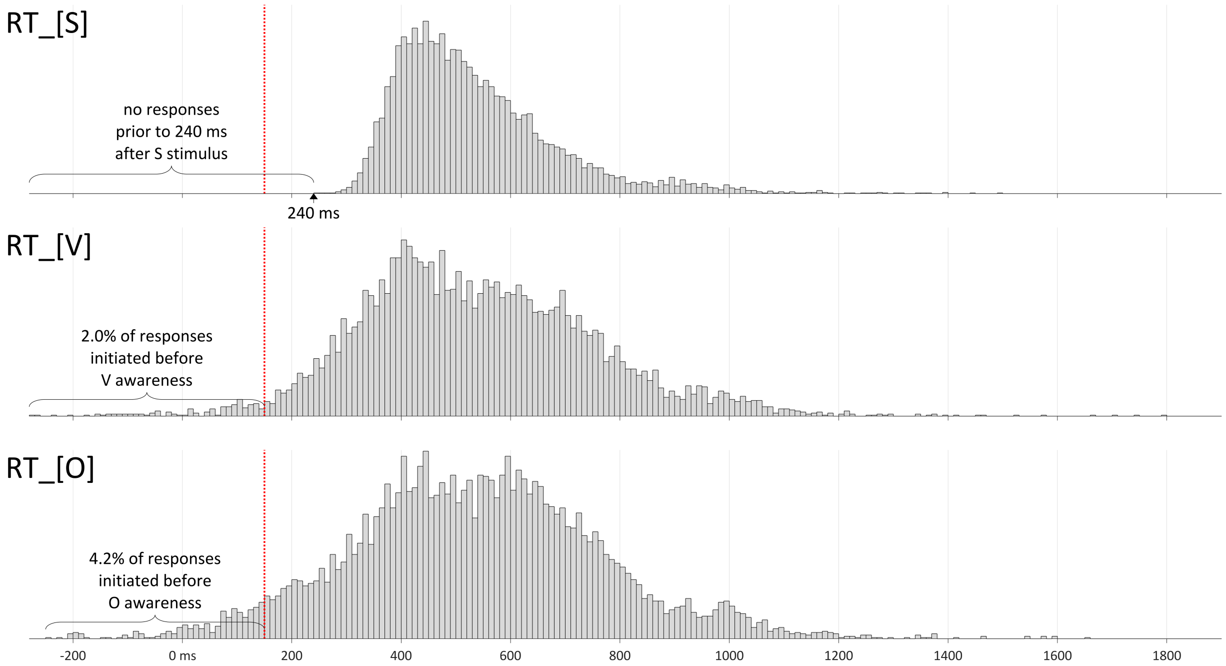
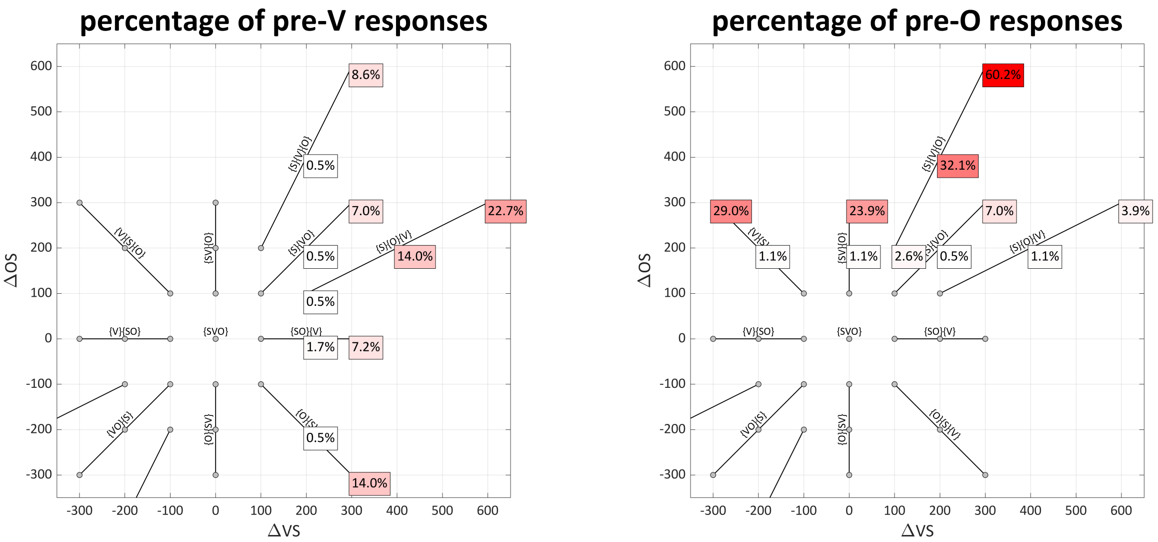


Fig. 15. Distributions of RT relative to stimuli. Vertical red line is the earliest possible time of category-specific awareness of the stimulus.

Furthermore, when the percentages of pre-V and pre-O responses are examined by condition (Fig. 16), we see that pre-V and pre-O responses are more frequent with large ΔVS and ΔOS. For example, in {S}{O}{V}/300, where ΔVS=600 ms, the response occurs before V awareness is possible on 22.7% of trials; in {S}{V}{O}/300, where ΔOS=600 ms, the response was initiated before O awareness on 60.2% of trials.



percentage of pre-V responses

percentage of pre-O responses

Fig. 16. Percentages of responses initiated before earliest awareness of V and O stimuli by timing pattern.

V and O stimulus delays induce response initiation delays

On average, delays of V and O relative to S result in longer RT_S, which suggests that the processes underlying response initiation typically involve organization of V and O systems. Looking within the {S}[1] group (Fig. 17), we can see several interesting patterns. Consider that both {SV}{O} and {SO}{V} orderings induce delays relative to {SVO}; yet the effect of ISI is stronger for {SO}{V} than for {SV}{O}—this suggests that V tends to be more influential for the initiation criterion than O: delay of V induces a greater increase of RT than delay of O. This is supported by the observation that the ISI effect in {S}{VO}, where both V and O are delayed, is quite similar to the effect in {SO}{V}. Similarly, notice that the ISI effect is greater for {S}{O}{V} than for {S}{V}{O}—this again follows from a greater role of V than O in response initiation.

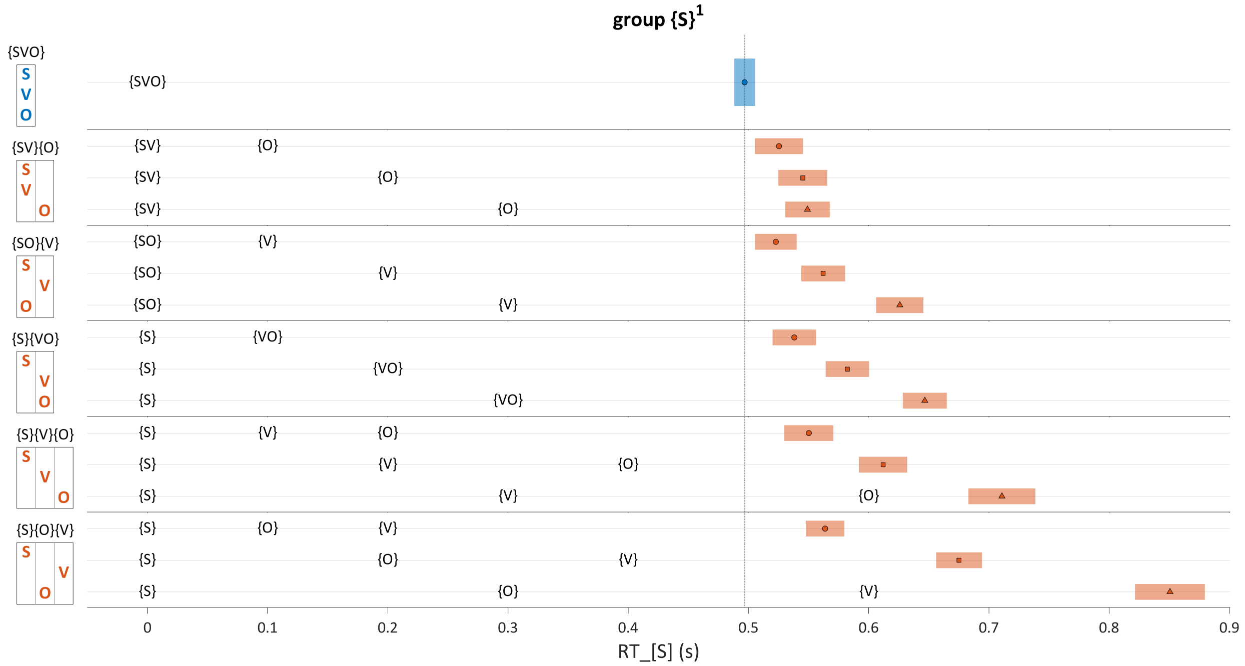


Fig. 17. RT_[S] for ordering group {S}^{1}, where S is in the first stimset. Group {S}^{0} is included for comparison. Filled areas are confidence intervals for the mean. Stimulus times are depicted for each timing pattern.

Response initiation is facilitated when V and O stimuli precede S

Examining RT_[S] patterns in the {S}^{2} and {S}^{3} groups sheds further light on the relative influences of V and O. Observe in Fig. 18 that when V precedes S (orderings {VO}{S} and {V}{SO}), there is a slight facilitation such that RT_[S] is faster. This facilitation is also present when only O precedes S (order {O}{SV}), although not quite as large. In addition, consider that the orderings {O}{S}{V} and {V}{S}{O} exhibit facilitation.

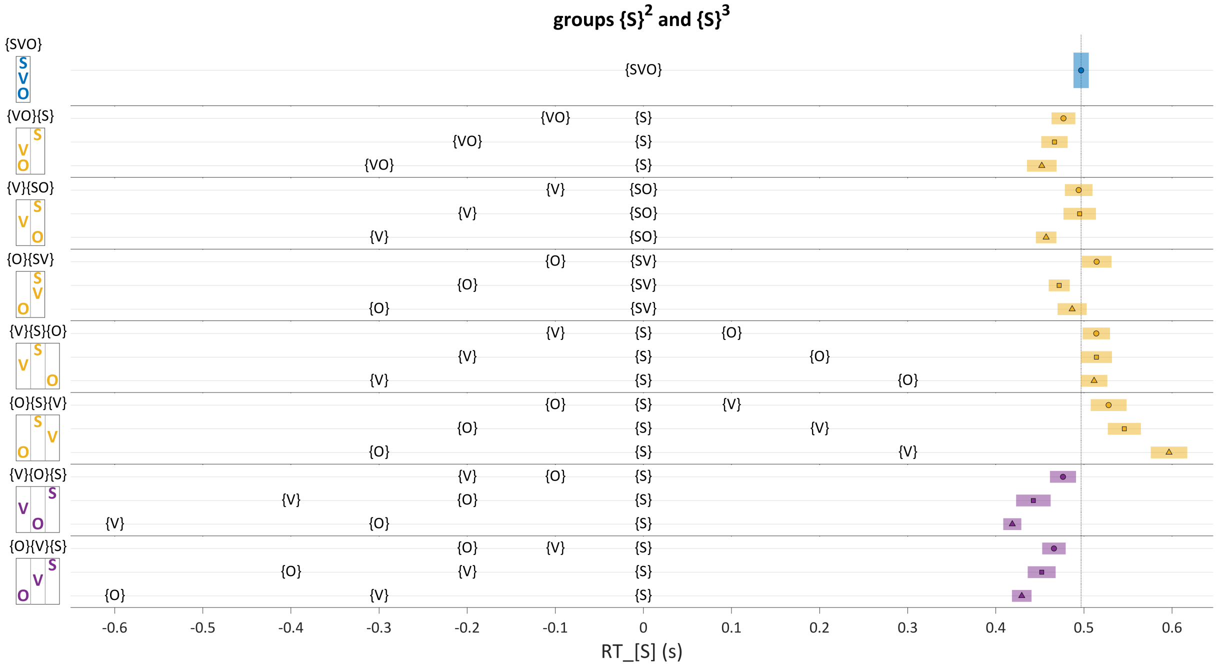


Fig. 18. RT_[S] for ordering groups $\{S\}^2$ and $\{S\}^3$, where S is in the 2nd or 3rd stimset. Group $\{S\}^0$ is included for comparison. Filled areas are confidence intervals for the mean. Stimulus times are depicted for each timing pattern.

Linear mixed effects models show ΔVS is more influential than ΔOS

The fixed effects coefficients of a mixed effects linear model of RT_[S] are shown in Table 8. This model considers only a subset of conditions in which ΔVS and ΔOS are orthogonal, which avoids the adverse effects of predictor collinearity. Note that a ΔVS-ΔOS interaction is included here, although this term does not significantly improve the fit.

Table 8. Fixed effect coefficient estimates

Name	**Estimate**	**SE**	**tStat**	**pValue**
(Intercept)	0.523	0.018	29.5	< 0.001
ΔVS	0.222	0.017	13.2	< 0.001
ΔOS	0.105	0.016	6.6	< 0.001
ΔVS:ΔOS	-0.029	0.059	-0.5	= 0.63

The main fixed coefficient estimates of the model can be interpreted as follows. The intercept is about 523 ms and is the model-predicted RT_[S] when all stimuli are simultaneous, i.e. {SVO}. Note that this differs from the grand mean of the data because it does not include random effects for subjects. The coefficients of ΔOS and ΔVS are the effects of delays of O and V relative to S. For example, in {V}{O}{S} order with an ISI of 300 ms, ΔVS = -0.600 s, and ΔOS = -0.300 s. These values are shown in the "predictors" columns in **Error! Reference source not found.** for three timing patterns: {SVO}, {V}{O}{S}/300, and {S}{V}{O}/300. The predicted effects on RT_[S] associated with each predictor are shown in the "effects" columns. For {V}{O}{S}, the combined effect of the predictors (not including the intercept) is -0.135 s. In other words, the model predicts that participants respond about 135 ms faster in the {V}{O}{S}/300 ms timing pattern than in {SVO}. Conversely, for {S}{V}{O}/300, the combined effect of the predictors is a 159 ms delay of RT_[S] relative to {SVO}.

Table 9. Examples of interpretation of fixed effect coefficients.

		{SVO}		{V}{O}{S}/300		{S}{V}{O}/300	
	coeffs.	**pred.**	**effects**	**pred.**	**effects**	**pred.**	**effects**
Intercept	0.523	1	0.523	1.000	0.523	1.000	0.523
ΔVS	0.222	0		-0.300	-0.067	0.600	0.133
ΔOS	0.105	0		-0.600	-0.063	0.300	0.031
ΔVS:ΔOS	-0.029	0		0.180	-0.005	0.180	-0.005
Sum of predictor effects			0		-0.135		0.159
Predicted RT			0.523		0.388		0.682

Notice that the coefficient of ΔVS is substantially larger than the coefficient of ΔOS: 0.222 vs. 0.105. This means that in the linear regression model, the timing of V relative to S is twice as influential on RT as the timing of O relative to S. This is consistent with our visual interpretations of the RT patterns above. To interpret the interaction term, we have to consider the units. The main fixed effect coefficients are in dimensionless units (proportional effects), whereas the interaction coefficients are in units of 1/s, because the predictors for these terms are in units of s^2. Thus the interaction coefficient cannot be directly compared to the main effect coefficients. Examining the effects of the interaction for the {V}{O}{S} and {S}{O}{V} conditions, we see that these are fairly small, about 5 ms. This value is also the largest predicted effect of the interaction for any condition, because the largest magnitude of this predictor value for ΔOS:ΔVS is ±0.180 s^2. Given the relatively small impact of the ΔOS:ΔVS interaction and its marginal significance, it is reasonable to focus our attention on the coefficients of a main-effects only model, in shown Table 10:

Table 10. Linear regression coefficients
of model without interaction

Name	**Estimate**	**SE**	**tStat**	**pValue**
(Intercept)	0.524	0.017	30.0	< 0.001
dVS	0.222	0.017	13.2	< 0.001
dOS	0.100	0.012	8.2	< 0.001

To better understand ΔVS and ΔOS effects, we visualize RT_[S] and linear and nonlinear model fits in Δ-space in Fig. 19. The nonlinear model includes ΔVS^2 and ΔOS^2 terms. The mean values for each timing pattern are represented by colors in Fig. 19A. The fixed effects of linear and nonlinear mixed effect models of RT with orthogonal predictors are shown in the heatmaps of panels (B)-(E). Notice that RT grows with both ΔVS and ΔOS in the linear model fit, but does so more quickly with ΔVS. This is because the coefficient of ΔVS is greater: 0.222 vs. 0.100. Fig. 19C shows the fixed effects predictions with the model intercept subtracted. This is equivalent to showing the difference between the model prediction and the predictions of an S-only model, because the intercept represents the mean value of the {SVO} condition (after random effects have been accounted for), and the S-only strategy predicts a constant surface in RT_[S] in Δ-space (see Fig. 13). Hence the positive (red) regions in Fig. 19C are Δ-values for which the model predicts a higher RT_[S] than the S-only strategy, and the negative (blue) regions are those where the model predicts a lower RT_[S].

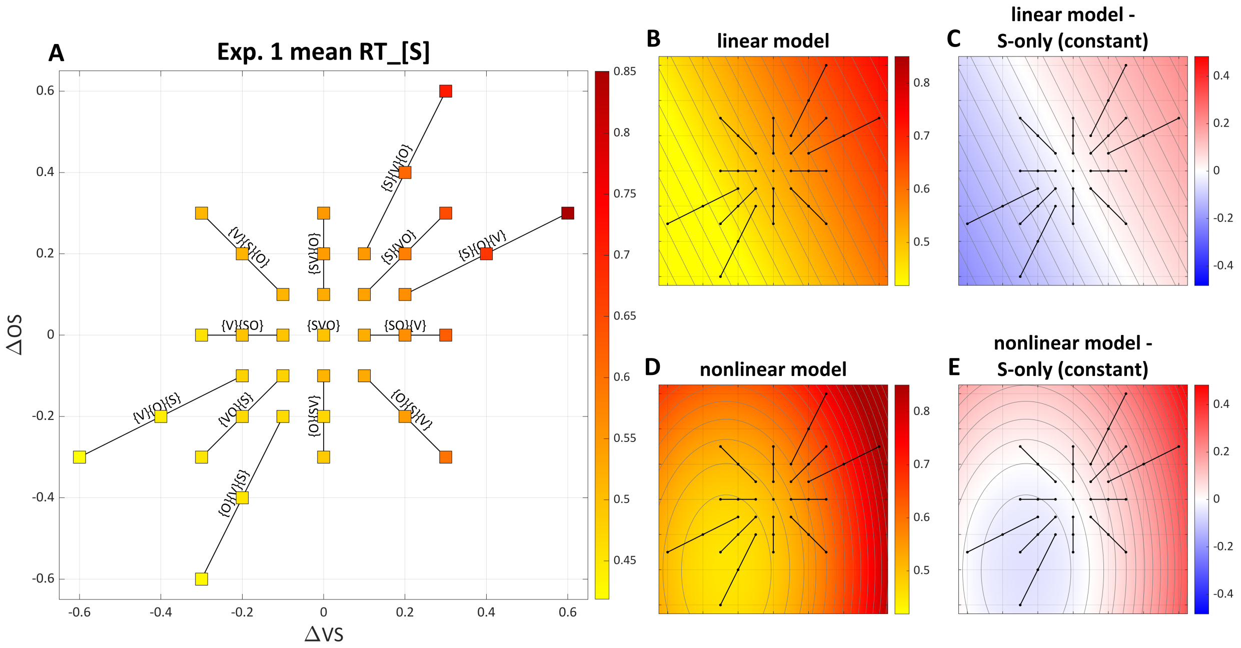


Fig. 19. RT_[S] for experiment 1 in Δ-space, along with linear and nonlinear model fits.

If we take the linear and nonlinear fits as decent approximations of the empirical data, the heatmaps show that RT_[S] is slower than predicted by the S-only model for large values of ΔVS and ΔOS (upper right quadrant), and faster than predicted for negative ΔVS and ΔOS (lower right quadrant). The quadratic terms of the nonlinear model (see Table 11 below) further indicate that the influence of ΔVS is stronger than ΔOS: the coefficient of the ΔVS^2 term is more than three times greater than the coefficient of the ΔOS^2 term. (Although note that the quadratic term coefficients are in units of 1/s.) Comparisons of mean absolute error (MAE) and Akaike information criteria (AIC) indicate that the nonlinear model provides a better fit, taking into consideration that it has more parameters.

Table 11. Coefficients of linear and nonlinear models

Model	**MAE**	**AIC**	**Intercept**	**ΔVS**	**ΔOS**	**ΔVS^2**	**ΔOS^2**
linear	0.066	-9297	0.524	0.222	0.100		
nonlinear	0.064	-9452	0.507	0.223	0.105	0.395	0.116

Inspecting the model errors by timing pattern (Fig. 20), observe that the linear model underestimates RT_[S] at extremal values of ΔVS, i.e. ΔVS = ±600 ms. The quadratic term of the nonlinear model corrects for these underestimates.

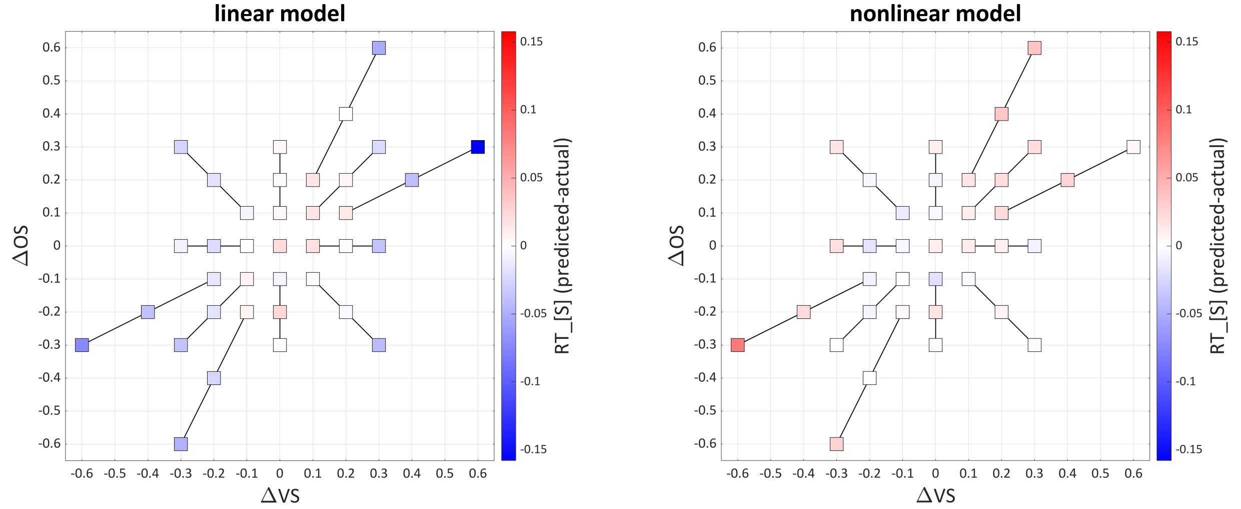


Fig. 20. Linear and nonlinear model errors by timing pattern.

There are two important phenomena which are apparent from the regressions analyses above. First, we see evidence of *late V delay*: response initiation is delayed substantially when the V stimulus occurs after S; to a lesser extent this is the case for O as well. Second, there is evidence of *early V and O facilitation*: V and O stimuli which occur before S are associated with lower RT_[S] than in the {SVO} pattern. This is illustrated by the location of the minimum in the nonlinear model heatmap in Fig. 19: the minimum value is located at large negative values of ΔVS and ΔOS, where the V and O stimuli precede S by a substantial amount.

Response initiation model

As an alternative to regression modeling, we analyze the performance of the dynamical models of response initiation under various parameter constraints. This analysis-by-synthesis approach shows that category-specific thresholds and interference are useful for understanding how stimulus timing influences response initiation. Furthermore, model optimizations suggest that interference effects are asymmetric, with S interfering with V and O to a greater extent than V or O interfere with S. Below we show the mean absolute error (MAE) of both the dynamical models and regression models, but it is important to keep in mind that these models cannot be directly compared: likelihood- and parameter penalization-based approaches to comparing models (such as AIC or BIC) cannot be applied to the dynamical models. Also, the parameters of the dynamical model are bounded in ways that the parameters of regression model are not. Ultimately, the dynamical models are much more appealing as analysis tools because they have a richer structure and a more limited predictive capacity than the regression models; we are particularly interested in what their optimized parameter values can tell us about the mechanisms responsible for organizing and initiating the response.

The dynamical models were optimized to fit the mean response initiation time for each timing pattern in Exp. 1, after subtraction of participant-specific intercepts (see Appendix: *Model optimization* for further detail). The linear regression model had fixed effects of ΔVS, ΔOS, and their interaction; the nonlinear regression model had fixed effects of fixed effects of ΔVS, ΔOS, ΔVS^2, and ΔOS^2; the dependent variable for the regression models was RT_[S] with participant-specific intercepts subtracted. The mean absolute errors of the models are shown in Table 12:

Table 12. Model error comparison

model	MAE	# params
τ3_g1_C0	0.0324	4
τ3_g3_C0	0.0324	6
linear	0.0192	4
nonlinear	0.0110	5
τ3_g1_C6	0.0102	10
τ3_g3_C6	0.0100	12

Not surprisingly, the models without interactions have relatively large errors. The linear regression, which has just four free parameters, exhibits lower error than the specific-thresholds and growth-rates model without interactions (τ3_g1_C0). The dynamical models with interactions slightly outperform the regression models. Unexpectedly, the models with category-specific growth rates do not substantially outperform their counterparts with uniform growth rates, despite having two additional parameters. This suggests that we can focus our attention on models with a uniform growth rate.

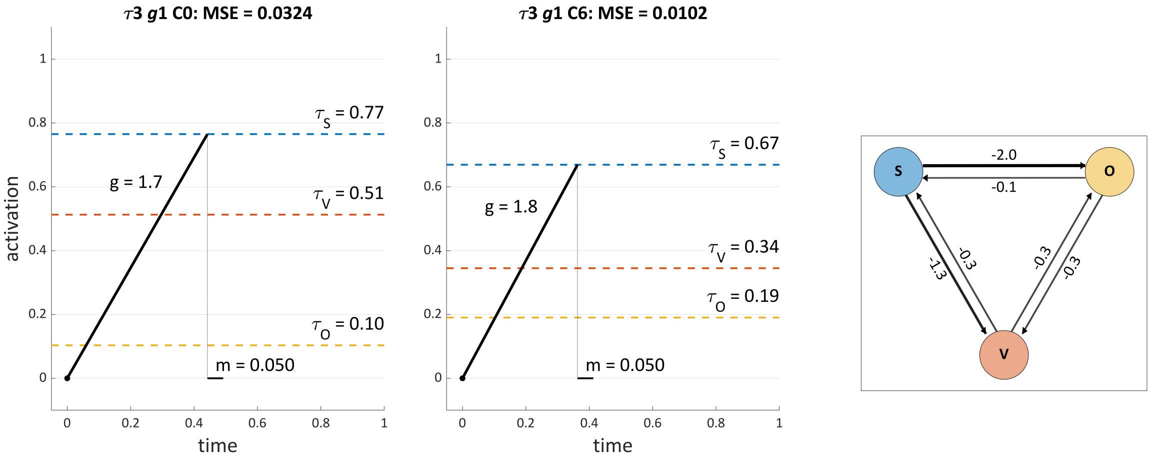


Fig. 21. Graphical depiction of parameters of optimized dynamical models. Left: diagonal lines show system activation growth for {SVO} ordering in the absence of interactions. Values of thresholds and growth parameters are labeled. Right: coupling interaction strengths for the model with interactions.

The parameter values of the dynamical models inform our conceptual understanding of how |S|, |V|, and |O| system states might interact and evolve over time. The parameters of models with uniform *g* and specific τ with interactions (i.e. τ3_g1_C6) and without interactions (i.e. τ3_g1_C0) are graphically illustrated in Fig. 21. The threshold values and growth rates are labeled in both panels. Note that a fixed motor initiation delay of 50 ms was imposed. The diagonal lines represent system activation growth for {SVO} ordering in the absence of interactions. Coupling strength parameters for the model with interactions (τ3_g1_C6) are shown on the right of the figure.

First, notice that in both models, the category-specific thresholds follow the order $\tau_S > \tau_V > \tau_O$. This is consistent with the empirical observations that response initiation is contingent on the S stimulus and depends more strongly on ΔVS than ΔOS. Second, notice that adding interference results in an increase in the growth rate, along with changes in the thresholds: τ_S and τ_V decrease, τ_O increases, relative to the model without interactions. Third, the strongest interference interactions are |S→O| = -2.0 and |S→V| = 1.3. This suggests that the process of organizing/preparing the |S| delays the organization/preparation of |O| and |V|, to a greater extent than vice versa.

The effects of including interference in the model can be better understood by comparing the model predictions and errors. Model-generated RT_[S] and errors relative to empirical means are shown in Δ-space Fig. 22. The model generated RTs are shown as heatmaps with equal-RT contour lines. The model without interactions can generate linear increases in RT_[S] in particular regions of Δ-space where ΔVS or ΔOS is large (these are determined by *g* and the relative values of τ_S and τ_V); outside of these regions the model can only generate constant RT_[S]. In other words, without interactions, the dynamical model predicts that in most timing patterns, early V and O will have no effect on response initiation since their thresholds are lower than the threshold for S.

The model with interactions generates more complicated, nonlinear variation in RT_[S] over Δ-space. The reason, which we examine below, is primarily that growth of V and O systems is slowed when the corresponding stimuli occur in the temporal vicinity of the S stimulus. The reductions of error in Δ-space that are obtained with interference are difficult to describe in a simple way. One clear difference is that error in the region of ΔVS, ΔOS < 0 is greatly reduced: without interactions the dynamical model overestimates RT values in this region.

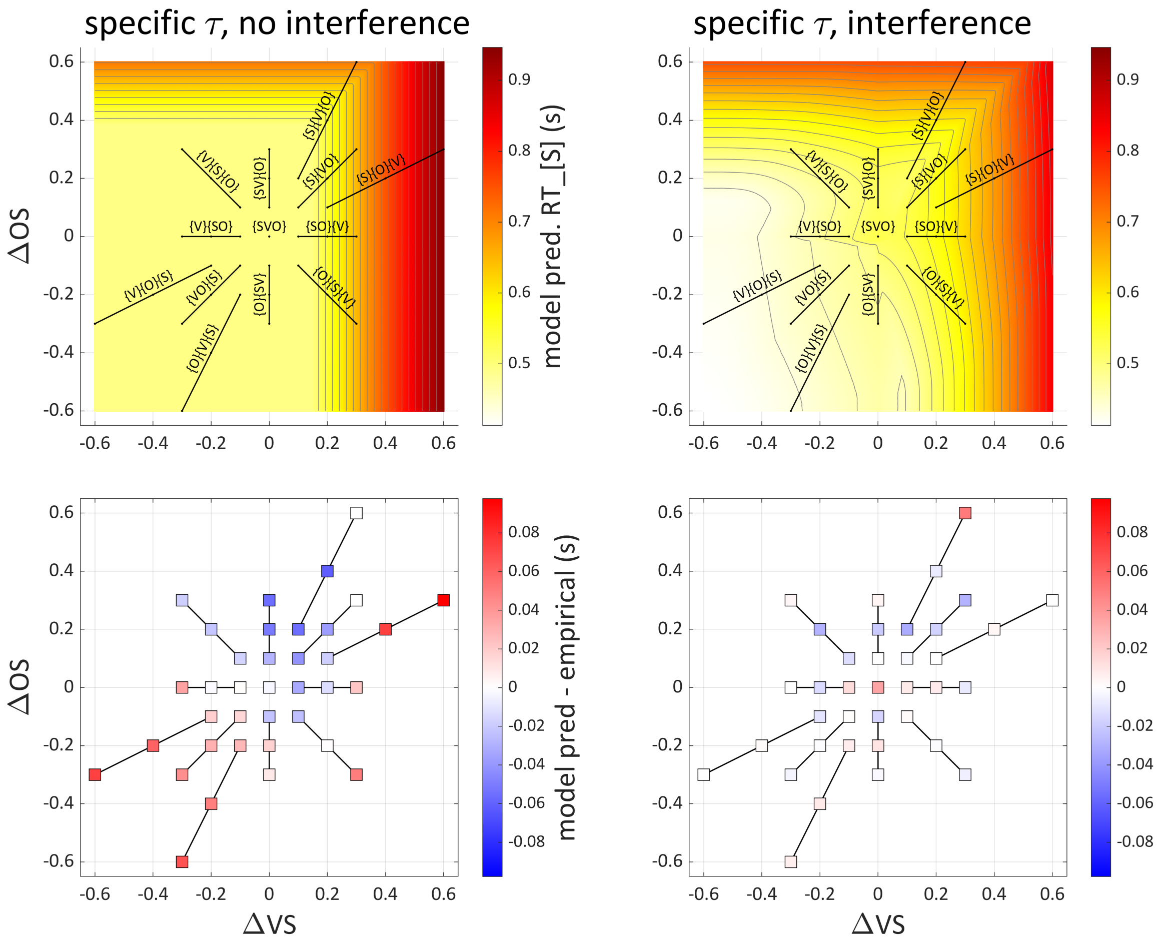


Fig. 22. Error by timing pattern for dynamical models with and without interactions. Top panels: heatmaps of model-generated RT with contour lines. Bottom: model errors for the timing patterns in Exp. 1; errors are indicated by colors.

To gain a better intuition for how the specific-τ model with interactions generates more accurate RTs, we examine the system activation time series for various stimulus timing patterns below. Fig. 23 shows system activations for {SVO} and the two timing patterns which represent extremal RT_[S]: {V}{O}{S}/300 and {S}{O}{V}/300. The simulations are aligned to the time of the S stimulus. In interpreting the activation time series, we distinguish between the intrinsic growth rate of a given system and the effective growth rate of that system: effective growth rate can vary over time and depends on both intrinsic growth and system interactions. Note that we refer to model systems between vertical bars, i.e. "the |S| system", and we refer to stimuli with the bare letter, i.e. "the S stimulus".

First, compare in Fig. 23 {SVO} and {V}{O}{S}/300 (maximally early V). Observe that the time for |V| and |O| to reach their thresholds is much longer in the {SVO} pattern than in the {V}{O}{S} pattern. This is because of interference exerted by |S| on |V| and |O|. The interference effects of |V| and |O| on |S| are smaller but not negligible. The last system to reach threshold in the {SVO} case is |O|, although all three systems achieve their thresholds around the same time. Overall, the interference effects delay RT in the {SVO} pattern compared to {V}{O}{S}. This is how the model generates the phenomenon of early V and O facilitation.

Next, compare {SVO} and {S}{O}{V}/300 (maximally delayed V). In {S}{O}{V}, the last system to reach threshold is |V|, simply because V comes late. Interference effects do not play a large role here, because |S| has already reached a high level of activation. In general, the effects of delayed V are larger than those of delayed O because |V| has a higher threshold than |O|.

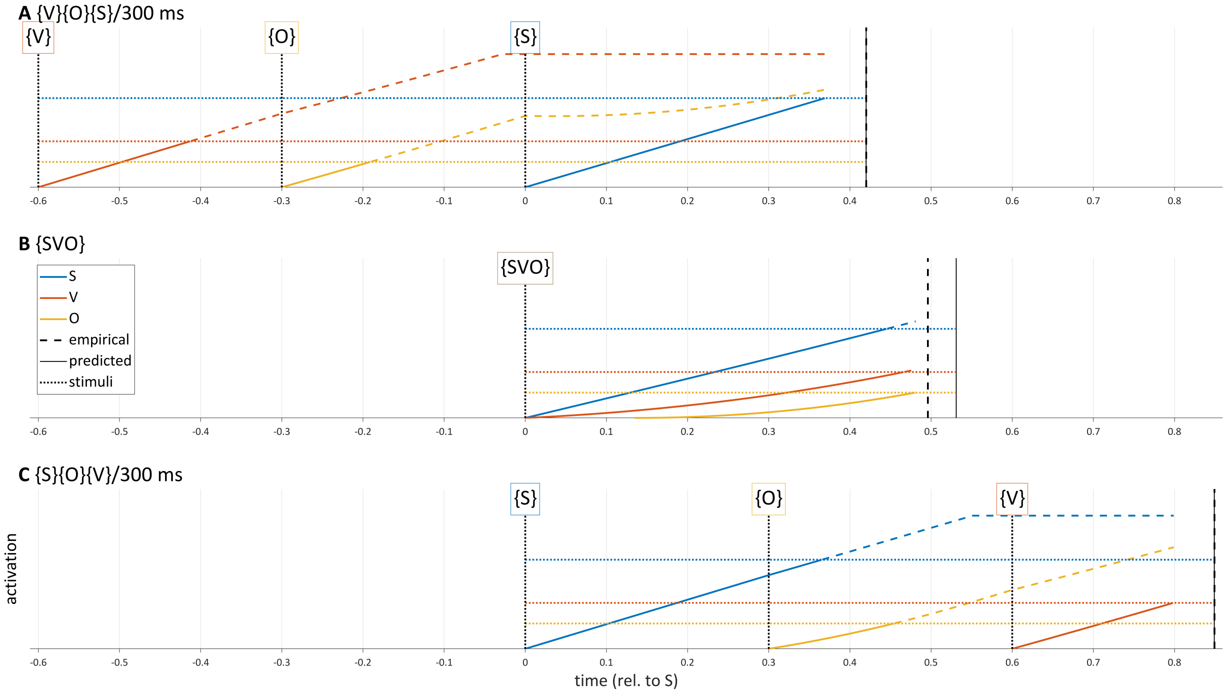


Fig. 23. Model simulations for {SVO} and for the patterns with extremal mean RT_[S]: {V}{O}{S}/300 and {S}{V}{O}/300.

Fig. 24 shows simulations for timing patterns where ΔOS = 0 and ΔVS varies. Note that panels are sorted by ΔVS from left to right and top to bottom. The fastest model-generated RT_[S] occurs when V precedes {SO} by 300 ms (A). Because |V| reaches threshold before |S| or |O| becomes active, it does not experience any interference; |O| does experience some interference from |S|, but the effects of this on RT are minor because τ_O is low. Note that (A-C) show how the model generates early V facilitation. In contrast, the late V delay pattern is generated in (E-G). Late V delay arises mostly because the V occurs late and therefore |V| reaches threshold later: |V| is the last system to reach threshold in all of the patterns where ΔVS>0.

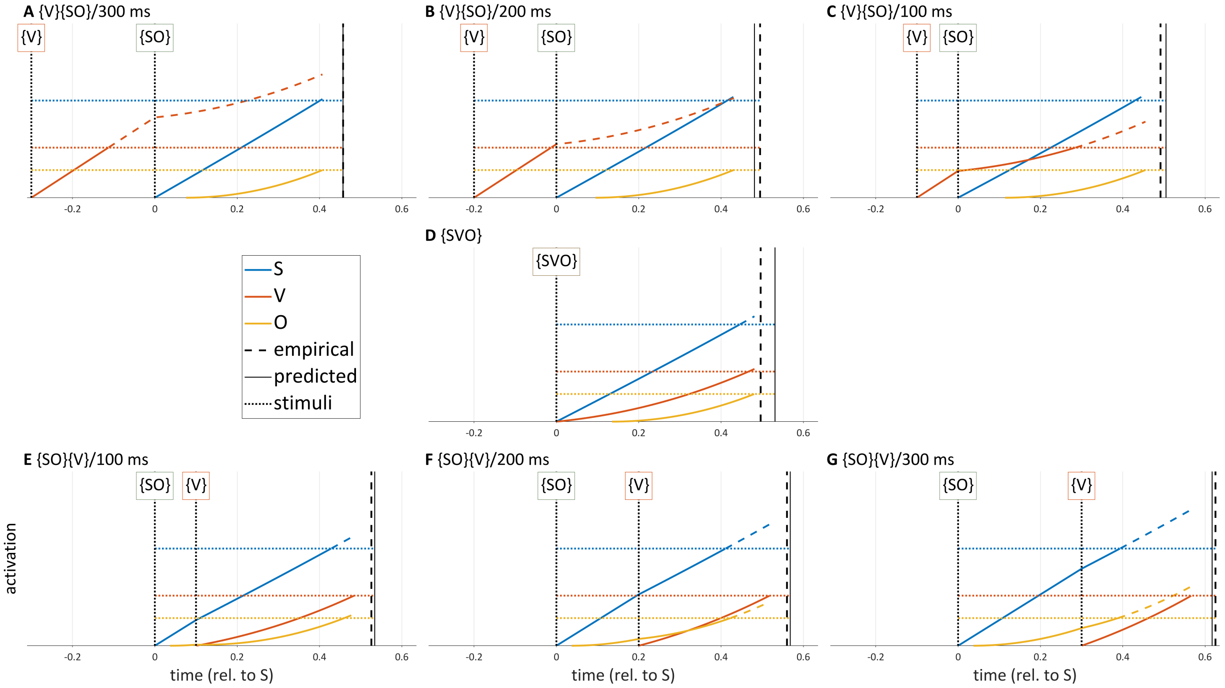


Fig. 24. Model simulations for timing patterns where ΔOS=0.

The optimized model thus generates early V/O facilitation and late V delay (see Fig. 2). Late V delay is generated straightforwardly from delays the timing of V, which delays the time of threshold achievement. The late V delay effect is larger than the late O effect because $\tau_V > \tau_O$. Early V/O facilitation is also generated by the model, but the way it accomplishes this is not necessarily what one would expect: rather than arising from a facilitatory mechanism, the early RTs arise from the absence of an inhibitory effect: when V or O stimuli occur early enough, |V| and |O| can reach threshold before |S| can interfere strongly with them. The effect of the absence of interference on |V| is more consequential than the effect of the absence of interference on |O|, because $\tau_V > \tau_O$ and because the interference exerted by |S| on |V| is greater than the interference exerted by |S| on |O|.

Experiments 2 and 3 Reaction times

Predictions

Experiments 2 and 3 were designed to obtain more accurate estimates of the effects of stimulus timing on response initiation time for three specific orderings: {SV}{O}, {SO}{V}, and {S}{VO}. These are the lines in the shaded regions of the Exp. 1 design in Fig. 25. By testing a finer grid of inter-stimulus intervals (ISIs), Exp. 2 allows for more a precise characterization the shape of the functions that relate RT to ΔVS and ΔOS. Note that {SVO} order is a limiting case of the non-synchronous orderings as ISI → 0.

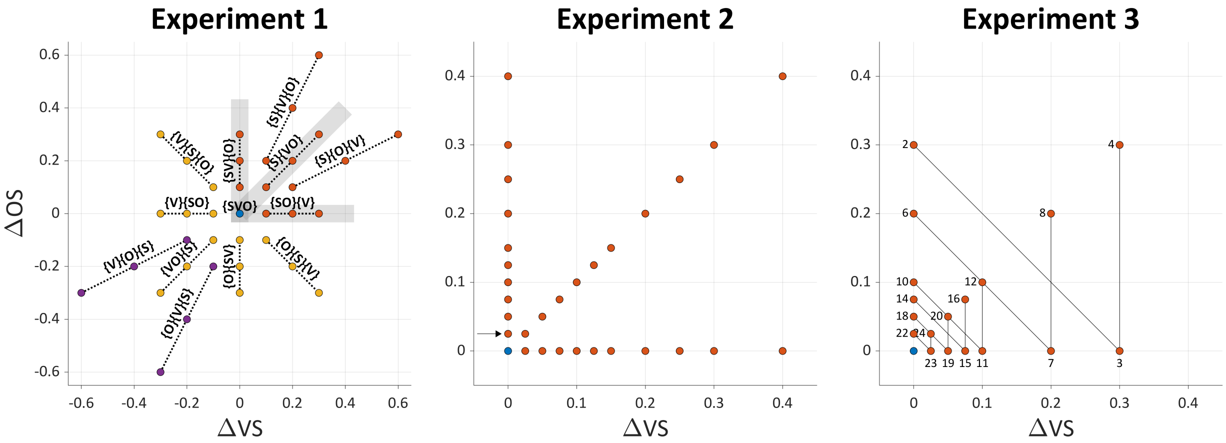


Fig. 25. Comparison of Δ-space sampling of Experiments 1-3

Consider the {SV}{O}/25 timing pattern (indicated by an arrow in the Exp. 2 panel of Fig. 25). In this condition the O stimulus appears approximately 25 ms after the {SV} stimulus set. Due to lack of precise control over the timing of graphics objects updates relative to screen refreshes, the actual ISIs can deviate up to ±8.3 ms $\left(= \frac{1}{2} \times \frac{1}{60Hz}\right)$ from the target ISI (see *Methods* for derivation of this). Psychometric studies have shown that the minimal ISI for correct detection of asynchronous order of visual stimuli is 20-40 ms (Hirsh & Sherrick Jr, 1961; Pöppel, 1997). Furthermore, it has been argued that sensory stimuli that co-occur in a temporal window of approximately 30 ms are integrated differently from stimuli with a larger ISI (Pöppel, 1997). This suggests that many of the asynchronous stimuli sets in the 25 ms ISI condition might not be perceived as asynchronous. Indeed, the psychometric threshold for visual asynchrony detection is typically calculated with relatively simple, non-linguistic stimuli, and since our stimuli are more complex, we might allow for a somewhat larger window of integration; on the other hand, the set of possible stimuli is small and this might reduce the asynchrony detection window. In any case, if visual asynchrony is a relevant factor, we would predict that for ISIs under a small threshold (such as 25 ms), there should be no effect of ISI on RT.

The dynamical model obtained from optimization of Exp. 1 data predicts a relatively constant relation between ISI and RT for a range of short, positive ISIs (predicted RT functions are illustrated later in Fig. 29). The source of this predicted effect is a trade-off between the inhibitory effects of interference between systems and the facilitatory effects of earlier stimuli. Specifically, when V or O occur simultaneously with S, a processing delay arises because of interference that |V| and |O| experience from |S|; however, this processing delay is counteracted by a facilitatory effect of allowing |V| and |O| activations to begin grow earlier in time. The predicted consequence of this is that there is a range of short ISIs in which RT_[S] is relatively insensitive to variation in ISI. We refer to this hypothesis as the *interference-parallelization trade-off*.

For subsequent comparison, the relevant RT patterns from Exp. 1 are shown in Fig. 26. The figure shows spline fits of RT_[S] from Exp. 1, after participant-specific intercepts have been removed. Note that the shortest ISI investigated in Exp. 1 was 100 ms, so the visual asynchrony effect cannot be assessed in this experiment. However, it is noteworthy that there are nonlinear relations between RT and ISI which have different forms for different orderings. For {SV}{O}, where ISI = ΔOS (orange line), there is a concave nonlinearity; for {SO}{V}, where ISI = ΔVS (blue line), there is convex relation. In other words, the RT effect of O delay is less than linear, while the RT effect of V delay is greater than linear. For {S}{VO} (yellow line), where increasing ISI entails both V and O delay, the ISI effects are less than additive.

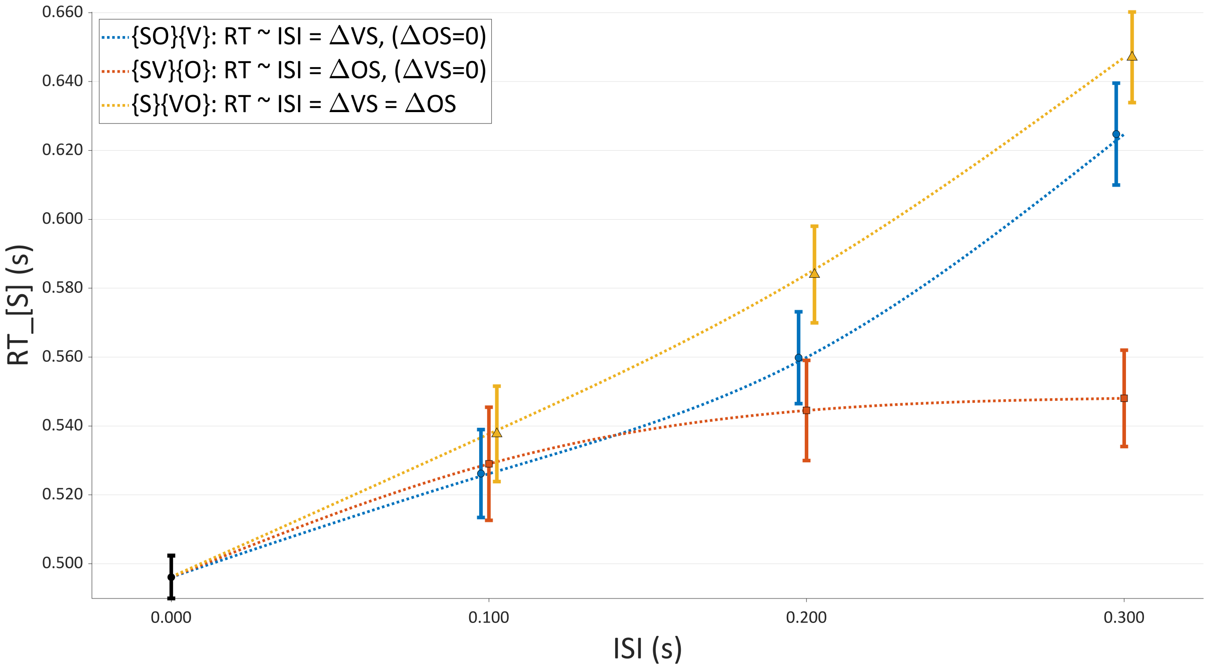


Fig. 26. Experiment 1 ISI effects on RT_[S] for {SO}{V}, {SV}{O}, and {S}{VO} orderings.

Another difference between experiments involves uncertainty in the ordering and timing of stimuli. We hypothesize that less uncertainty in when and where stimuli occur facilitates visual processing and possibly also response organization processes. Hence Both Exps. 2 and 3 are predicted to exhibit faster RTs than Exp. 1, because there is less uncertainty regarding the ordering of stimuli in Exps. 2/3, due to the smaller set of possible orderings.

Furthermore, this hypothesis predicts that RTs in Exp. 3 should be faster than in Exp. 2 because there is less uncertainty in Exp. 3 regarding the timing and ordering of stimuli. Exp. 3 differed from Exp. 2 in that the timing patterns were separated by blocks. The blocking of timing patterns was designed to optimize participant performance for the shortest ISIs. The blocks were ordered such that the four orderings—{SVO}, {SV}{O}, {SO}{V}, and {S}{VO}—occurred in that order, and ISI decreased over the experiment (excepting in the recurring synchronous {SVO} condition). The block numbers are labeled for each ISI in Fig. 25, except for the recurring {SVO} blocks, which were block numbers 1, 5, 9, …, 25. Hence the first block was {SVO}, blocks two through four corresponded to each of the three orderings with a 300 ms ISI, the fifth block was {SVO}, blocks six through eight were the three orderings with a 200 ms ISI, the ninth block was {SVO}, etc. The blocking design removes uncertainty about the timing and ordering of stimuli. The purpose of decreasing ISI in non-synchronous blocks over the session was to allow the participant to gradually adapt to faster ISIs over the course of the session. (Note that there was likely some uncertainty in the initial several trials of each block in Exp. 3, since participants were not explicitly informed of the timing pattern changes from block to block). Note that in all experiments there is the same amount of uncertainty regarding the lexical instantiation of the syntactic categories—the same sets of lexical items in the same proportions were used.

To summarize, the following hypotheses are tested in Exps. 2 and 3:

(i) *Visual asynchrony threshold hypothesis*: there is a minimum absolute value of ISI below which effects of stimulus asynchrony are not observed. This range is expected to be ±25 ms, on the basis of psychometric studies.

(ii) *Interference-parallelization tradeoff hypothesis*: there is a period of short ISIs in which the adverse effects of interference are counteracted by the facilitatory effects of organizing systems in parallel, resulting in a relative insensitivity of RT to ISI. This hypothesis predicts that RT will be relatively constant over a range of short ISIs. Based on the Exp. 1 optimized model, this range should be larger than the visual asynchrony threshold range. There should be a range of ISIs over which stimulus timing has no effect or minimal effect on RT.

(iii) *Uncertainty hypothesis*: uncertainty in the ordering and timing of stimuli slows response preparation. This hypothesis predicts that less uncertainty (or, greater predictability) in stimulus timing patterns will be associated with faster RTs. Both Exp. 2 and Exp. 3 have less uncertainty than Exp. 1, and thus speakers in Exp. 2 and 3 will on average exhibit faster RTs than in Exp. 1. Furthermore, Exp. 3 has less uncertainty than Exp. 2, and thus RTs will be faster on average in Exp. 3 than Exp. 2.

Results

RTs for Exps. 2 and 3 provide support for all three of the above hypotheses. First we examine Exp. 2 RTs, which are shown in Fig. 27. The upper left panel of the figure shows smoothing spline fits of the mean RT_[S] and 95% confidence intervals for each ordering in Exp. 2. The other panels show the same spline fits separately for each ordering, along with the means and 95% confidence intervals of RT_[S] for each ISI from Exps. 1 and 2.

The visual asynchrony threshold hypothesis was supported by Exp. 2: the mean RT at 25 ms ISI for all three asynchronous orderings did not differ significantly from the mean RT for {SVO} ordering.

The interference-parallelization tradeoff hypothesis was also supported: all three orderings have a range of ISIs in which RT is relatively insensitive to change in ISI. These insensitivity ranges are indicated by horizontal braces above which the mean RT over the range is shown. For {SV}{O} and {S}{VO}, the insensitivity range is 50-150 ms; for {SO}{V} the range is 50-125 ms. The mean values in these ranges are approximately the same for the asynchronous-SV and asynchronous-SO conditions (about 510 ms), and are about 20 ms higher than the RT for {SVO} order. In contrast, the value when both V and O are asynchronous (i.e. {S}{VO}) is about 35 ms higher than the synchronous order. These patterns suggest that the cost of O and V delays in the insensitivity range are about the same, and are close to additive.

The uncertainty hypothesis was supported by Exp. 2 as well: RTs were faster in Exp. 2 than in Exp. 1 RTs. Note that these differences between the experiments increase with ISI. Mean RTs are nearly the same for {SVO}, but differ by about 50 ms for {SO}{V}/300 ms.

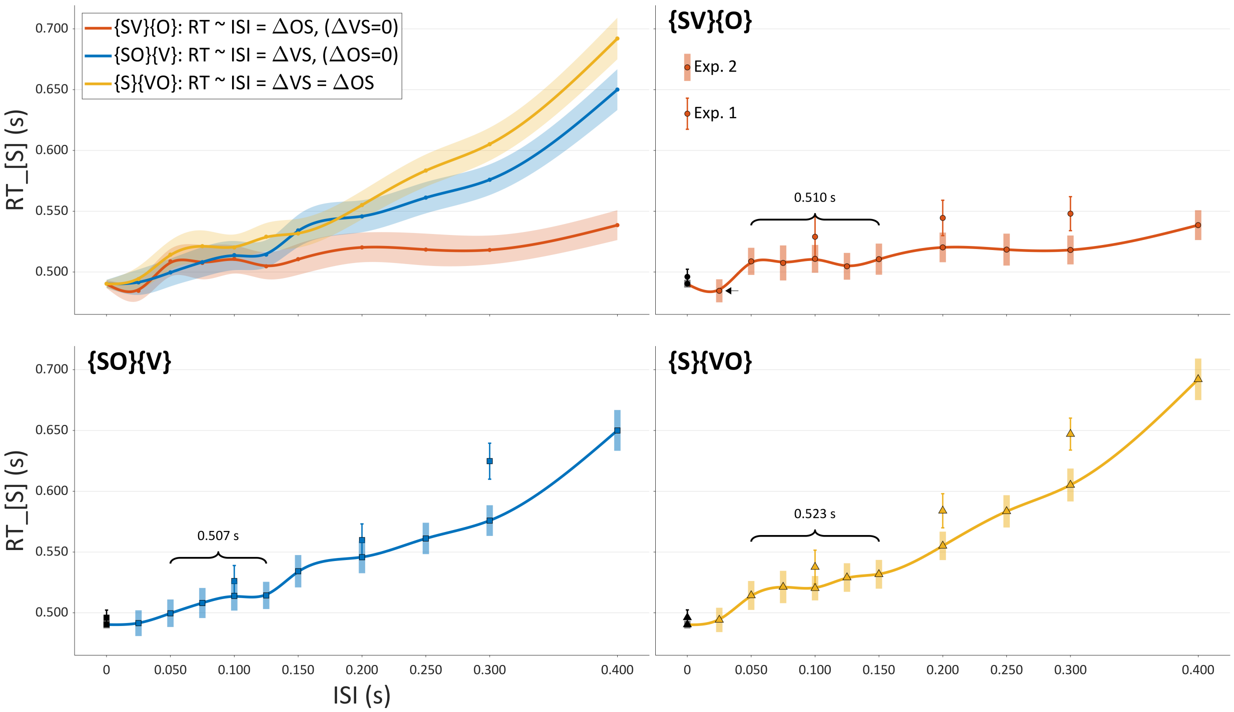


Fig. 27. Experiment 2 ISI effects on RT_[S]. Top left panel: smoothing spline fits for all three orderings with 95% confidence intervals. Other panels: smoothing spline fits for each ordering, with Exp. 1 means shown for comparison.

Exp. 2 replicates the Exp. 1 finding that response preparation depends more strongly on V timing than on O timing. The functions relating ISI and RT have similar shapes in Exps. 1 and 2; specifically, for {SV}{O} ordering the function is relatively flat, while for the {SO}{V} and {S}{VO} orders the functions are convex, i.e. exhibit a greater-than-linear increase of RT with ISI. The observation that ISI has a smaller effect for O delay than for V delay supports a model in which response initiation depends more strongly |V| organization than on |O| organization.

A potentially interesting trend shown in Fig. 27 is that RT for {SV}{O}/25 is faster than for {SVO}. The value at 25 ms is indicated by an arrow in the figure. This decrease is somewhat unexpected, since this ISI is within the hypothesized visual integration window. However, a post-hoc t-test of the difference did not find it significant (p=0.25, (t=1.15,df=218.7), diff=0.006, 95% c.i.=[-0.004, 0.016]).

The uncertainty hypothesis was further supported by RTs from Exp. 3, which are shown in Fig. 28, along with patterns from Exps. 1 and 2. RT_[S] was much lower in Exp. 3 than in Exp. 1 or 2. This effect appears to diminish for large ISIs (≥ 300 ms), but recall that in Exp. 3 block order is a confounding factor: RTs are relatively high for large ISIs because those blocks were performed earlier in the experiment, when participants were still learning to achieve optimal performance in the task. Note also that mean RT_[S] for the {SVO} ordering was calculated only from the last two {SVO} blocks of the session, in order to get an estimate of RT when the participant is most adept at the task.

In comparing Exp. 2 and 3 RT_[S] for ISIs below 300 ms, there is a fairly constant difference of approximately 50 ms. This indicates that there is a quite large effect of uncertainty about when the V and O stimuli will appear relative to S: when the timing of V and S stimuli is predictable, participants can initiate the response about 50 ms faster than when they are uncertain about stimulus timing. An interpretation of this in model terms is that participants may adjust their category-specific thresholds to minimize the time required for response preparation: for example, if it is known that the V will occur a specific period of time subsequent to the S, the participant can adopt a lower τ_S and initiate the response earlier, without risking a loss of fluency.

Evidence in support of the interference-parallelization tradeoff hypothesis was also observed in Exp. 3: ranges of insensitivity in RT to variation in ISI were again observed. However, the exact ranges and offsets differed from those of Exp. 2. Specifically, in the {S}{VO} condition, the insensitivity range is lower and smaller, from 25 to 75 ms, and the RTs do not differ substantially from {SVO}. The same is mostly the case for the other two orderings.

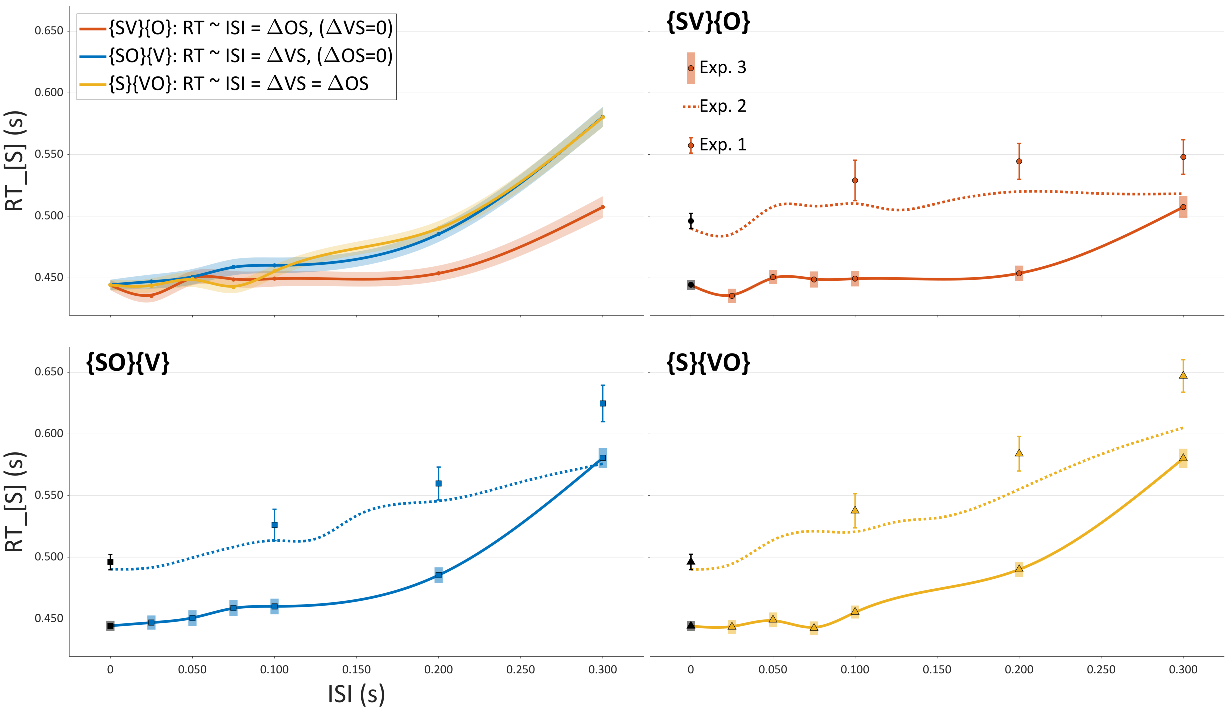


Fig. 28. Experiment 3 ISI effects on RT_[S]. Top left panel: smoothing spline fits for all three orderings with 95% confidence intervals. Other panels: smoothing spline fits for each ordering, with Exp. 1 means and Exp. 2 spline fits shown for comparison.

The visual asynchrony threshold hypothesis was partly supported by Exp. 3: the mean RT at 25 ms ISI for two of the three orderings ({SO}{V} and {S}{VO}) did not differ significantly from the mean RT for {SVO}. However, as in Exp. 2, Exp. 3 shows an unexpected response facilitation at 25 ms in the {SV}{O} condition. A post-hoc t-test showed that mean RT_[S] for {SV}{O}/25 and {SVO} are significantly different ($p<0.02$, ($t=2.49$, $df=734.8$), diff=0.009, 95% c.i.=[0.002, 0.016]. The effect is fairly small (9 ms), but the fact that it was observed in both Exps. 2 and 3, suggests that it may be attributable to mechanisms of response preparation, rather than stochastic variation. Moreover, the effect is clearer in Exp. 3, where stimulus timing is less uncertain.

Response initiation model

Here we examine how models which are optimized to fit data for Exp. 1 and Exp. 2 differ. The models have the same parameters structure, i.e. specific thresholds and interaction parameters, but their parameter values are different, because they were optimized to generate mean response initiation times for Exp. 1 and Exp. 2, respectively. Fig. 29 compares the model-generated RT patterns for the three orderings of Exp. 2. One clear difference in the model-generated RT functions is that for {SV}{O} and {SO}{V} orderings, the

Exp. 2 model (solid lines) is relatively insensitive to ISI variation below some particular ISI value. This value is much higher for {SV}{O} ordering than for {SO}{V}. In contrast, the Exp. 1 model exhibits a complex form of ISI dependence for all orderings, which results in a nonmonotonic relation between ISI and RT. Note that the Exp. 2 model does not generate early V and O facilitation effects. The models have the same relative values of threshold parameters, and the two strongest interference interactions (|S|→|O|,|V|) are the same. The main differences in the Exp. 2 model parameters relative to the Exp. 1 model parameters are: (i) slightly higher growth rate and (ii) much higher τ_S. In addition, |S|→|O| interference is about 20% weaker, |O|→|V| interference is twice as strong, and |O| nor |V| interference with |S| is negligible.

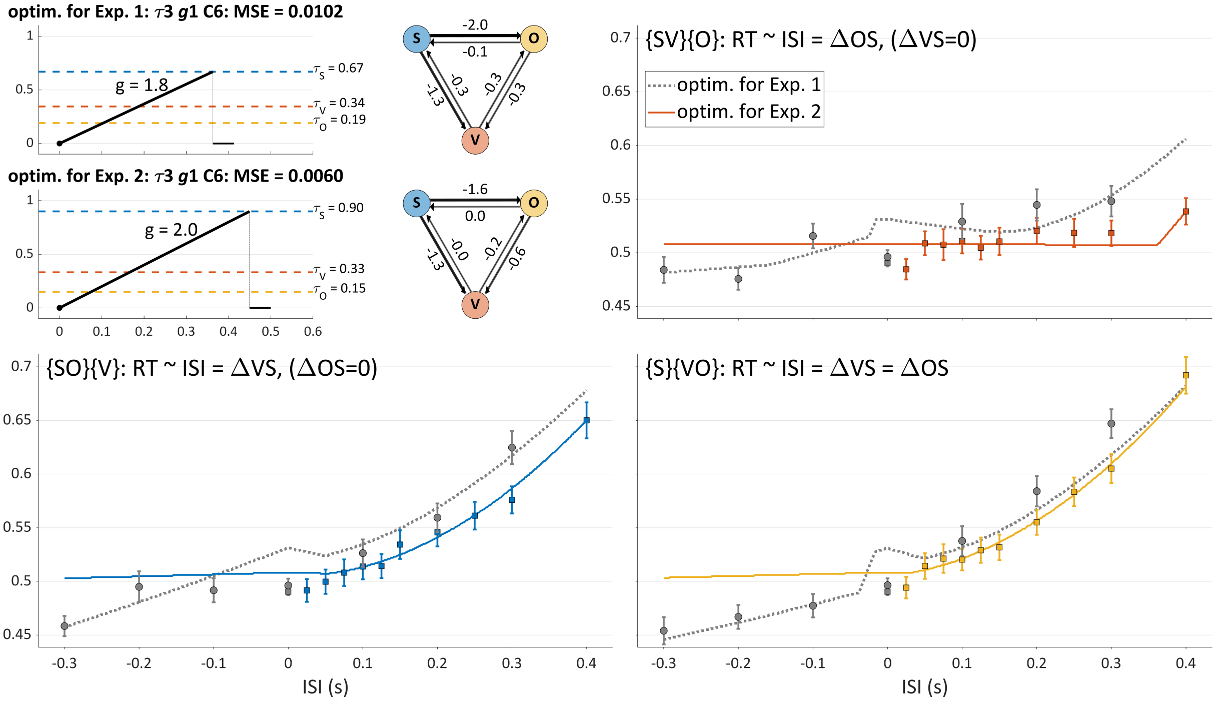


Fig. 29. Comparison of Exp.1 and Exp. 2 models. Top left: model parameters. Other panels show predicted RT_[S] for each ordering, for models optimized on data from Exp. 1 (gray dotted lines) and Exp. 2 (solid lines); also shown are means and 95% confidence intervals of RT_[S] from Exp. 1 (gray circles) and Exp. 2 (squares).

On a qualitative assessment, both models are somewhat unsatisfactory, but for different reasons. Observe that the Exp. 1 model generates an RT bump around ISI = 0 ms. Because of this, its predictions for {SVO} ordering and for short ISIs are too large. This bump is a consequence of relatively strong |S|→|O| interference and a relatively low threshold for |S|. In contrast, the Exp. 2 model is unsatisfactory because it does not generate early V/O facilitation. This is not surprising, given that the Exp. 2 model is not optimized on data where any stimuli precede S.

Response analyses

Here we analyze some additional aspects of the response, specifically word duration and response errors. Note that square brackets are used to refer to the production of a word associated with a given syntactic category, i.e. [S] refers to productions of the word form associated with the S stimulus.

Word duration

The only substantial effect of stimulus timing on word duration is that [S] duration increases with ΔVS and ΔOS. Word durations from each experiment are shown in Fig. 30. The scales in all panels of the figure have the same range (about 65 ms), but differ in their offsets. Thus the relative differences between colors in the figure have the same meaning across panels, despite the fact that [S], [V], and [O] have different average durations. Note that Exp. 3 has durational effects that are most likely attributable to block order, rather than stimulus timing.

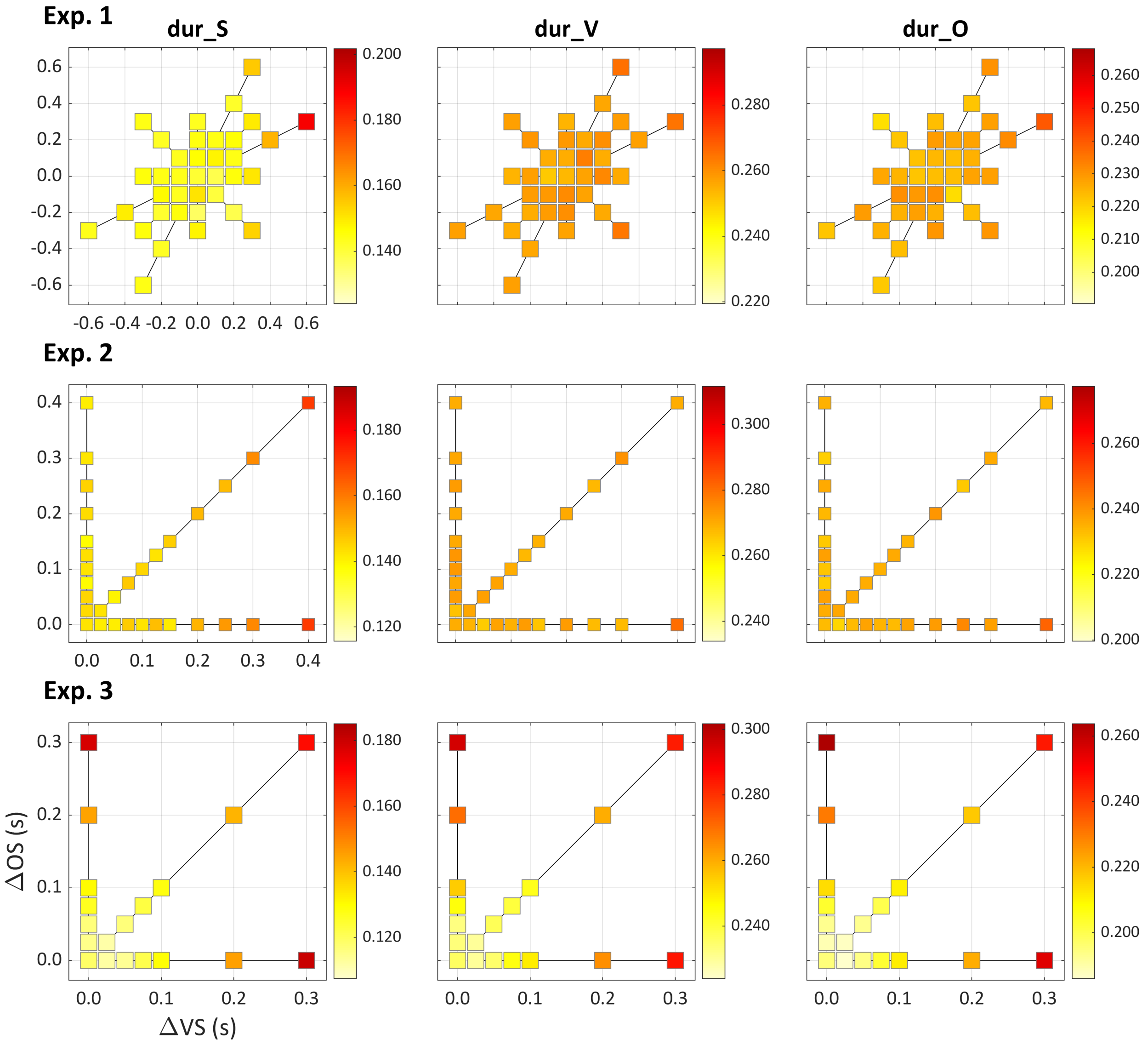


Fig. 30. Word durations in Exps. 1-3. The scales in all panels of the figure have the same range (about 65 ms), but differ in their offsets. Note that the experiments probe different values in Δ-space and Exp. 3 has effects that are attributable to block order.

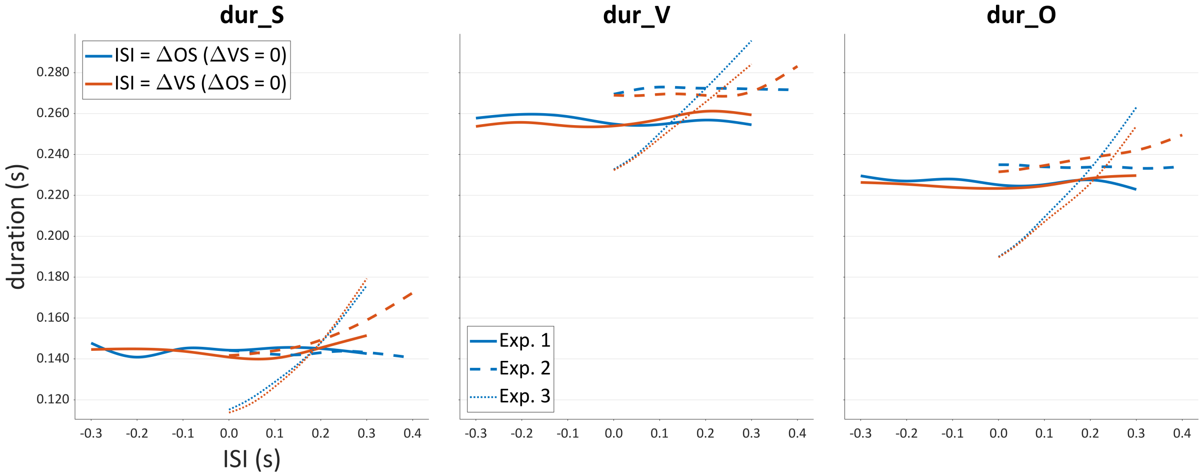


Fig. 31. Spline fits of ISI effects on word durations of each syntactic category, for all three experiments. Blue lines: ΔOS, orange lines: ΔVS.

When we compare ISI effects across experiments (Fig. 31), one notable difference is that [V] durations and to a lesser extent [O] durations are longer in Exp. 2 than Exp. 1. Specifically, [V] durations are about 15 ms longer in Exp. 2 than in Exp. 1. One possible explanation for this is that by eliminating uncertainty in the timing of S stimuli, Exp. 2 drew more attention to V and O stimuli, and this had the effect of increasing the corresponding word durations. This explanation requires a mechanism whereby attention to stimuli increases the corresponding production duration. An alternative possibility is that Exp. 2 participants happened to speak more slowly than Exp. 1 participants, but this is not consistent with the similarity in durations of [S].

For Exp. 3, word durations for small ISIs are shorter for all three syntactic categories than in Exps. 1 or 2. This is consistent with the explanation proposed above: with no uncertainty about stimulus timing, word durations are shorter. However, Exp. 3 conflated block order with ISI, and this could account for why Exp. 3 durations increase with ISI.

Another interesting pattern is a subtle difference between effects of Δ measures on [S] duration in Exps. 1 and 2. Observe that [S] duration increases for large ΔVS (orange lines) but not for large ΔOS (blue lines); this difference may arise because speakers slow down [S] production to buy time for the processing of the V stimulus or the motor preparation of [V]. This is an interesting effect because it suggests that the timecourse of [S] production can be modulated by the state of the |V|.

Word durations which are abnormally long can be interpreted as a form of hesitation, and it makes sense that the most substantial effects on duration are observed in [S] for large ΔVS: when V is delayed the speaker may have initiated [S] production without being prepared to produce [V]. The occurrence of silent pauses—another form of hesitation—might also reflect this scenario. However, silent pauses were quite infrequent in the absence of a response error. The percentages and counts on non-error trials of silent pauses between S and V and between V and O are shown in Table 13.

Table 13. Silent pause rates

	S-V		V-O	
Exp. 1	0.17%	12/7199	0.21%	15/7199
Exp. 2	0.14%	10/7378	0.03%	2/7378
Exp. 3	0.04%	4/9358	0.09%	8/9358

Response errors

Response errors in the experiments are important because they show how the simplified dynamical model is insufficient, and therefore provide guidance on how the model should be elaborated. The analysis here is focused on the lexical and syntactic aspects of errors, rather than the detection or repair of errors. To that end, we adopt the following classification and terminology:

Lexical substitution errors: these are errors in which the wrong lexical item is produced, i.e. a source item substitutes for the target item. 268 lexical substitution errors were identified across experiments. These can also be described as *overt* errors because the identity of the source is evident in the utterance. Lexical substitution errors are subclassified as internal or external depending on whether the source is or is not part of the target sentence. For example:

Internal lexical substitution error: “Mo-(Moe) Lee saw Moe” TARGET: *Lee saw Moe*

> Note that “Mo-(Moe)” indicates that this pronunciation of Moe was cut off before completion; the parenthetical specifies the inferred word form. Cutoff disfluencies are very common when lexical errors are repaired, and in general, speakers usually detect and repair such errors. Nonetheless, there are some cases in which lexical errors are not repaired. Internal substitutions can be classified according to the syntactic categories of the source and target items. Here the source is O and the target S. To represent this we use the abbreviation: S←O.

External lexical substitution error: “Ray saw Moe” TARGET: *Lee saw Moe*

> In this example, the source item (Ray) is not part of (i.e. is external to) the target sentence. This variety of lexical substitution is less frequent than internal substitutions for noun targets. Note that all lexical substitutions of V are necessarily external, because there is only one V target in each sentence.

Lexical substitutions are the most frequent types of errors, and the analyses below focus primarily on these errors. Notably, it does not seem to be the case that an S or O was substituted for a V, or vice-versa. In other words, substitutions never involved words of different lexical categories (i.e. NOUN and VERB). There was one instance of an error in which the first word produced was the V: “hear-(heard) SIL Ray heard Moe” for the target utterance *Ray heard Moe*. This could be analyzed as lexical substitution of V for S, but alternatively it can be analyzed as a failure to select S, which was subsequently repaired.

Covert errors: these occur when a participant produces a correct form/forms, and then hesitates and/or repeats that form/forms. Often there is a silent pause between the initial productions and repetitions, and sometimes the initial production may be cutoff. Such errors may arise from difficulty in preparing an upcoming form, or could reflect uncertainty regarding the stimulus, or could arise from an errorful plan detected by an internal monitor. There were 24 cases of covert errors in which a form/forms were repeated; disfluent responses in which there were filled pauses or hesitations without repetition may also be covert errors.

Blends: these are errors in which a form that is produced combines articulatory gestures associated with two different lexical items. For example, in the sentence *Lee saw Moe*, a speaker might produce the target subject *Lee* as [mi:], combining the onset of the target object, [m], and the vowel of the target subject, [i:]. An error was only labeled as a blend if the form was clearly identifiable as the combination of two forms. 25 blends were identified across the experiments, although some pronunciation errors may also be blends. All but two blends involved targets of the same syntactic category. Blends could be interpreted as mechanistically related to internal lexical errors, but it is ambiguous whether they arise from errorful selection of lexical items or errorful selection of articulatory gestures.

Other errors: various other types of errors occurred but are not analyzed here because they are too infrequent or not suitable for analysis. These include trials where the participant failed to respond, word mispronunciations (which were not interpretable as blends), and the presence of filled pauses or non-speech vocalizations (i.e. coughs, yawns).

The lexical substitution error percentages and counts by experiment are shown in Table 14. The rates were around 1% in all experiments and did not differ significantly between experiments (p = 0.102, $\chi^2(2)$ = 4.6). The lexical error counts by class are shown in Table 15, along with percentages. V substitutions were the most common lexical substitution, accounting for about half of all substitution errors. All lexical V errors are external by definition, because there is only one V in each sentence. The next most common class of lexical substitutions were internal S←O errors, where the object was produced in place of the subject. External substitutions for S occurred somewhat less frequently. In contrast, for substitutions of target O, there were more external than internal sources (34 vs. 11). Of the 11 internal substitutions for O, 10 of these were “exchange” or “transposition” errors (S←→O) where the target O also served as an internal source of substitution for the target S. There was only one case where a S that was correctly produced was repeated in place of the target O.

The asymmetry in the S and O substitution error subclasses can be reasonably interpreted as a predominance of anticipatory errors over perseveratory errors. More usefully, this asymmetry can be understood to follow from a tendency for participants not to substitute a previously produced form for a target form, in combination with a propensity to substitute one argument in the target sentence for another (i.e. internal substitution).

Table 14. Lexical substitution error rates by experiment

Exp. 1	0.8%	60/7293
Exp. 2	1.1%	82/7497
Exp. 3	1.1%	101/8802

Table 15. Substitution error subclass counts and percentages

external V	113	51.6%
internal S<--O	39	17.8%
external O	34	15.5%
external S	22	10.0%
internal S<-->O	10	4.6%
internal S-->O	1	0.5%

Table 16. Error subclass proportions by experiment

	Exp. 1	Exp. 2	Exp. 3
external O	0.13	0.17	0.16
external S	0.08	0.10	0.11
external V	0.56	0.60	0.43
internal S<-->O	0.02	0.04	0.07
internal S<--O	0.21	0.09	0.24

Table 17. Lexical substitutions by lexical source and target

		SOURCE				
		Lee	Moe	Ray	saw	heard
TARGET	Lee		17	23		
	Moe	7		17		
	Ray	29	9			
	saw					22
	heard				95	

Examination of the distribution of error classes by experiment (Table 16) did not reveal strong asymmetries. A contingency analysis conducted after excluding the sole S→O error did not find a significant interaction between experiment and error type ($\chi^2(8) = 10.2$, $p = 0.25$).

There were asymmetries in the lexical identities of the sources and targets of substitutions. These are shown in Table 17. The target *Ray* was more frequently substituted with *Lee* than with *Moe*, and to a lesser extent *Lee* was more frequently substituted with *Ray* than with *Moe*. This could be due to the greater degree of similarity between the initial consonants /l/ and /r/ than between /l/ or /r/ and /m/. Another somewhat more puzzling asymmetry was that target *heard* was substituted with *saw* much more frequently than the reverse.

To assess whether substitution error likelihood was influenced by Δ-values, logistic regressions of error likelihood as a function of ΔVS and ΔOS were conducted for each of three most frequent substitution error subclasses, for each experiment separately. In all three experiments, the only error subclass for which Δ terms were significant predictors were external V errors (Exp. 1: $\chi^2(2) = 13.06$, $p < 0.001$; Exp. 2: $\chi^2(2) = 91.33$, $p < 0.001$; Exp. 3: $\chi^2(2) = 16.49$, $p < 0.001$). In all of these cases, increases in ΔVS increased the likelihood of a substitution for V.

The occurrence of lexical substitution errors, and in particular the syntactic category asymmetries between internal and external sources of these errors, are important phenomena because they place strong requirements on an adequate model of response initiation and production. The simplified dynamical models considered above do not generate lexical substitution errors. Hence error phenomena indicate that major extensions to the model are necessary. Specifically, the simplified models do not have an explicit conceptual-syntactic association mechanism, nor do they have an error detection mechanism.

Discussion and conclusion

Summary of main findings

The main findings of the experiments are listed below and summarized in Fig. 32.

(i) *Greater sensitivity to V delay than O delay*. The timing of V relative to S is generally more influential than the timing of O relative to S. This is particularly evident for ΔVS or ΔOS > 100 ms. The regression analyses conducted in preceding sections also support this conclusion: ΔVS coefficients had greater magnitudes than ΔOS coefficients in both linear and nonlinear regressions of Exp. 1. The simplified dynamical model is able to capture this difference via a higher initiation threshold for V than for O.

(ii) *Region of relative insensitivity*: response initiation is relatively insensitive to the timing of V and O stimuli relative to S for asynchronies in the range of -125 to 125 ms. This is illustrated in Fig. 32 as shallower slopes of the lines connecting mean RTs in Exps. 2 and 3. It was also manifested in the quadratic terms of the nonlinear regression of Exp. 1. The model is able to capture this as follows. For ΔVS<0 and ΔOS<0 in this range, the advantage of activating |V| and |O| systems earlier (i.e. in parallel with |S|) is counteracted by the interference effects they experience from |S|.

(iii) *Early V and O facilitation*. For large negative ΔOS and ΔVS in Exp. 1—and particularly with ΔVS = -300 ms, there is a facilitatory effect on response preparation, such that RT_[S] is nearly 40 ms faster in {V}{SO} than in {SVO} of Exp. 1. The model generates this effect via the reduction of interference effects of |S| on |V| and |O|.

(iv) *Stimulus timing uncertainty effect*. Uncertainty in stimulus timing leads to longer response initiation times. In Exp. 1, there were 13 unique orderings and 37 unique timing patterns which could occur on each trial. In Exp. 2, there were 4 unique orderings and 31 unique timing patterns. In Exp. 3, there was 1 unique timing pattern in a given block of trials. The relative entropies of ordering and timing pattern across experiments are correlated with the relative RTs of the experiments. The simplified model has no mechanism for generating these effects.

(v) *The 25 ms anomaly: object-delay facilitation.* When the O stimulus occurred approximately 25 ms after S, response initiation was facilitated. This effect was fairly small—on the order of 10 ms—and was only significant in a post-hoc test in Exp. 3. Our simplified dynamical models do not predict this anomaly.

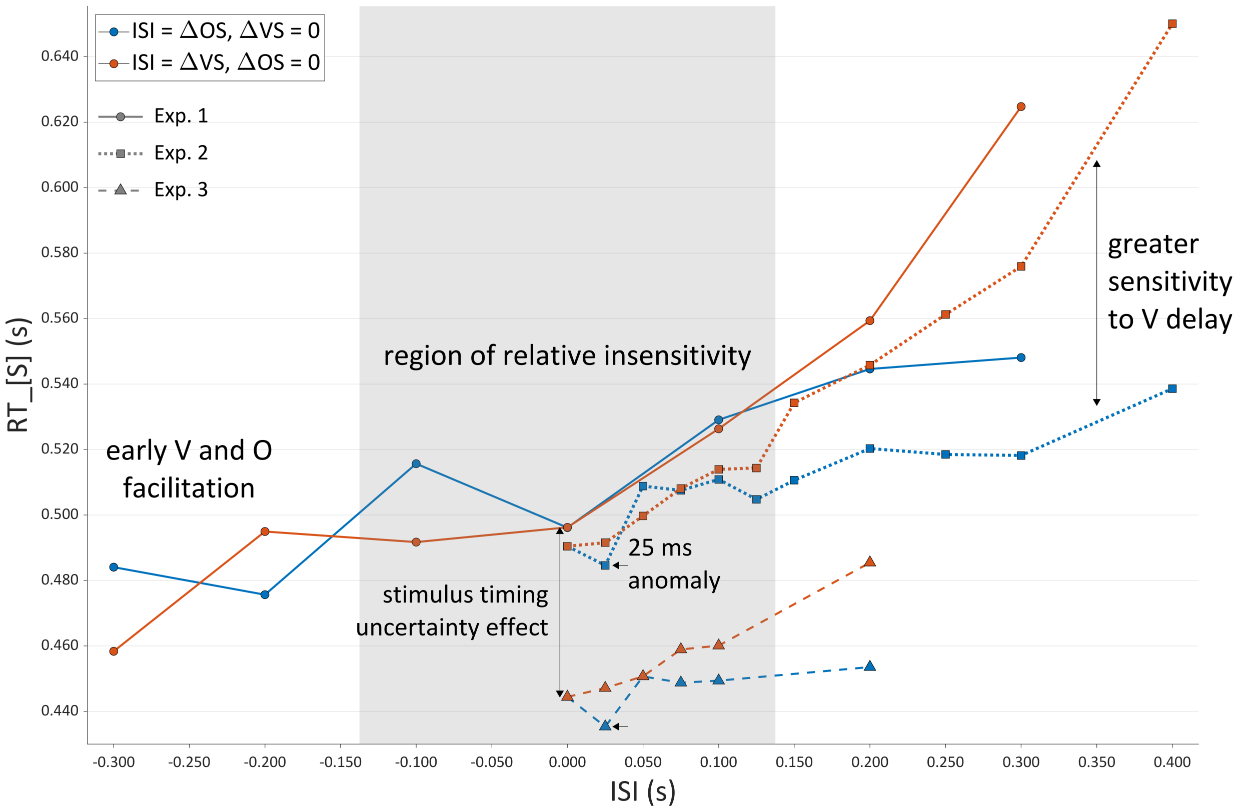


Fig. 32. Summary of main findings.

Evidence for syntactic categories?

Do the experimental findings support the idea that there are “syntactic” categories of S, V, and O, as opposed to generic ordering categories of first word, second word, and third word (i.e. W1, W2, and W3)? It is logically possible that the RT patterns are not specific to the categories S, V, and O, but rather, are simply attributable to the order in which the words are produced. For instance, consider that ΔVS was found to be more influential than ΔOS on RT_[S]. This could be alternatively interpreted to mean that the timing of W2 relative to W1 (Δ21) was more influential than the timing of W3 relative to W1 (Δ31). Indeed, it is not easy to identify any particular RT pattern that cannot be viewed in this way.

To address this question, a control experiment could be conducted in which a list of three words is produced which does not constitute a sentence and does not contain verbs. Replacing the verbs with two names (e.g. *Will*, *Ned*) would be a sensible approach, as this would maintain the same informational content of the stimuli. These alternative names would be exclusive with one another and would always appear second in the list.

One potential source of evidence in favor of the syntactic interpretation is an interaction between category and stimulus timing uncertainty. Specifically, consider the comparison of Δ effects between experiments in Fig. 33. The Δ-RT relations are plotted without intercepts on the right of the figure. The coefficients of nonlinear regressions of the form RT = 1 + Δ + Δ^2 are shown for each subset. Participant-specific ISI effects were residualized, and only ISIs less than or equal to 200 ms are included, to avoid the influence of the block order confound of Exp. 3. The cases in which the nonlinear terms are significant ($p<0.05$) are indicated with an asterisk. Notice that the coefficients of the nonlinear term are positive for regression by ΔVS and negative for regression by ΔOS. This means that delay of V is associated with a larger than linear increase in RT, while delay of O induces a smaller than linear increase in RT. This contrast does not seem to gel with the notion that it is word order rather than syntactic identity that drives RT patterns. If word order were the sole influence on RT we would have no reason to expect that the signs of the nonlinear coefficients would differ between W2 and W3. On the contrary, something else beyond word order seems necessary to account for the difference.

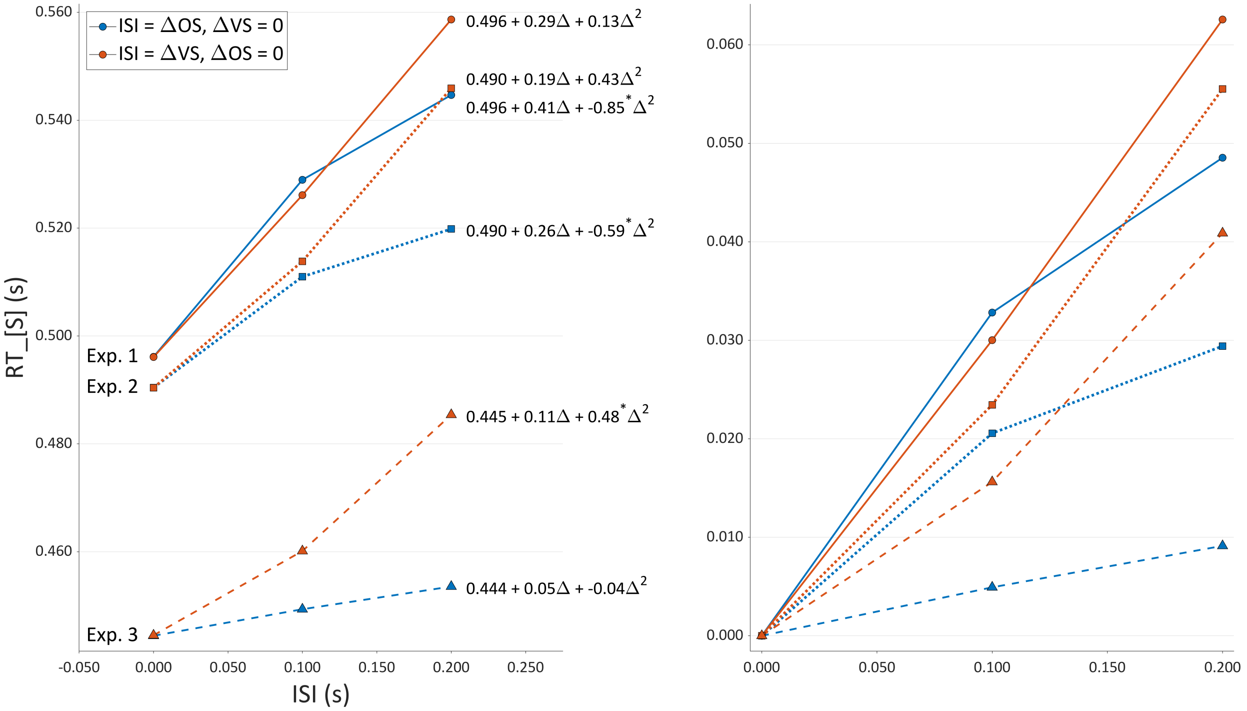


Fig. 33. Comparison of ISI effects by predictor (ΔOS or ΔVS), for all three experiments. Left: intercepts included. Right: intercepts removed.

Furthermore, it is evident that the both the relative influences of the linear and nonlinear terms differ across experiments and by Δ-measures: the nonlinearity of the ΔVS effect becomes more influential as uncertainty regarding stimulus timing is reduced, while the both the linear and nonlinear ΔOS effects become less influential as uncertainty is reduced. This difference is somewhat puzzling: why does uncertainty reduction affect V and O stimulus timing effects in qualitatively different ways? Because our dynamical model does not draw explicit connections between uncertainty and model parameters, it does not offer much help in reasoning about such patterns.

Information production/entropy analysis

To assess the role of uncertainty (i.e. entropy) and information production in RT patterns, we must consider the timecourse of information produced by the observation of the stimuli. Recall that we refer to a set of simultaneous stimuli as a *stimset*, and we will represent the first, second, and third stimsets as $\{\}^1$, $\{\}^2$, and $\{\}^3$, respectively. Here we distinguish between several different types of information: lexical information, spatial information, and temporal information:

Lexical information: information about the identities of the lexical items of the stimuli: {*Lee*, *Moe*, *Ray, saw, heard*}. There are $N \ = \ 3 \times 2 \times 2 = 12$ different combinations, which are equally probable ($p = \frac{1}{N}$). The lexical entropy in bits for equiprobable alternatives is defined simply as $-\sum_i p_i \log_2(p_i) = -\log_2\left(\frac{1}{N}\right)$. Recall that V ∈ {*saw*, *heard*}, and S,O = N ∈ {Lee, Moe, Ray}, with the constraint that S≠O. The total lexical entropy before the first stimset is thus 3.585 bits for all experiments.

As stimuli appear, the lexical entropy decreases. The rate of entropy reduction (i.e. information production) differs by timing pattern and differs on average across experiments. As shown in Table 18, more lexical information is produced on average by $\{\}^1$ in Exp. 2 than in Exp. 1. This is because the S always appears in $\{\}^1$ in Exp. 2. Accordingly, there is on average more lexical uncertainty remaining after $\{\}^1$ in Exp. 1 than in Exp. 2 (column H^1). Furthermore, slightly more information is produced on average by $\{\}^2$ in Exp. 2, and no uncertainty remains after $\{\}^2$ in Exp. 2 (column H^2). Note that the averages of information production are calculated only over those timing patterns in which a stimset occurs, and the averages of remaining entropy are calculated only over timing patterns in which there is any remaining lexical uncertainty—this is why the entropy remaining after a stimset does not equal the entropy before that stimset minus the information produced. For Exp. 3 the information production and remaining entropy depend on the stimulus ordering of a given block and hence are not statistically stationary on intermediate timescales.

Table 18. Average lexical entropy (H) and information production by stimset

Exp.	H^0	$\{\}^1$	H^1	$\{\}^2$	H^2	$\{\}^3$
1	3.58	1.83	1.90	1.40	1.00	1.00
2	3.58	2.58	1.33	1.33	0	n/a
3 {SVO}:	3.58	3.58				
3 {SO}{V},{SV}{O}:	3.58	2.58	1.00			
3 {S}:	3.58	1.58	2.00			

Spatial information: what will be the spatial arrangement of information in view of the participant after a stimset appears? For $\{\}^1$ in Exp. 1 there are 6 possibilities: {SVO}, {SV}, {SO}, {S}, {V}, {O}, and for Exp. 2 there are 4 possibilities: {SVO}, {SV}, {SO}, {S}. As shown in Table 19, the initial spatial entropy is greater for Exp. 1 than Exp. 2 (column H^0), but more information is produced on average by $\{\}^1$ in Exp. 2 than in Exp. 1. In fact, in Exp. 2, there are always 2 bits of spatial information produced by $\{\}^1$, and there is never any remaining uncertainty in the spatial arrangement of $\{\}^2$. There is no spatial uncertainty remaining here because the spatial arrangement of $\{\}^2$ is fully determined by $\{\}^1$ in Exp. 2. Note that {SVO} is not equally probable with asynchronous orderings in either experiment, with p({SVO}) = 0.077 and p({SVO}) = 0.25, in Exps. 1 and 2, respectively. For timing patterns with two or three stimsets in Exp. 1, $\{\}^2$ produces information. Exp. 3 never produces spatial information, because there is no uncertainty about the spatial arrangement of information due to the blocking of timing patterns.

Table 19. Average spatial entropy and information production by stimset

Exp.	H^0	$\{\}^1$	H^1	$\{\}^2$	H^2	$\{\}^3$
1	2.60	0.87	1.88	1.88	0	0
2	2.00	2.00	0	0	n/a	n/a
3	0	0	0	0		

Temporal information: what will be the interval of time between stimuli (ISI)? In Exp. 1 there are four possibilities: [0, 100, 200, 300 ms]; in Exp. 2 there are 10 possibilities: [0, 25, 50, 75, 100, 125, 150, 200, 300, 400 ms]; recall that in both cases these are not equiprobable. The first stimset always resolves some uncertainty because it informs the participant whether the ISI is zero or non-zero; the entropy reduction (information production) associated with $\{\}^1$ is greater in Exp. 2 than in Exp. 1, because the set of possibilities is larger (although the larger probability of {SVO} actually decreases the entropy reduction of $\{\}^1$ somewhat in Exp. 2). In Exp. 2, the average temporal entropy is larger after $\{\}^1$ (column H^1) than it is before $\{\}^1$ (column H^0), because the averages are calculated only over timing patterns where there exists a second stimset (this restriction makes sense if we interpret the guaranteed non-occurrence of a stimset as a non-event, i.e. one does not "observe" a non-event, especially given that there is no chance that some event will occur). In that case, in Exp. 2, consider that although after $\{\}^1$ the set of possible ISIs has decreased, the probabilities of the remaining ISIs are much more uniform (and in fact equiprobable)—this accounts for the slight increase in entropy (column $H^1 > H^0$). Once again, Exp. 3 produces no temporal information, because the blocking of timing patterns eliminates temporal uncertainty.

Table 20. Average temporal entropy and information production by stimset

Exp.	H^0	$\{\}^1$	H^1	$\{\}^2$	H^2	$\{\}^3$
1	1.85	0.39	1.58	1.58	0	0
2	3.30	0.81	3.32	3.32	n/a	n/a
3	0	0	0	0		

It is important to consider that the above analyses assume an observer with exact knowledge of event probabilities. Human participants, on the other hand, do not have this knowledge. Let us assume that participants develop estimates of probabilities for the lexical, spatial, and temporal properties of stimuli, based upon relative frequencies they experience over the course of the experiment. For Exps. 1 and 2, these estimates will necessarily be biased early on in a session, because participants have not been exposed to the full set of timing patterns and lexical items. They may also develop estimates that are biased toward properties of recent exemplars as opposed to estimates that reflect the full distribution of properties experienced up to a given point in an experiment.

Even if the estimate is an exact reflection of the current distribution, it is important to note that the relative frequencies of stimulus properties rarely exactly match their probabilities. For lexical information, the normalized frequency distribution approaches these probabilities more closely as the experiment progresses. For temporal and spatial information, the distribution exactly matches the probabilities at the end of each block (in Exps. 1 and 2), and within blocks, the deviations between the normalized frequency distribution and the probabilities becomes smaller over the course of the session.

In addition, for Exp. 3 the initial several trials of each block may in fact contain some temporal and spatial uncertainty, even though we have stated above that there is no such uncertainty. The reason for this is that participants may require a trial or two to recognize that there is a new timing pattern.

Despite the above issues, we will assume that, at least after the first block of trials (which are excluded in most of the preceding analyses), the expectations of human participants are close to those of an ideal observer with exact knowledge of probabilities; in that case, the information calculations above are still useful for reasoning about differences between experiments.

Now we examine the timecourse of lexical information production in more detail, in order to calculate a rate of information production. Fig. 34 shows time courses of the lexical entropy for all 37 unique stimulus timing patterns in Exp. 1 (many of the lines overlap). Information is produced whenever entropy (H) is reduced, so the information production events shown in the bottom panel correspond to reductions of entropy. As illustrated in the figure, the lexical information produced by the first stimulus (at time 0) takes one of four values. If the stimulus is {SVO}, then the lexical entropy drops to 0 and 3.585 bits of information are produced. If only the V stimulus appears, then the number of possible word sequences is reduced to $N/2$, and hence the entropy is reduced by 1.0—in other words, the V stimulus provides 1 bit of information. If the first stimset is either {S} or {O} (i.e. {N}), then the number of possible lexical sequences is reduced to $N/3$, so the stimset provides 1.585 bits of information; at this point, the remaining entropy is 2 bits because there are two possible V stimuli and two possible N stimuli. If any two stimuli appear (i.e. {XX}), then there is just one bit of entropy remaining.

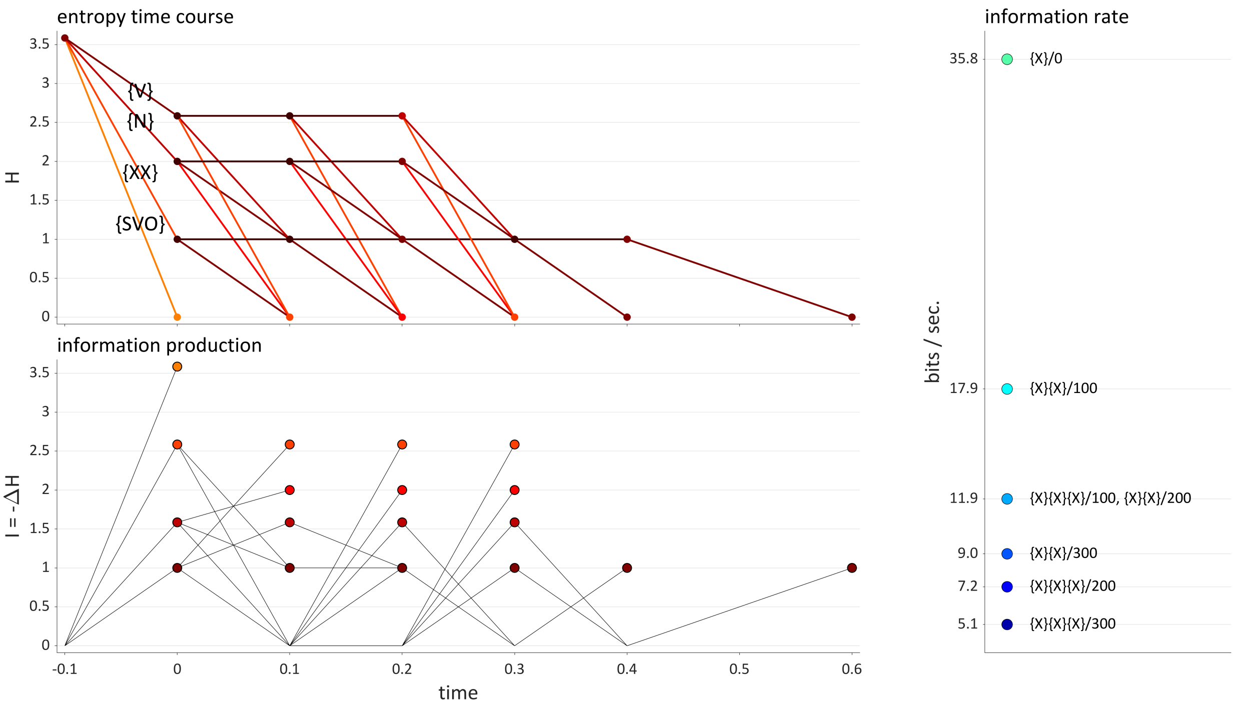


Fig. 34. Lexical entropy time-course in Exp. 1, along with information production and information production rate for each timing pattern.

Thus, the information produced by a single S or O (i.e. N) stimulus depends upon whether another N stimulus has already appeared. If it is the first N stimulus, then the entropy reduction/information produced is $-log_2\left(\frac{1}{3}\right) = 1.585$ bits. If it is the second N stimulus, then the lexical information produced is only 1 bit, because there are only two possible N items available (this is because one of the three has already been used for the first N stimulus). The information rate for a given stimulus timing pattern is defined here as the ratio of information produced to a period of time in which the entropy is non-zero. Here the period of time is defined to extend from 100 ms before the first stimset to the time of the last stimset. Note that it is necessary to select an arbitrary starting time of this period that is prior to the first stimset, otherwise the information rate for the {SVO} pattern would be undefined. Thus the maximal information rate is 3.585 bits / 100 ms = 35.85 bits/s.

As illustrated in Fig. 34, there are six unique information rates in Exp. 1, and all but one of them is associated with a unique combination of the number of stimsets and ISI. Specifically, the third highest rate (11.9 bits/s) is common to timing patterns which have three stimsets at a 100 ms ISI or two stimsets at a 200 ms ISI. This shows that the relative information rates between timing patterns correspond to the time of the last stimset. The same relation holds for the average lexical information rate of timing patterns in Exp. 2, shown in Fig. 35.

Differences in the average lexical information rate (i.e. lexical information rate averaged over all timing patterns) could be implicated in between-experiment differences in RT. The average lexical information rate in Exp. 1 (12.5 bits/s) was less than the average information rate in Exp. 2 (20.9 bits/s). Perhaps a higher average information rate leads participants to adopt processing/preparation strategies that allow lexical information to be processed more quickly. This could account for the faster RTs in Exp. 2 than in Exp. 1.

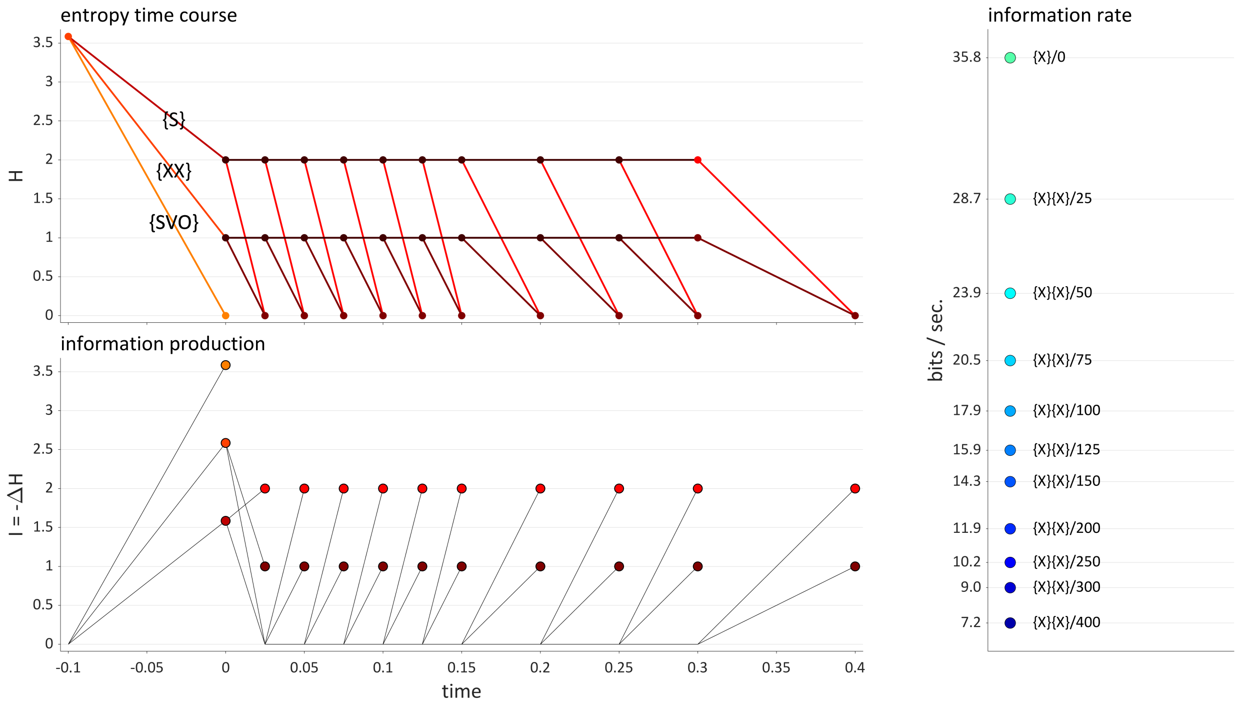


Fig. 35. Lexical entropy time-course in Exp. 2, along with information production and information production rate for each timing pattern.

However, there is reason to doubt that lexical information rate is an important factor in RT patterns. As shown in Table 21, for Exp. 1 the lexical information rate is not nearly as good a linear predictor of RT_[S] as either of the Δ measures; nor is it as good as t_S for RT_first. In contrast, for Exps. 2 and 3 (where RT_[S]=RT_first), the lexical information rate is nearly as predictive as ΔVS, having comparable R^2 values. However, this may simply be attributable to the fact that information rate is more closely related to ISI in Exps. 2 and 3. In Exp. 1, where three-stimset/100 ms patterns and two-stimset/200 ms patterns have the same information rate, we see a weaker predictive value of information rate.

Table 21. Comparison of information rate, ISI, and Δ-measures as predictors

Exp1.					**Exp. 2**			**Exp. 3**		
	RT_[S]		**RT_first**							
predictor	**R^2**	**AIC**	**R^2**	**AIC**	**predictor**	**R^2**	**AIC**	**predictor**	**R^2**	**AIC**
ΔVS	0.51	-9116	0.24	-4624	ΔVS	0.64	-8483	ISI	0.55	-14182
ΔOS	0.41	-8238	0.30	-5064	ISI	0.63	-8369	ΔVS	0.52	-14134
t_S	0.37	-7959	0.59	-7959	rate_I	0.61	-8211	rate_I	0.50	-13814
ISI	0.27	-7254	0.43	-6245	ΔOS	0.59	-8023	ΔOS	0.44	-13449
rate_I	0.27	-7231	0.41	-6113						

What effects on RT might we expect of informational differences between experiments? The differences between experiments are summarized in Table 22. For exposition we refer to table rows (a)-(j). In general, it could be hypothesized that when there is more uncertainty, the processing of the stimuli is more difficult, and thus RT should be longer. This prediction is consistent with the empirical differences for lexical uncertainty after $\{\}^1$ (b), spatial uncertainty before $\{\}^1$ (e), and spatial uncertainty after $\{\}^1$ (g). The prediction is not consistent with differences in temporal uncertainty before and after $\{\}^1$ (h,j), where Exp. 2 has more temporal uncertainty than Exp. 1. This could suggest that spatial and/or lexical uncertainty is more influential than temporal uncertainty.

Table 22. Comparison of experiments with respect to information production and entropy

		Exp. 1	Exp. 2	Exp. 3	matches empirical
(a)	$\{\}^1$ lexical information produced	2 > 1		*	y
(b)	H^1 lexical uncertainty remaining after $\{\}^1$	1 > 2		*	y
(c)	$\{\}^2$ lexical information produced	1 > 2		*	n
(d)	Average lexical information rate	2 > 1		*	y
(e)	H^0 spatial uncertainty before $\{\}^1$	1 > 2		*	y
(f)	$\{\}^1$ spatial information produced	2 > 1		= 0	y
(g)	H^1 spatial uncertainty after $\{\}^1$	1 > 2		= 0	y
(h)	H^0 temporal uncertainty before $\{\}^1$	2 > 1		= 0	n
(i)	$\{\}^1$ temporal information produced	2 > 1		= 0	y
(j)	H^1 temporal uncertainty after $\{\}^1$	2 > 1		= 0	n

*depends on block

An alternative hypothesis specific to temporal information is that greater temporal uncertainty facilitates stimulus processing. Requiring participants to process stimuli in a larger region of timing space might promote more rapid stimulus integration—perhaps more attention is devoted to predicting stimset timing. In that case, the empirical patterns are consistent with the greater degree of temporal uncertainty in Exp. 2 than in Exp. 1.

Another sensible hypothesis is that when more information is produced by a stimset, the response can be prepared more quickly, and so RTs should be faster when more information is produced earlier on. This prediction is consistent with the lexical, spatial, and temporal information produced by $\{\}^1$ (a, f, i), but not the lexical information produced by $\{\}^2$ (c), where Exp. 1 produces slightly more information on average than Exp. 2. The average lexical information rate is also higher for Exp. 2 than Exp. 1, consistent with the hypothesis that greater information production facilitates response preparation.

An alternative hypothesis is that, when more information is produced by a stimset, it is more difficult to process that information, in which case RT should be slower. This is not consistent with the empirical differences between Exp. 1 and 2 for $\{\}^1$ (a, f, i). It could be argued that because Exp. 3 produces no spatial or temporal information and has the shortest RTs, it provides support for the hypothesis that more information is more difficult to process. However, it is worth considering that on a larger timescale, Exp. 3 provides the spatial and temporal information at the beginning of each block.

Why are RTs much faster in Exp. 3 than Exp. 2? Possibly it is because Exp. 3 has no spatial or temporal uncertainty. Consider that the average lexical information produced by $\{\}^1$ Exp. 3 varies by block. For {SVO} blocks in Exp. 3 there is more information produced by $\{\}^1$ than the average information for Exp. 2; for {SV} and {SO} blocks it is the same amount of information produced; for {S} blocks there is less information produced by $\{\}^1$ in Exp. 3 than the Exp. 2 average. Despite this variation across blocks, Exp. 3 RTs are faster for all of the orderings (excluding the early blocks for which practice effects are a confound). This could suggest that the lexical information typically produced by the first stimulus does not have as large of an effect on response preparation as uncertainty regarding the spatial arrangement of the first stimset (or relatedly, how much spatial information is produced by $\{\}^1$).

To summarize, it seems likely (i) that lexical and/or spatial uncertainty have greater slowing effects than temporal uncertainty; (ii) that spatial uncertainty has greater effects than lexical uncertainty; and (iii) that earlier information production (or greater information production rate) facilitates response preparation. It is also possible that temporal uncertainty actually facilitates stimulus processing, perhaps by encouraging participants to shift attention more rapidly after stimuli. The relative importance of spatial uncertainty over temporal uncertainty might be explained by the dynamics of gaze control, which we consider next.

Speculations on gaze control

Future studies could benefit from analysis of gaze dynamics during experimental trials. Here we will speculate on where participants might be looking during the experiment, and raise a variety of questions regarding how gaze may be controlled to facilitate information processing.

A preliminary question is whether participants need to saccade during the stimulus presentation. It is logically possible that participants are able to process all stimuli para-foveally from a central gaze location without needing to saccade. Studies of visual attention/processing have found that the spatial dimensions of the window in which information can be processed can vary with task (LaBerge, 1983; Ludwig et al., 2014) and may range from 0.5° to 2°. Here participants generally sat approximately 0.75 to 1.25 m from the stimulus monitor, which was a 19 inch (0.48 m) monitor with horizontal and vertical dimensions of (0.34 m). The horizontal offsets of the centers of the S and O stimuli were 10% of horizontal screen width, which is 0.034 m. Thus the visual angle between S and O was between $2\tan^{-1}(0.034/1.25) = 3.1°$ and $2\tan^{-1}(0.034/0.75) = 5.2°$. This suggests that it may be necessary, or at least advantageous, to saccade from one location to another to process information. However, the size of the window may be larger if the relevant para-foveal information is easy to distinguish, so the necessity of a saccade in the current contexts remains unclear. Nonetheless, impressionistic observation of the participants during pilot studies indicates that eye movements do occur during stimulus presentation, and so we will subsequently assume that participants are in fact controlling eye movements, whether it is necessary or not.

A second question is what are the most effective gaze control strategies, when it comes to minimizing RT. Participants are probably not randomly directing gaze, and there is most likely some correlation between the time-course of gaze and the timing pattern of stimuli. First we consider the gaze location prior to the first stimset; note that we refer to the *locations* of stimuli as <S>, <V>, and <O>, to distinguish locations from the stimuli themselves. Several different strategies seem sensible here. One strategy (*<S>-first*) is to direct gaze to <S> before the first stimset. This might be an effective strategy because the initiation of the response depends on knowledge of the lexical identity of S, and therefore places a hard constraint on response initiation. By initially locating gaze on <S>, the participant avoids the need to saccade to <S> should the S appear. An alternative strategy (*central fixation*) is to direct gaze to the location of the fixation cross <+>, which is in the center of the stimuli triangle. The central fixation strategy minimizes the (angular) distance of the saccade to any particular stimulus location. Yet another strategy might be to direct the gaze to a location that is between stimuli locations, such as halfway between <S> and <V>. If the gaze location is close enough to <S> and <V> that either stimulus can be processed without a second saccade, then this might be preferable to either the <S>-first or central fixation strategies.

One consideration in developing hypotheses about initial gaze is that the differences between experiments are likely to affect gaze control strategies. For instance, in Exp. 2 where S always appears in the first stimset, the <S>-first strategy might be most effective; in contrast, in Exp. 1, where S only appears in the first stimset on 18/39 = 46.2% of trials, central fixation might be more effective. In Exp. 3 where the timing and location of all stimuli is known in advance, a strategy that locates initial gaze between stimuli might be most effective.

A third question is the extent to which saccades subsequent to the initial fixation are influenced by the timing/ordering of stimuli. It is logically possible that there is no influence: for example, participants might cyclically saccade through locations and only process lexical information when a stimulus is present. However, this would not be a very effective strategy. Another logical possibility is that saccades occur in a fixed order but are entirely contingent on stimulus processing. For example, in an <S>-first strategy, participants might wait until S appears before saccading to <V> (regardless of whether V was visible), then wait for V (if necessary), then once V is processed, saccade to <O>. This strategy would likely be suboptimal for timing patterns in which V or O precedes S by a substantial amount of time. Indeed, the early V/O facilitation effect suggests that participants do not adopt this approach, because such an approach predicts no difference between {V}{O}{S} and {SVO} orderings.

A more optimal strategy for gaze control would be one in which gaze control is governed both by a predictive model of stimuli and a model of the relative importance of the stimuli for response initiation. For example, on a {V}{O}{S}/300 trial in Exp. 1, imagine that initial gaze is on <S> (because of its relative importance), but when O appears the participant saccades to <O> and processes O. Subsequently, the participant saccades back to <S>, because S is more important than V. However, V appears and this induces a saccade to <V> and processing of V. This is followed by a saccade to <S> because it is predictable with 100% certainty that S will appear there shortly. Now consider a {VO}{S}/300 trial, again with initial gaze on <S>. Here the participant saccades after the first stimset to <V> and then to <O>, and then to back to <S>. In both cases, the control of gaze is driven both by stimuli appearance, but also by anticipation of where subsequent stimuli may appear and a consideration of the relative importance of the stimuli for response production. We must assume in these cases that participants can detect the spatial pattern of stimuli in their parafoveal vision—hence the participant can see that stimuli have appeared at both <V> and <O> locations even when their gaze is on <S>.

Some important considerations in developing a model of gaze control during the task are the fixation time required for recognition of lexical identity of stimuli and the time required for planning saccades to a subsequent location. Because the set of stimuli in the experiment is small and the locations of those stimuli are known, it is likely that fixation time required for processing stimuli is relatively small, most likely on the short end of reported fixation durations in reading (50-600 ms). Moreover, although saccades to unexpected stimuli may take from 200-300 ms to plan, in the current experiments strong expectations can be formed regarding the timing and location of stimuli. Hence it may be the case that saccade planning time is greatly reduced and/or parallelized. Perhaps a small repertoire of saccades is learned by participants to facilitate performance in the task.

Ultimately, there are many unknowns regarding gaze control in the current experiments. Obtaining gaze trajectories throughout the experiment is important because these can be used to help interpret RT patterns. It is possible that RT patterns are strongly influenced by gaze dynamics, in which case it is important to factor out those dynamics to determine what role, if any, syntactic organization has.

A more comprehensive model

The simplified model can be optimized to generate mean reaction times on an experiment-by-experiment basis with a fair degree of accuracy, but it is not sufficiently powerful in a number of ways. One shortcoming is the absence of mechanisms which associate lexical items with motor plans, and which govern the order in which motor plans are selected along with the relative timing of their execution. Such mechanisms are necessary to account for durational effects of ΔVS on [S], specifically the lengthened duration of [S] when V is delayed relative to S. They are also necessary to account for the occurrence of blends, where the motor actions of two distinct targets are produced together. Although blends are somewhat rare, they occur frequently enough that they should not be ignored; they show that selection of motor plans is not simply a mapping from selected lexical items to articulatory gestures. A more comprehensive model should include mechanisms of motor sequencing and coordination (see Tilsen (2019) for ideas on how such blends and other articulatory errors might arise).

Another shortcoming of the simple model is the absence of mechanisms for detecting and repairing errors. Hesitation, repetition, and cutoff disfluencies are common in association with error patterns; this indicates that a more comprehensive model should include self-monitoring mechanisms which can lead to various repairs. The simplified model also does not generate effects of spatial-temporal uncertainty: differences in RT between experiments are not predicted. Such effects may be due to between-experiment differences in the strategies participants use to control gaze, and thus a more comprehensive model would benefit from an explicit representation of gaze control, which in turn constrains when task-relevant information becomes available.

The shortcoming of the model which seems most pressing to address is its inability to generate lexical substitution errors, and specifically the asymmetry between internal vs. external sources of S and O lexical substitutions: internal substitutions of S were more frequent than external ones, while external substitutions of O were more frequent than internal ones. The simplified model cannot generate substitution errors because it does not have an explicit mechanism whereby lexical items cued by the stimuli are associated with S, V, and O syntactic systems (or more neutrally, the 1st, 2nd, and 3rd words of the response). As implemented above, the model merely assumes that the correct lexical items are correctly associated with the three words of the response. Lexical substitution errors show that this assumption is incorrect, but in a way that only allows for substitutions between S and O, and not between V and S, or V and O.

Despite its many shortcomings, the simplified model does give us a way of reasoning about how the experimental phenomena arise. The model can generate the RT patterns which exhibit Δ-insensitivity for small |ΔVS| and |ΔOS|, (ii) early V/O facilitation, and (iii) late V delay. Specifically, the model holds that early V/O facilitation arises from an absence of interference effects, and late V delay arises from the dependence of the initiation criterion on the |V| system state. Δ-insensitivity arises because the interference and initiation-dependence effects oppose one another. The simplified model implements the idea from Tilsen (2019) that conceptual systems interfere with each other in the process of forming stable resonances with syntactic systems. In that way, the experimental findings can be interpreted as evidence in support of the hypothesis that response initiation can be understood as the relaxation of a system toward an attractor, which amounts to an increase in order and stability of the system.

One of the key principles of this view is that the experience of meaning requires that the relevant conceptual systems obtain a stable, coherent state space trajectory, which is accomplished via their interactions with syntactic systems. The current experiments do not guarantee that participants engage in meaning experiences, only that they select and execute the relevant sets of articulatory gestures in the correct order. It seems possible that the relevant meaning experiences, if they do arise in the experiments, may stabilize after response initiation; all that is logically required for response initiation is stabilization of the conceptual-syntactic system associated with S, which is a pre-requisite for selection of S-associated motor plans. Nonetheless, the fact that V- and O- stimulus timing do have substantial influences on response initiation shows that there are interactions between the verb and the arguments; whether those interactions are interactions between conceptual-syntactic systems or gestural-motoric systems remains an open question.

There are many potentially interesting questions that can be investigated in the current experimental paradigm. For one, will Δ-space RT patterns differ in languages with other word orders? An SOV language like Japanese might be expected to show early O facilitation and late O delay, in which case we might conclude that RT patterns are driven more by sequencing mechanisms that syntactic organization; alternatively, if the Japanese RT patterns are similar to those observed here, it would suggest that the RT patterns are attributable to syntactic organization, rather than word order. Another question is whether RT patterns will differ from those observed when the subject and object arguments are allowed to be identical, i.e. *Moe saw himself*. It will also be helpful to investigate the effects of variation in the informational content of the stimuli, i.e. allowing for larger or smaller sets of lexical items for each syntactic role, and exploring the consequences of manipulating statistical dependencies between those items. Perhaps a task which elicits questions by presenting an ambiguous noun or verb stimuli may shed light on the organization of interrogatives.

Finally, my hope in conducting the current experiments and in presenting the above analyses is to stimulate interest in experimental paradigms which investigate utterance *generation* (or production), as opposed to utterance comprehension. The vast majority of psycholinguistic/experimental syntax research is focused on how speakers comprehend (i.e. parse) written or spoken language. Here our interest is in the conditions necessary for producing speech. Of course, the content of that speech is determined by external stimuli, but this is ultimately not different from what induces us to speak in everyday contexts—we experience the world around via our sensory organs, and this sensory information can lead—more or less directly—to cognitive states in which we generate speech. It is also important, in studying utterance production, to start small: the focus here on basic SVO utterances with a sparse lexicon helps us conduct more powerful analyses, allowing for the detection of small effects which might be obscured with more complicated designs.

Methods

Participants, task, and stimuli

Participants were undergraduates and staff at Cornell University, Ithaca Campus. All were native speakers of English, with no self-reported speech or language disorders. A total of 56 speakers participated in three experiments (18, 18, and 20 participants in Exps. 1, 2, and 3, respectively). Data from one participant in Exp. 2 were excluded from all analyses because they fell asleep several times during the experiment. Data from one participant in Exp. 3 were excluded because the speaker appeared to intentionally produce incorrect responses during part of the experiment.

Participants were seated in a chair in front of a computer monitor in a sound-attenuating booth in the Cornell Phonetics Lab. Prior to beginning the experiment, participants went through an instructions interface with the experimenter present. The experimental task was framed as speaking to an AI system, and the instructions placed emphasis on responding both quickly and clearly. The instructions for all three experiments are provided in Table 23 (Each cell of the table contains the text that appeared on a separate screen. Line breaks are removed to save space; bolded and italicized text appeared as such on the screen; numbers are added for reference):

Table 23. Instructions for experiments

(1) In this experiment, you will be speaking to an A.I. system. On each trial, three words will appear on the screen. Your task is to say a sentence with those words. Click "Next" to see an example.	(2) The words will appear in a pattern such as below: [image of stimulus pattern] The sentence that you should say in this case is: *Lee saw Ray* Click "Next" to see another example.
(3) [image of stimulus pattern] The sentence that you should say in this case is: *Moe heard Lee* Note that the subject of the sentence is always shown on the left, and the object is always shown on the right. Click "Next" to proceed.	(4) Note that a cross appears before the words do on each trial. You should look at the cross at the start of every trial. Click "Practice" to practice.
(5) You will be given a score on every trial of the experiment except for the first trial. The score ranges from 0-100 and is based on how quickly the A.I. system recognized the sentence that you produced. Click "Next" to proceed.	(6) Click "Practice" to practice. You will be shown random scores for these practice trials only.
(7) In order for the A.I. system to correctly recognize the sentence **you must speak clearly**. But, in order to receive a high score, **you must speak quickly**. Click "Next" to proceed.	(8) Whenever you get a high score, you will receive a bonus to your compensation. You get the bonus whenever your score is over 50, and the higher your score is, the larger the bonus is. Every once in a while you will be shown the total amount of bonus compensation that you have earned. Click "Next" to proceed.
(9) Remember: To achieve high scores, you must **speak both clearly and quickly**. In order to achieve high scores, you should always **begin producing the sentence as soon as you can**, and **produce the sentence quickly**. The score is based on when you finish producing the sentence. Note that over the course of the experiment, it will become more difficult to achieve high scores. Click "Next" to proceed.	(10) Please keep the following in mind: 1. Never touch the microphone that you are wearing. 2. Avoid making extra noises during the trials. 3. Try to keep the volume of your voice at a normal level. If you speak too quietly the A.I. system may have difficulty recognizing the sentence. Click "Next" to proceed.
(11) IMPORTANT: Try to say the sentences the same way throughout the experiment. If you purposefully change how you say the sentences, the A.I. system will have difficulty recognizing them. Do not add extra emphasis to any particular word in the sentence. Say the sentence in a plain manner. The experimenter will be monitoring your responses and will stop you if you fail to follow these instructions. If you make an error, try to finish saying the sentence correctly. If you have any questions, please ask the experimenter now. Click "Next" to proceed.	(12) You are almost ready to start the experiment. The experiment will last for about 50 minutes. If something seems to go wrong during the experiment, or the experiment unexpectedly halts, you can let the experimenter know. Remember: respond quickly and speak quickly. Click "Next" to proceed to the experiment.

Where indicated above (screens 4, 6), participants performed a set of five practice trials with stimuli/timing patterns that were randomly selected from the set of all stimuli/timing patterns in the experiment (these differed between experiments). If participants did not produce the sentence with normal declarative sentence intonation, the experimenter demonstrated the desired pronunciation for them.

Each trial during the experiment began with the appearance of fixation cross at the center of the screen, which remained visible for 750 ms until the first stimulus set appeared. Audio was recorded with a AKG C520 headset microphone at 22050 Hz from 500 ms prior to the first stimulus until 2000 ms after the last stimulus. Stimuli remained visible until the end of the recording.

The design variables of the experiments are summarized in Table 24. The main differences between experiments were (i) that Exp. 1 used all 13 orderings while Exps. 2 and 3 used only the four S-first orderings; (ii) Exp. 1 used a smaller set of ISIs than Exps. 2 and 3; and (iii) Exp. 3 blocked timing patterns to eliminate spatial and temporal uncertainty in stimuli. The number of unique timing patterns in each experiment was different. Exp. 1 tested 37 unique timing patterns. Each block of Exp. 1 contained 39 trials, resulting from crossing the 3 non-zero ISIs with the 13 orderings; because {SVO} has an ISI of zero, {SVO} ordering was repeated three times in each block. Exp. 2 tested 31 unique timing patterns. Each block of Exp. 2 contained 40 trials, resulting from crossing the 10 non-zero ISIs with the 4 orderings; thus {SVO} ordering was repeated ten times in each block. Exp. 3 tested 19 unique timing patterns. The {SVO} pattern was tested in five blocks. Blocks were ordered by decreasing ISI (for non-synchronous orderings) and following the order: {SVO}, {SV}{O}, {SO}{V}, {S}{VO}; hence the timing patterns by block in Exp. 3 were {SVO}/0, {SV}{O}/300, {SV}{O}/300, {SV}{O}/300, {SVO}/0, {SV}{O}/200, …, {SVO}/0, {SV}{O}/25, {SV}{O}/25, {SV}{O}/25.

Table 24. Comparison of experiment design variables

	Experiment 1	**Experiment 2**	**Experiment 3**
description	all orderings	S-first orderings, randomized	S-first orderings, blocked timing patterns

orderings	{SVO},
	{SV}{O}, {SO}{V}, {S}{VO}
	{VO}{S}, {V}{SO}, {O}{SV}
	{S}{V}{O}, {S}{O}{V}
	{V}{S}{O}, {V}{O}{S}
	{O}{S}{V}, {O}{V}{S}
		{SVO},
		{SV}{O}, {SO}{V}, {S}{VO}
			{SVO},
			{SV}{O}, {SO}{V}, {S}{VO}

ISIs	0, 100, 200, 300	0, 25, 50, 75, 100, 125, 150, 200, 250, 300, 400	0, 25, 50, 75, 100, 200, 300
num. timing patterns	37	31	19
num. trials/block	39	40	22
num. {SVO}/block	3	10	n/a
num. blocks/session	12	12	24
num. trials/session	468	480	528
num. participants	18	17	19

The word forms used in the experiment were {*Lee*, *Moe*, *Ray*, *heard*, *saw*}. There are 3 × 2 × 2 = 12 unique SVO sequences of these word forms, under the constraint that the O is never identical to the S. In Exps. 1 and 2, all word form combinations were presented once with each timing pattern, distributed randomly over the 12 blocks. In each block of Exp. 3, each of the 12 unique word form sequences was included once, along with an additional 10 selected randomly without replacement from a bag of unused word form sequences. The bag was refilled with all 12 sequences whenever it was emptied, thus ensuring that the word form sequences were as evenly distributed as possible over the entire session.

The noun forms {*Lee*, *Moe*, *Ray*} were chosen because they had simple CV forms and were comprised of sonorant onset consonants; because these sounds are normally voiced, estimates of response onset are expected to be closer in time to the initiation of articulatory gestures. Note there is unavoidably some delay between gestural initiation and its acoustic consequences. It was desired that the vowel and consonant categories of the forms differ as well, in order to diminish potential confounds from phonological similarity. The verb forms {*heard*, *saw*} were chosen to be monosyllabic and to have similar semantic qualities, both being frequent verbs of sensory perception. One reason for choosing these verbs is that in future versions of the experiment it is desired to cue the sentence context with visual scenes rather than orthographic forms. A drawback of using *heard* is that it has an obstruent coda, unlike *saw*. Both verbs nonetheless are heavy syllables—a short vowel-codas sequence [ɚd] in *heard* and the long vowel [a:] in *saw*. Past tense was used rather than present tense because the past tense seems to be a more natural way to describe a visual scene.

One thing to take note of is that the ISIs specified in the experiments do not exactly correspond to actual ISIs of visual stimuli, because the screen only refreshes at specific intervals. To characterize the discrepancies that arise from this, we distinguish between CPU clock-time and screen refresh times. Commands to display stimuli and to begin audio recording are synchronized in CPU clock-time, and relatively precise synchronization of these commands can be achieved. However, the current experiments did not synchronize these commands with screen refreshes. The refresh rate of 60 Hz restricts changes to visual objects on the screen to occur at intervals of δscr = 16.667 ms. The stimuli in the experiments were controlled with Matlab timer objects, which operate in CPU clock time and have a precision of 1ms. The timers were used to issue commands to make visible pre-constructed graphics objects for the fixation cross and for the S, V, and O stimuli. However, since the timing of these commands relative to screen refresh times is not controlled, it is reasonable to expect that the actual time a stimulus becomes visible after issuing the command will be delayed from 0 to δscr, with the delays being uniformly distributed.

To see the consequences of lack of synchronization of stimuli commands with screen refreshes, consider a series of two stimuli, stim1 and stim2, with an ISI of 25 ms. Fig. 36 shows how the actual times of stim1 and stim2 vary as a function of the stim1 command time, which we define relative to a periodic screen refresh which occurs at time 0 and multiples of 16.7 ms (i.e. 60 Hz).

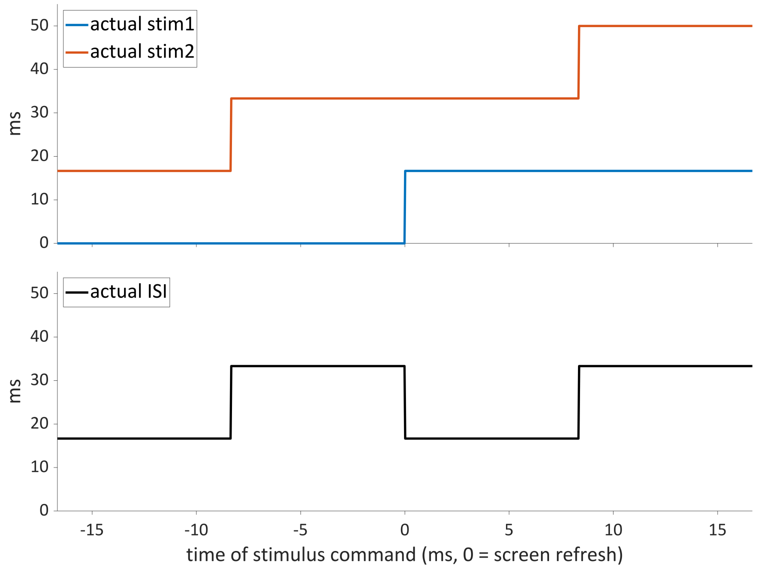


Fig. 36. Illustration of effects of screen refresh on actual ISI.

By subtracting the actual stim1 time from actual stim2 time, we can see that the actual ISI is either one or two times the screen refresh period, i.e. 16.7 or 33.3 ms, and that we should expect about equal numbers of both (Fig. 36, black line). Note that these are equal to the target ISI ± δscr/2. In general, whenever the ISI is not an integer multiple of δscr, we can expect evenly distributed actual ISIs equal to the target ISI ± δscr/2 (when ISI is an integer multiple of δscr, we can expect actual ISI to equal target ISI). Whether these deviations of the actual ISI from the target ISI are problematic is hard to know. The deviations are fairly small, amounting to ±8.3 ms. Moreover, since they are in opposite directions and evenly distributed, their effects relative to the target ISI might cancel out. However, the above analysis does not consider other sources of variation such as buffering of graphics object updates. In future experiments, it may be helpful to ensure that stimulus commands are synchronized to screen refreshes.

Online response processing

During the experiment, the acoustic recording of each trial was processed in order to give the participant feedback on response speed, with accuracy being taken into account. The following procedure was used. The absolute value of the recorded audio signal was zero-phase lowpass filtered (10 Hz cutoff, 200-point FIR filter). The response offset was defined as the last sample of the lowpass filtered signal above 5% of the maximum. The audio signal was converted to a matrix of MFCC vectors, using the following parameters: window 30 ms, step: 5 ms, frequency range: [300, 5000] Hz, coefficients: 18, pre-emphasis argument: 0.97, channels: 20, lifter coefficients: 22. Signals in each frame were hamming-windowed. Delta MFCC coefficients were included. The MFCC matrix was then provided as input to the recognition network (see below). For each time frame, the recognition network outputs a vector of probabilities associated with the response categories and silence: {*Lee*, *Moe*, *Ray*, *heard*, *saw*, SILENCE}. These probabilities were smoothed with a moving average window of 25 ms, and a timeseries of detected categories was defined as the category with the highest probability in each frame.

To assess whether the response was correct, evidence for the target categories (i.e. the words that comprise a correct response on a given trial) was sought in the detected category timeseries, using the following algorithm. For the first target category (i.e. the subject), frames of the category detection timeseries were looped through and matches of the target were counted. Once 10 matches were counted (i.e. 50 ms), matches of the next target (i.e. the verb) were counted, and so on. Thus a response was counted as correct if the target words were detected for at least 50 ms each in the correct order. The recognition time—time at which the correct response is fully detected—was defined as the time at which the third target form was identified. Thus the recognition time was typically about 50 ms after the acoustic onset of the object word form. The RT for online feedback was defined as the delay between the recognition time and whichever stimulus came later, S or V. Hence the online RT measure is related to RT_last_[SV] (i.e. t_{REF} = max(t_S, t_V)), except that it encompasses additional time associated with the production of the first two words of the target sequences. The relativization of the measure to max(t_S, t_V) was used because pilot tests showed that this resulted in a low variance measure. It is important that the speaker feels that their RT is being measured accurately. Measures relativized to S or max(t_S, t_V) are the two most appropriate measures in this regard because response initiation is absolutely contingent on t_S and to a lesser extent depends on t_V. There is a logical possibility that participants became aware of the fact that RT was measured relative to max(t_S, t_V) and that this influenced their behavior; however, this seems very unlikely given the random variation in stimulus timing patterns from trial to trial. Moreover, no participants in pilot testing reported awareness of this. If the correct response was not detected, the RT was defined as 250 ms after the acoustic offset of the response. Participants were not directly informed when the correct response was not detected, but detection failures led to low feedback scores, as described below.

Beginning after the third trial of each session, the participant received a feedback score. They were instructed that the score "is based on when you finish producing the sentence", and that to achieve high scores they "should always begin producing the sentence as soon as you can, and produce the sentence quickly" (see Table 23, screens (7,9)). The score was derived from the online RT after each trial as follows. First, the set of all RTs from the session were transformed to z-scores, and any values in the upper or lower 1% of the distribution (i.e. $|z| > 2.326$) were excluded. The mean (μ_{RT}) and standard deviation (σ_{RT}) of the remaining values were calculated. The score of the current trial was then defined as one minus the value of the cumulative density function of the RT of the current trial *i*, assuming a normal distribution with μ_{RT} and σ_{RT}, i.e. $\mathrm{score}_i = 1 - \mathrm{normcdf}(RT_i, \mu_{RT}, \sigma_{RT})$. Hence the score is confined to the interval [0, 1] and values > 0.5 are faster than the mean of the distribution of all previous RTs. The score displayed to the participant was multiplied by 100 and rounded to the nearest integer. This score was presented for 1 s after the trial.

Every 24 trials, participants were also informed of a bonus to their compensation, based upon the most recent 24 trials. The bonus calculation was defined as $\$0.02\sum_{i-23}^{i}[2\max(score_i - 0.5{,}0)]$. Thus participants received up to 2 cents per trial when their RT was below the mean. The exact amount depends on how far the RT is below the mean, relative to the standard deviation. An important aspect of this design is that over the course of a session, as participants become more adept at the task, it becomes more difficult to receive higher scores because the mean of their RT distribution becomes lower. Hence participants are encouraged to respond as quickly as possible in a way that depends on a criterion that is specific to their performance. During the bonus presentation, participants were informed of the total bonus earned up to that point, the bonus over the last 24 trials, their average score over the last 24 trials, and the number of trials remaining. This information remained visible for 5 seconds and during a 5 s countdown to the beginning of the next trial. Every two bonus periods (48 trials), there was a longer countdown of 10 seconds.

The online recognition network was initially trained on hand-labeled data from pilot tests, and subsequently retrained with additional hand-labeled data from early participants in Exp. 1. Only word labels were used for this purpose. The initial training dataset included three repetitions of each of the 12 unique lexical sequences, randomly selected from 12 pilot sessions with different speakers, thus a total of 432 sentences. The retraining included three repetitions of the 12 unique sequences from an additional 7 speakers who had participated in Exp. 1, thus a total of 720 sentences. Of particular importance was speaker generalization, i.e. the ability of the network to perform well on new speakers. Some hyperparameter testing was conducted to attempt to optimize generalization performance. However, because the hyperparameter space is very large, it was not possible to systematically explore anything but a small portion of that space. To assess generalization, training was conducted with one session held out, repeating for each session. Accuracy on held-out data was assessed using the correct response detection procedure described above. The final network design and parameters (see below) had an average accuracy of 98.6% correct on held out sessions.

The final optimized network consisted of three 400-unit bidirectional LSTM layers, each followed by a 50% dropout layer; these were followed by a fully connected layer, softmax layer, and classification layer. Input MFCC matrices (described above) were zero-centered by dimension. Input sequences were sorted by longest and shuffled every epoch. The following training parameters were used: optimizer: Adam; gradient threshold: 1.0; initial learning rate: 0.001; learning rate drop period: 50; maximum epochs: 200; mini-batch size: 64; validation frequency: 20; validation patience: 20. Half of the hand-labeled data were used for training, the other half for validation.

Data processing and analysis

Trials were segmented for offline analysis by forced alignment with Kaldi (Povey et al., 2011). Monophone 5-state HMMs were trained on MFCCs with a subset of hand-segmented data from all sessions (including pilot sessions). One trial of each of the 12 unique lexical sequences was randomly selected for hand-labeling from each session, excluding trials from the initial block of the session. There were a total of 720 hand-labeled tokens used for the training, from 60 different speakers. MFCCs had window sizes of 25 ms and frame steps of 5 ms, with 16 cepstral coefficients over the range [50, 10000 Hz]. The monophone HMMs were used for forced alignment of all data. Visual inspection of a random subset of trials showed that the forced aligned was highly accurate when participants produced the correct response, especially in locating the beginning of initial segment of the response, which is important for robust RT estimation. Some systematic variation was present between speakers in locations of boundaries of non-initial segments, but this variation is only potentially problematic for durational analyses, and less so if speaker is included as random factor in regressions.

On trials with silences, non-speech noise, or incorrect productions, the forced alignment results in abnormal interval durations. To correctly identify errors and exclude data when warranted, trials with alignments that met either of the following two criteria were visually and auditorily inspected: (i) there was a silent period detected between response words; (ii) the response initiation time relative to S was abnormally early or late (beyond ±2.32 st. dev. from the mean, calculated by participant). When appropriate, the forced alignments of these trials were either corrected by hand or the trials were labeled as errors. A total of 358 trials (1.5% of all trials) were identified as having errors and were excluded from analyses of RT or duration. A little more than three quarters of these (280, 78.2%) were identified as lexical substitution errors, blends, or covert errors. Other exclusions were trials in which (i) the participant produced a non-speech vocalization (i.e. a cough, laugh, or yawn), 23 trials; (ii) failed to complete the response in the recording period (which is indicative a very late response initiation), 14 trials; or (iii) an intermittent malfunction in the audio recording hardware occurred, 44 trials.

The first block of trials in each session of Exp. 1 and 2 was excluded from all analyses (39 and 40 trials in Exps. 1 and 2, respectively), because participants are becoming familiar with the task and their RT patterns are highly nonstationary during these trials. Exp. 3 analyses are treated differently because of the blocking of timing patterns. For all RT and duration analyses, RT outliers were excluded on a by-subject, by-timing pattern basis; specifically, trials with an RT which were outside of ± 2.32 st. dev. from the mean (i.e. the upper and lower 1% of a normal distribution) were excluded. Because RT distributions are generally skewed leftward with longer right tails, this exclusion strategy primarily excludes late responses. For RT_[S], there are 488 outlier exclusions (2.0%) across all three experiments for abnormally long RTs, and 26 outliers exclusions (0.11%) for abnormally short RTs. The percentages of exclusions are similar for RT_first.

The overview of results in Fig. 2 shows RT_[S] means and confidence intervals derived from combining Exps. 1 and 2 data; to do this, a mixed effects regression was calculated with participant and experiment as random effects, with an intercept as a fixed effect. The means shown in Fig. 2 are the residuals of this regression with the intercept added; hence the subject- and experiment-effects are subtracted out of these data. Elsewhere regressions were conducted within experiment and subject intercept and slope terms were included as random effects, unless otherwise indicated. Note that the full random effects structure was justified, i.e. a participant-specific intercept and participant-specific slopes for tS, ΔVS, and ΔOS; correlations between these were included as well.

For optimization of models, a particle swarm global optimization was used with a swarm size of 1000 and cost function step tolerance of 0.001. The cost function was the mean absolute error between the model-generated response initiation time and the response initiation time specified in the hypothesized behavioral patterns. The RTs modeled in association with Exp. 1 and Exp. 2 were derived from the residuals of mixed effects regressions of RT_first_[SVO] with only a fixed intercept and random intercepts for participants. The fixed intercept was added to the residuals and the mean was calculated for each timing pattern. The same data were used for the linear and nonlinear regressions which are compared with the optimized models.

Acknowledgements

Thanks to research assistants Shuya Zhao, Joshua Marker, and Matthea Shames for conducting the experiments. Shuya, Josh, and Matthea, along with graduate student Forrest Davis, also contributed in testing and refining various aspects of the experimental design.

References

Hirsh, I. J., & Sherrick Jr, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, *62*(5), 423.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(3), 371.

Ludwig, C. J., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences*, *111*(2), E291–E299.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*(2), 56–61.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., & Schwarz, P. (2011). The Kaldi speech recognition toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.

Tilsen, S. (2019). *Syntax with oscillators and energy levels*. Language Science Press.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, *71*, 249–267.

Vanrullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.