PIPS: a parallel planning model of sentence production

Laurel Brehm[1], Pyeong Whan Cho[2], Paul Smolensky[3,4] and Matthew A. Goldrick[5]

[1]Max Planck Institute for Psycholinguistics

[2] Department of Psychology, University of Michigan,

[3] Department of Cognitive Science, Johns Hopkins University, [4]Microsoft Research AI,

[5]Department of Linguistics, Northwestern University

Author Note

This research was funded by INSPIRE Grant BCS1344269 and the College of Literature, Science and Arts, University of Michigan. We would like to thank the GSC team, the MPI Simulation and Modeling User Group, and Antje Meyer for helpful feedback.

PIPS: a parallel planning model of sentence production

Laurel Brehm[1], Pyeong Whan Cho[2], Paul Smolensky[3,4] and Matthew A. Goldrick[5]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, NL

[2] Department of Psychology, University of Michigan, Ann Arbor, MI, USA

[3] Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA

[4] Microsoft Research AI, Redmond, WA, USA

[5] Department of Linguistics, Northwestern University, Evanston, IL, USA

Corresponding Author: Laurel Brehm (laurel.brehm@gmail.com; PO Box 310, 6500 AH Nijmegen, NL)

Abstract

Subject-verb agreement errors are common in sentence production. Many studies have used experimental paradigms targeting production of subject-verb agreement from a sentence preamble (*The key to the cabinets*) and eliciting verb errors (... **were shiny*). Through reanalysis of previous data (50 experiments; 102369 observations), we show that this paradigm also results in many errors in preamble repetition, particularly of local noun number (*The key to the *cabinet*). We explore the mechanisms of both errors in PIPS (Parallelism in Producing Syntax), a model in the Gradient Symbolic Computation framework. PIPS models sentence production using a continuous-state stochastic dynamical system that optimizes grammatical constraints (shaped by previous experience) over vector representations of symbolic structures. At intermediate stages in the computation, grammatical constraints allow multiple competing parses to be partially activated, resulting in stable but transient conjunctive blend states. In the context of the preamble completion task, memory constraints reduce the strength of the target structure, allowing for co-activation of non-target parses where the local noun controls the verb (notional agreement and locally-agreeing relative clauses) and non-target parses that include structural constituents with contrasting number specifications (e.g., plural instead of singular local noun). Simulations of the preamble completion task reveal that these partially-activated non-target parses, as well the need to balance accurate encoding of lexical and syntactic aspects of the prompt, result in errors. In other words: because sentence processing is embedded in a processor with a finite memory and prior experience with production, interference from non-target production plans causes errors.

PIPS: A parallel planning model of sentence production

1. Introduction

Speaking is a hard act that feels deceptively easy: producing an utterance requires that the speaker select, plan, and articulate only one out of many possible ways of expressing any given message. Alternate production plans — things that could have been said — influence speech at multiple levels of representation, with effects of co-activated lexemes/lemmas on the phonological level (synonym effects: Jeschniak & Schriefers, 1998; cognate facilitation: Costa, Caramazza, & Sebastian-Galles, 2000; phonological facilitation: Morsella & Miozzo, 2002) and effects of co-activated phonological representations on the articulatory/acoustic properties of speech (Goldrick, Keshet, Gustafson, Heller, & Needle, 2016). There is also evidence for the influence of alternate sentence plans on production (see e.g. idiom blends: Cutting & Bock, 1997; blends of two syntactic formulations attested in Frazier & Clifton, 2014), but this phenomenon has received less attention in the literature. We use a computational model to show how alternate production plans elicit verb and noun errors in a sentence completion task.

We begin the paper with an overview of nearly thirty years of experimental data showing that subject-verb number agreement errors and the mis-recall of a sentence prompt (‘preamble’ errors) both occur frequently in the same experimental paradigm. We then model these data using PIPS (Parallelism in Producing Syntax), a model in the Gradient Symbolic Computation (GSC) framework (Cho, Goldrick, Lewis, & Smolensky, 2018; Cho, Goldrick, & Smolensky, 2017, 2020; Smolensky, Goldrick, & Mathis, 2014). Using PIPS, we demonstrate how verb errors and some types of preamble errors are consequences of a human processor operating in real time with memory constraints. Specifically, transient blend states during the planning process allow portions of target and non-target structures to interact, causing interference.

1.1 Experimentally eliciting agreement errors

Errors in subject-verb number agreement are relatively common in everyday language production, occurring approximately once in every 6000 words in English (estimated from Strang, 1966). The typical laboratory paradigm targeting subject-verb agreement errors uses a *preamble completion* task (e.g. Bock & Miller, 1991) in which a participant hears or reads a sentence fragment containing inflected nouns (*The key to the cabinets*) and uses it to form a complete sentence (e.g. *The key to the cabinets is shiny*). While much of the literature reviewed here focuses on English, this task reliably elicits agreement errors in a wide assortment of languages, including Dutch (Anton-Mendez & Hartsuiker, 2010; Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Hartsuiker, Schriefers, Bock, & Kikstra, 2003), German (Hartsuiker, Schriefers, Bock, & Kikstra, 2003); French (Franck, Vigliocco, & Nicol, 2002), Hebrew (Deutsch & Dank, 2009, 2011), Italian (Franck, Lassi, Frauenfelder, & Rizzi, 2006; Vigliocco, Butterworth, & Semenza, 1995), Portuguese (Acuña-Fariña, 2018), Russian (Lorimor, Bock, Zalkind, Sheyman & Beard, 2008), Serbian (Mircovic & MacDonald, 2013) and Spanish (Acuña-Fariña, 2018; Bock, Carreiras, & Meseguer, 2012; Foote & Bock, 2012; Vigliocco, Butterworth, & Garrett, 1996), and analogous findings appear in agreement comprehension (see e.g. Lago, Shalom, Sigman, Lau & Phillips, 2015; Tanner, Nicol, & Brehm, 2014; Wagers, Lau and Phillips, 2009). Agreement production errors occur in a wide assortment of structural and semantic configurations with different lexical factors in play, such as co-occurrences between nouns and predicates, word regularity, and word frequency. Agreement production errors also occur on other parts of speech (e.g., noun-pronoun agreement; Bock, Eberhard, & Cutting, 2004), and for other morphosyntactic features such as grammatical gender (Badecker &

Kuminiak, 2007; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Siloussar & Malko, 2016; Vigliocco & Franck, 2001). Because agreement is so pervasive across and within languages, and because it is prone to error, it is an important aspect of sentence production to explore experimentally and computationally.

Existing experimental work on English subject-verb agreement production has focused on three core phenomena. The first is errors in selecting the correct verb form for production where the verb agrees with a linearly intervening *local* noun instead of the *head* of the phrase which is the typical[1] controller of agreement (e.g. *The key to the cabinets *were shiny*). The increase in agreement errors for head-mismatching local nouns compared to head-matching ones (e.g., versus *The key to the cabinet *were shiny*) is called *attraction* in the literature (e.g. Bock & Miller, 1991). The verb error rate in American English is highest for preambles with a singular head and plural local noun (henceforth $N_s\,N_p$: *The key to the cabinets... *were*), compared to preambles with a plural head and singular local noun ($N_p\,N_s$: *The keys to the cabinet... *was*). This *mismatch asymmetry* is often attributed to grammatical markedness: in English, plural nouns receive an inflection, making their number grammatically 'marked' relative to the singular default (e.g. Eberhard, Cutting, & Bock, 2005). These patterns are summarized in Table 1: when examining error rates out of what in the literature are called 'valid' trials (trials with correct preamble repetitions and an inflected verb completion), the attraction effect in American English is estimated to be about 12% and the mismatch asymmetry effect is estimated to be about 10%. Importantly, attraction also generalizes to so-called *non-intervening* or *remote* configurations,

[1] Following the literature, 'head noun' refers to the first noun of the sentence, and the noun that nearly always controls agreement — except in a pseudopartitive parse, when the agreement controller is the second ('local') noun.

where the attractor is the first noun in the sentence (*The cabinets that the key *were used to open*) and the preamble contains an embedded relative clause (see e.g. Bock & Miller, 1991; Franck & Vigliocco, 2002; Franck et al, 2006; Staub, 2009, 2010; and Santesteban, Pickering & Branigan, 2013). Though error rates are lower in this configuration, in English a similar mismatch asymmetry appears with plural attractors stronger than singular ones.

The second core phenomenon relates to the conceptual number of the phrase. Phrases that refer to referents that are conceptually (often called *notionally*) plural (e.g. *The picture on the postcards; The gang on the motorcycles; The apple with the fresh peach*) elicit reliably more plural verb completions than those referring to referents that are conceptually singular (e.g. *The key to the cabinets*) regardless of the grammatical number marked on the nouns in the phrase. This is termed *notional agreement* (see e.g. Brehm & Bock, 2013, 2017; Eberhard, 1999, Humphreys & Bock, 2005; Vigliocco et al., 1995). Notional agreement is extremely common, especially for collective noun heads like *gang* or *staff*. As shown in Table 1, collective-headed phrases with plural local nouns (e.g. *The gang on the motorcycles*) elicit plural verb completions at rates of 60% of all 'valid' trials, despite the fact that they are prescriptively singular in American English (i.e. are 'supposed' to take singular agreement).

On top of attraction and notional agreement, a third phenomenon has been examined in experiments on subject-verb agreement production: the role of experience. Recent exposure to specific sentence types changes rates of notional agreement and attraction (Haskell, Thornton, & MacDonald, 2010). It also changes rates of plural agreement with conjoined noun phrases, which are inherently flexible in their verb agreement because the number of the phrase sometimes conflicts with notional number (e.g. *The name and address is/are*; Lorimor, Adams, & Middleton, 2018). Word frequency, an important factor in long-term language experience, also impacts attraction such that high frequency local nouns do not cause agreement attraction in comprehension (correct and erroneous verbs are read equally fast no matter the local noun number; Brehm, Hussey, & Christianson, 2019). Base rates of structure frequencies also change what structure is comprehended, such that infrequent agreement patterns are sometimes initially misread (Keshev & Meltzer-Asscher, 2021). Because of the observed effects of priming and structure frequency in agreement, MacDonald and colleagues (Haskell et al., 2010; see also Thornton & MacDonald, 2003) propose that past experience with grammatical sentences exhibiting notional agreement or attraction errors following local plurals is what drives the mismatch asymmetry effect in American English. The authors show that non-canonical plural agreement is high for $N_s N_p$ phrases — comprising 21% of all completions for all $N_s N_p$ items in the Brown corpus — but low for all other phrases. Indeed, many of these items are *pseudopartitive:* the second noun can control agreement of the phrase. Their claim is that exposure to notional agreement, pseudopartitives, and errors enhances the mismatch asymmetry effect, leading to high rates of plural verb production errors for $N_s N_p$ items.

1.2 Modeling subject-verb agreement

Existing models of agreement have mainly focused on the mechanisms by which attraction, the mismatch asymmetry, and notional agreement occur. One class of model explains agreement production via a feature-based mechanism that appeals to syntactic and semantic properties of sentence preambles. The earliest and still most-commonly cited model of agreement production is Marking and Morphing ('M&M', Eberhard et al., 2005). M&M generates verb or pronoun completions to phrases based upon spreading activation within a noun phrase; while it was designed to account for both subject-verb and pronoun agreement, we focus here only on what it predicts for subject-verb agreement.

Table 1. Aggregated data from spoken preamble repetition studies of agreement attraction in American English using preambles with the structure $NP_1[PP\ P\ NP_2]$, where the head of NP_1 is the head noun and the head of NP_2 is the local noun, tabulated in terms of which response types are aggregated into proportions per row and raw counts. Singular V = correct repetition of preamble plus inflected singular verb; Plural V = correct repetition of preamble plus inflected plural verb; Preamble = erroneous repetition of preamble; Uninfl = correct repetition of preamble plus non-overtly inflected past tense lexical verb (e.g. *seemed*) ; NR = no response. Note that not all papers reported full breakdowns of response types; these studies are omitted from any tabulations including preamble error, uninflected, and NR responses, which leads to a larger number of data points included in 'Valid' trials overall. See Appendix A for counts by experiment.

Head Noun	Local Noun	Source	Singular V		Plural V		Preamble		Uninfl		NR	
			Prop	*Count*	*Prop*	*Count*	*Prop*	*Count*	*Prop*	*Count*	*Prop*	*Count*
Singular count	Plural count	Valid trials	0.86	*15651*	0.14	*2612*						
		Trials with inflected verbs	**0.74**	*14408*	**0.13**	*2433*	**0.13**	*2546*				
		All trials	0.60	*14408*	0.10	*2433*	0.11	*2546*	0.17	*4112*	0.02	*533*
Singular count	Singular count	Valid trials	0.98	*17943*	0.02	*363*						
		Trials with inflected verbs	**0.89**	*16439*	**0.02**	*335*	**0.09**	*1647*				
		All trials	0.71	*16439*	0.01	*335*	0.07	*1647*	0.19	*4379*	0.02	*477*
Plural count	Singular count	Valid trials	0.04	*221*	0.96	*5102*						
		Trials with inflected verbs	**0.04**	*196*	**0.82**	*4249*	**0.14**	*749*				
		All trials	0.03	*196*	0.67	*4249*	0.12	*749*	0.18	*1148*	0	*2*

Plural count	Plural count	Valid trials	0.03	*137*	0.97	*4662*						
		Trials with inflected verbs	**0.03**	*127*	**0.80**	*3653*	**0.17**	*801*				
		All trials	0.02	*127*	0.64	*3653*	0.14	*801*	0.2	*1151*	0	*0*
Collective	Singular count	Valid trials	0.87	*984*	0.13	*151*						
		Trials with inflected verbs	**0.73**	*906*	**0.09**	*107*	**0.18**	*226*				
		All trials	0.47	*906*	0.06	*107*	0.12	*226*	0.35	*681*	0	*0*
Collective	Plural count	Valid trials	0.40	*301*	0.60	*450*						
		Trials with inflected verbs	**0.28**	*244*	**0.41**	*364*	**0.31**	*278*				
		All trials	0.15	*244*	0.22	*364*	0.17	*278*	0.46	*746*	0	*0*

M&M generates preamble completions based upon a valuation of the head that combines grammatical and notional number into a single, continuous index ranging from −1 (specified singulars such as *One key*) to 1 (unambiguous plurals; *The keys*), with singular count nouns (*The key*) receiving a value of 0 and notionally plural phrases (e.g., *The picture on the postcards*) receiving moderately positive values. This valuation is combined with the sum of grammatical number values for all nouns in the phrase, weighted based upon the noun's distance from the root of the tree and the so-called 'contrastiveness' of singular to plural usages for that class of noun, such that nouns that are syntactically closer to the root and nouns that have no singular form (e.g. *pants*) contribute more plural value to the whole phrase. This means that M&M accounts for notional, lexical, and syntactic factors in agreement, and for the relative frequency of singular versus plural forms used with particular lexical items. Verb number is then generated by taking the logistic transform of the combined number value, with an additional bias included to elicit more singular agreement as it is considered the 'unmarked' form in English.

However, later work has shown M&M to fail in a variety of cases. It fails to characterize non-intervening attraction because the attractor is outside of the noun phrase; however, this is not problematic if non-intervening attraction is caused by a different mechanism (see Staub, 2009, 2010). Most crucially, it fails for constructions where both singular and plural forms are available (e.g., for Serbian nouns with quantifiers: Mirković & MacDonald, 2013; for conjoined nouns in English: Keung & Staub, 2018), and when there are more than two options for agreement (e.g. Slovak gender agreement: Badecker & Kuminiak, 2007). It also cannot clearly account for form-based effects on agreement, which include reduced attraction rates for irregular plurals (Haskell & MacDonald, 2003) and reduced attraction rates within paradigms where singular and plural forms have some overlap (Lorimor, Jackson, Spalek, & van Hell, 2016; Mirkovic & MacDonald, 2013). These suggest that the fact that M&M requires an additional bias towards singular forms — a way of accounting for frequency patterns of American English — in fact prevents the model from generalizing across structures and languages.

A more recent model that uses a feature-based mechanism to account for notional and grammatical effects in agreement production uses a self-organizing probabilistic dynamical system framework (SOSP; Smith, Franck, & Tabor, 2018; 2021). This model derives notional number from a set of syntactic and semantic features such that more plural agreement occurs for collective heads and other grammatically *pseudopartitive* phrases (e.g. *A lot of postcards are*) because in constructions like this, the second noun can be probabilistically used as the grammatical head of the phrase. In this model, sentences are built out of self-assembling *treelets* (atomic tree stuctures consisting of a parent and one or more child nodes, corresponding to rules in a context free grammar). The probabilistic activation of treelets causes a graded pattern of plural verb completions based upon the features of the lexical item in the phrase. This model uses a single, simple mechanism that synthesizes both notional and grammatical factors in agreement, as conceptually suggested by MacDonald and colleagues, doing away with the multi-stage framework in M&M that separates grammatical and notional number. SOSP (Smith et al., 2021) also covers encoding interference effects, where semantic similarity at item encoding leads to later errors, better than any other model (see Barker, Nicol, & Garrett, 2001; Smith et al., 2021; Villata, Tabor, & Franck, 2018).

A second class of models of agreement, those using ACT-R, focus primarily on the role of lexical factors and memory retrieval in attraction. These models use a domain-general framework that ascribes errors or processing difficulty to memory retrieval; these models were first applied to agreement comprehension but have since been adapted for production. Models using the ACT-R implementation state that attraction arises from noun mis-retrieval in a content-addressable (or cue-based) memory framework (e.g. Lewis & Vasishth, 2005; see Jäger, Engelmann, & Vasishth, 2017, for meta-analysis, but see Hammerly, Staub, and Dillon, 2019, for counter-evidence in agreement comprehension). Existing data support the role for content-addressable memory dynamics in production, such that attraction reflects when the local noun is mis-retrieved as the agreement controller instead of the head. Patterns consistent with mis-retrieval in production include the increased attraction rate for case-ambiguous head nouns and the graded attraction pattern across the gender paradigm hierarchy in Slovak (Feminine > Masculine > Neuter), where attraction occurs in proportion to how much more marked the local noun is than the head (Badecker & Kuminiak, 2007), the increased attraction rate when gender and number cues both mismatch in Spanish (Lorimor, Jackson, & Foote, 2015), and the increased rates of singular agreement for conjoined noun phrases matching in determiner gender in Dutch and German (Lorimor et al., 2016).

The primary error mechanism of ACT-R is a domain-general aspect of memory: mis-retrieving the number of the verb controller. This reliance on memory dynamics in eliciting agreement errors also has broad ecological validity. There is a known link between working memory capacity and success in agreement production for special populations (bilingual children: Veenstra, Antoniou, Katsos, & Kissine, 2018; older adults: Fyndanis, Arcara, Christidou, & Caplan, 2018; aphasic patients: Fyndanis et al., 2018; Slevc & Martin, 2016), and in some populations of healthy young adults (e.g. Hartsuiker & Barkhuysen, 2006, though cf. Bock & Cutting, 1992). This underscores the importance of retrieving nouns from memory while performing subject-verb agreement. It also highlights that errors might occur because of the need to manipulate multiple simultaneously active lexical elements while planning to prepare speech, suggesting a plausible role for parallel planning in eliciting morphosyntactic errors.

1.3 Memory errors in agreement as a key phenomenon

Importantly, the existing literature may have under-estimated the role of memory in agreement production. This is because memory failures might explain an often-overlooked aspect of the observed data from production error elicitation experiments. As shown in Table 1, the preamble completion paradigm elicits many more types of responses than the ‘valid’ trials reported in the literature (correct preamble repetitions followed by overtly inflected plural or singular verbs). Here, we focus on the production of what we term *preamble errors,* where the speaker has mis-produced the sentence preamble they were prompted with; these are also often called *miscellaneous* errors in the literature. Many types of preamble errors occur. Instead of producing *The key to the cabinets,* speakers might change the inflection on the local noun (*The key to the *cabinet_*) or the head (*The *keys to the cabinets*), might alter lexical items while preserving the original inflections (*The key to the *locks*), or might change the preamble’s structure entirely (e.g. *The key *and the cabinets*).

Taking the mean of the count-noun trials in Table 1 shows that preamble errors reflect 11% of all trials, while agreement errors reflect only 4% of all trials. This means that the single most common error in subject-verb agreement paradigms is not in selecting a verb for production, but in recalling and repeating the words correctly from the prompt. Critically, preamble errors also co-vary with agreement, with more preamble errors typically appearing when either the head noun or the local noun is plural; this fact has gone largely unnoticed in the literature (though cf. Bergen & Gibson, 2012, and Thornton & MacDonald, 2003).

One important subtype of preamble error is those in which the lexical items and structure are repeated veridically, but inflections on the head or local noun are changed. We term these *head* and *local* errors respectively.

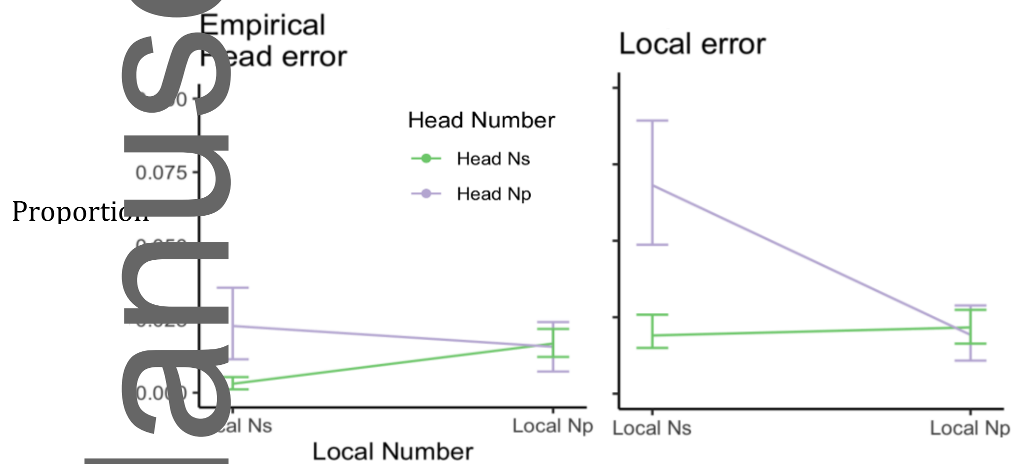

Figure 1. Empirical preamble error probabilities tabulated over preambles' original head and local number. Error bars show 95% confidence intervals (CI)s generated by 5000 runs of a non-parametric bootstrap, sampling with replacement. Labels refer to the target production, not the error production.

Table 2. Preamble error breakdown for all preamble completion studies for which data were available. Proportions are taken out of total trial numbers (N_sN_p = 2627; N_sN_s = 2627; N_pN_s = 616; N_pN_p = 852); counts follow in parentheses. Head and local noun types refer to the target production.

Head Noun	Local Noun	Preamble error total	Head num error		Local num error		Other
		Prop Count	Form	Prop Count	Form	Prop Count	Prop Count
Singular count	Plural count	0.144 (379)	$N_sN_p \rightarrow N_pN_p$	0.017 (44)	$N_sN_p \rightarrow N_sN_s$	0.022 (57)	0.106 (278)
Singular count	Singular count	0.102 (267)	$N_sN_s \rightarrow N_pN_s$	0.003 (8)	$N_sN_s \rightarrow N_sN_p$	0.019 (50)	0.080 (209)
Plural count	Singular count	0.157 (97)	$N_pN_s \rightarrow N_sN_s$	0.023 (14)	$N_pN_s \rightarrow N_pN_p$	0.068 (42)	0.067 (41)
Plural count	Plural count	0.089 (74)	$N_pN_p \rightarrow N_sN_p$	0.016 (13)	$N_pN_p \rightarrow N_pN_s$	0.019 (16)	0.054 (45)

Figure 1 and Table 2 show the preamble error rates reported in Thornton and MacDonald (2003) combined with two data sets made available to the authors (Humphreys & Bock, 2005 and Brehm & Bock, 2013) to estimate how often the inflections are changed. (This represents all of the preamble errors which were available for re-coding from any of the studies reported in Appendix A.) Though data are sparse—particularly for plural-headed phrases— three patterns emerge. First, local errors (right panel of Figure 1) are more common than head errors (left panel). Second, plural heads (purple lines) are broadly associated with more errors (becoming singular) for both head and local errors. Finally, head and local number interact for both error types, with the consequence that mismatches are often eliminated between head and local number (so that *The key to the cabinets* becomes *The key to the cabinet* or *The keys to the cabinets*). Plural-headed preambles show more head and local errors for singular than plural local nouns, whereas singular-headed preambles tend to show the opposite pattern. This interaction is particularly strong for local errors, with notably high error rates for $N_p N_s$ items, so that *The keys* *to the cabinet* frequently becomes *The keys to the cabinets*. Because of the high rates of preamble errors in preamble completion tasks and their relationship to grammatical number, we test how a model fitted to verb errors generalizes to preamble errors, rather than discarding preamble errors as failures to perform the task. This serves as a test of the model.

1.4 PIPS: A parallel planning model of sentence production

We present here PIPS (Parallelism in Planning Syntax): a model of structure generation in sentence production. In PIPS, agreement errors and preamble errors both reflect a broader phenomenon within sentence production: the presence of transient blend states during planning. While planning to speak, the temporary activation of components of alternative structures and alternative lexical items allows elements of other sentences to compete for production, producing errors. Our goal is to describe production errors as a consequence of these transient blend states in a model that produces sentences based on the structural frequencies of American English. This will serve to capture effects due to competition from structural and lexical elements within a system that is based on the constraints of the language, is fallible in memory, and has to plan over time. As such, our model incorporates many of the key strengths of SOSP and ACT-R models of agreement while providing a fuller description of the dynamical computation of sentence structure during production.

We adopt the Gradient Symbolic Computation framework (GSC; Cho et al., 2018; Cho et al., 2017, 2020; Smolensky et al., 2014). In GSC, sentence production is represented as a continuous-time, continuous-state stochastic dynamical system (similar to the underlying dynamics of SOSP, Smith et al., 2018). Symbolic constituents are represented by vectors (e.g.,

N_s corresponds to a list of numbers denoted here by $\mathtt{N_s}$). Because all constituents are represented by vectors, we can directly encode similarity between representations in a graded manner. For example, rather than treating $\mathtt{N_s}$, $\mathtt{N_p}$, $\mathtt{V_s}$, $\mathtt{V_p}$ as separate, discrete symbols, we can encode the greater similarity of elements of the same type (such that $\mathtt{N_s}$ is more similar to $\mathtt{N_p}$ than to $\mathtt{V_p}$). This allows us to manipulate the lexical and structural similarity between elements of a sentence in order to test how varying representations affects errors.

In the implementation of PIPS used here, there are only two types of words: nouns and verbs, each inflected for number as singular or plural. The model has, however, a rich representation of elements at the sentence level: sentences are combinations of nouns and verbs in a variety of structures such as NP, RC and S(entence). Sentence representations are structured combinations of syntactic constituents, constructed using tensor (outer)[2] products of vectors representing fillers (lexical items and structural constituent categories) and vectors representing roles (positions within hierarchical sentence structures). This allows the model to represent the tree structure of a sentence and its lexical items as a single vector: the sum of the tensor products binding roles with fillers (Smolensky, 1990, 2006). For example, consider the treelet [S_s N_s V_s], representing a simplified tree structure for a sentence comprised of a singular noun followed by an agreeing intransitive singular verb; S_s represents the parent node and N_s and V_s are the child nodes in order from left to right [3]. Each element of the treelet can be represented as the tensor

[2] Also called *outer product*; at the most basic level, the tensor product of two vectors a and b is a matrix in which the *i*th, j^{th} element is the product of the i^{th} element of vector a and the j^{th} element of vector b (see Smolensky, 2006, for more detailed discussion).

[3] Because of the high-dimensional problem space required to represent fillers and roles separately, we

product ($\otimes$) of a filler vector ($\texttt{N}_\texttt{s}$) and a role vector ($\texttt{left-child}$; $\texttt{N}_\texttt{s} \otimes \texttt{left-child}$). The entire treelet is simply the sum of all the tensor products ($\texttt{Sentence} \otimes \texttt{root} + \texttt{N}_\texttt{s} \otimes \texttt{left-child} + \texttt{V}_\texttt{s} \otimes \texttt{right-child}$). Note that representing linguistic structure as a tensor product of fillers and roles also allows the model to generalize fillers to roles they did not appear in within the model's training data, including ungrammatical sentence types with verb errors (e.g., [$S_s\, N_s\, V_p$]: $\texttt{Sentence} \otimes \texttt{root} + \texttt{N}_\texttt{s} \otimes \texttt{left-child} + \texttt{V}_\texttt{p} \otimes \texttt{right-child}$). This means that PIPS represents errors when well-supported by other constraints in the input or grammar.

These discrete combinations of tensor products representing discrete symbolic structures identify a subset of states within the continuous representational space of the model. The remainder of the space consists of *conjunctive blend* states where multiple symbols are simultaneously partially active, e.g., a state where both a singular and plural noun occupy the first position of the sentence to varying degrees $(0.25 \cdot \texttt{N}_\texttt{s} + 0.6 \cdot \texttt{N}_\texttt{p}) \otimes \texttt{left-child}$, or a state where both a singular and a plural verb occupy the final position of the sentence to varying degrees: $(0.25 \cdot \texttt{V}_\texttt{p} + 0.6 \cdot \texttt{V}_\texttt{s}) \otimes \texttt{right-child}$). Activations in the model can be any real number, but two constraints restrict activations (one prevents extreme activations and another pushes symbol activations towards 0 or 1). See Cho et al. (2018, 2020) for more details.

While output blends are rare, they have been consistently found in real production, motivating us to explore the consequences of a fully gradient representational space. An example of a *full* conjunctive blend of sentences in real production would involve simultaneously

---

simplified the number of nodes in the structure, making it as compact as possible while representing the essential constituent structure of the phrase.

producing two different sentences by speaking one and signing the other; while unusual, this does occur (e.g. Emmorey, Borinstein, Thompson & Gollan, 2008). In the model, this would correspond to all units representing both sentences being activated at values close to 1.

A *partial* conjunctive blend would involve combining elements from multiple formulations, so that some of the units from multiple sentences are activated at values close to 1. Examples of this include:

- Combining multiple elements of idiomatic expressions (e.g., *kick the maker,* from *kick the bucket* and *meet your maker*; Cutting & Bock, 1997).
- Combining multiple elements of distinct formulations conveying similar meanings (e.g., *Many students often turn in their assignments late*, where *many* and *often* are two ways to express a similar meaning; Coppock, 2010; Frazier & Clifton, 2014).
- 'Doubling' in code mixing, where an element appears in two languages (e.g., *They gave me a research grant koḍutaa*, where the final word is the Tamil translation equivalent of 'gave'; Sankoff, Poplack, & Vannianiarajan, 1990; see Goldrick, Putnam, & Schwarz, 2016, for review and discussion).

GSC claims that all conjunctive blends are surface manifestations of intermediate states arising ubiquitously within cognitive processes. In sentence production, conjunctive blends represent multiple choices of formulations – for example, what the grammatical number of the verb or the subject noun should be. Since PIPS only has one token of each word class, many of its blend states are analogous to being undecided whether to use a singular or plural noun (e.g. *pasta*$_s$ or *noodles*$_p$) or being undecided as to which verb inflection is appropriate (e.g. *The police* *is/are*). However, conjunctive blend states cover the full range of possible formulations, and can include parts of ungrammatical sentences.

At the outset, the model is undecided as to which out of all possible combinations of words and structures should be produced – existing in a conjunctive blend of all possible formulations. Production is then a winnowing of these possibilities. Modulo the blend errors noted above, processing typically moves from considering multiple possible utterance formulations to selecting a single well-formed utterance for production. The selection of one plan instead of a blend of multiple plans is driven by a dynamic control parameter, *commitment strength*, which increases over the course of processing, pushing the model towards states that correspond to discrete sentence structures. Commitment strength is the magnitude of a constraint that forces the system to commit to the production of a single discrete structure.

Changes in commitment strength are unique to GSC models and might be analogized as a helium balloon (the current state of the model) beginning the computation by floating at the center of a tent's ceiling. The space on the fabric between tent poles is the model's representational space, with discrete structures mapping to the top of poles and conjunctive blends of utterances mapping to points in middle. Increasing commitment strength changes the goodness of symbolic states by altering their *harmony* (here, the height of the tent at each point). States with higher harmony values are preferred by the model; the processing system's state movement within the continuous representational space is continuously updated so as to maximize this value (as SOSP does in a continuous but non-changing space; Smith et al., 2018). The increase in commitment strength is like lifting the tent poles while anchoring the center to create distinct peaks, so that the balloon tends to settle in one symbolic state, floating to the top of one peak, despite the inherently gradient representational space. Random fluctuations in traversing the space causes the model to sometimes choose a peak that is not the highest one, making errors proportional to their harmony. See Cho et al, 2017 for more details.

The selection of a particular structure is guided by the model's experience. During training, the model encounters some structures more than others: for example, sentences corresponding to different 'preamble' inputs, or singular versus plural verbs. The relative probability of different structures forms a second key part of the model's dynamics; it is more likely to commit to higher probability sentences, which have higher harmony values, than low probability sentences. This allows us to account for the influence of structure frequencies and asymmetries in the distribution of fillers (e.g., plural verbs versus singular verbs) and to explicitly test the role of grammar frequencies in production. In addition to overall frequency differences, an error that obeys local dependencies (while violating non-local dependencies; e.g., N_s N_p V_p) has higher harmony and therefore is more likely than an error that does not (e.g. N_s N_s V_p), following earlier work (see e.g., Goldrick & Daland, 2009; Smolensky et al., 2014). Continuing the analogy, the higher harmony of some states causes the initial starting state of the model to be raised towards one side; the balloon will settle in those spaces preferentially. This provides a mechanism for the higher probability of some errors, such as the increase in agreement errors when the head and local noun mismatch.

In addition to context-free probability, the model's selection of a specific structure is guided by the task. In the experimental preamble completion task, we assume that participants first encode the preamble, then attempt to retrieve the correct elements from their memory. The model implements memory encoding by partially pre-activating fillers from the model's grammar — both words (e.g., N_s, N_p) and structural constituents (e.g., NPC_s). The activation of constituents during processing, based upon that input, corresponds to memory retrieval. Noise in the system means that the model will make errors in encoding and retrieval, like people do. Structural elements are activated in addition to lexical elements following the assumption that in the preamble completion task, speakers need to parse the input and then produce a sentence with the input structure. Partial activation of structures corresponds to the intuition that this process can be faulty.

A final simplifying assumption that we take in PIPS is that production happens after all utterance planning is complete. While individuals have a fundamentally flexible planning scope (e.g. Konopka & Meyer, 2014), there is existing evidence that the planning scope of simple sentences often includes a verb (e.g. Kuchinsky & Bock, 2010; Momma & Ferreira, 2019). PIPS uses a fairly wide planning scope of two nouns and up to two verbs. We took this assumption as a starting point: since we do not have data on the typical planning scope for sentences of the elicited type; incorporating variations in incrementality would require an extra free parameter.

Combined, these properties allow us to build a model that characterizes preamble and agreement errors as arising from the same dynamical mechanisms: language production processes that are sensitive to both globally- and locally-driven agreement. The model's dynamics are sensitive to the distribution of structures in its experience — here, based on American English — accounting for the role of past experience in current production, and production is supported by an inherently noisy memory encoding and retrieval process.

The approach we take in the rest of this paper is to qualitatively fit a model in the GSC framework to the two critical verb error patterns displayed in human data for canonical attraction constructions: *attraction* (more errors for $N_s N_p$ than $N_s N_s$ inputs) and the *mismatch asymmetry* (more errors for $N_s N_p$ than $N_p N_s$ inputs). We fit these data by fitting noun and structural constituent similarity as well as the strength of structural encoding of an input preamble. We then investigate how the same model dynamics gives rise to errors of other types. First, if verb errors in canonical constructions and non-intervening constructions are generated by the same dynamics, we should be able to capture a mismatch asymmetry in both cases. Second, if verb and preamble errors are generated by the same model dynamics, then we should also be able to model preamble errors with the same model settings. We end with an exploration of the model space in order to demonstrate how each parameter contributes to model behavior.

2. Method

Code and results from all simulations reported here are archived on the Open Science Framework: https://osf.io/3udb8/.

2.1 Structure of training grammar

A GSC model was trained to implement a probabilistic context-free grammar (PCFG) G (1) that generates sentences of up to 4 words in length (omitting function words). The key sentences in the grammar were complex noun phrases (NPC): preverbal noun phrases with an inflected count noun head and an inflected 'local' prepositional phrase complement (e.g. *The key(s) to the cabinet(s)*; the inflection on each noun is reflected by the subscript s(ingular) or p(lural). Noun phrases with multiple expansions were represented with multiple distinct non-terminal (i.e., phrasal) representations, following Cho et al. (2018). This means that, for example, the sentence preamble N_s N_p was represented separately from the preamble N_p N_s at non-terminal nodes as well as at the terminals (words).

All probabilities within the grammar follow American English (see Appendix B for full calculations). Singular nouns were twice as probable as plural nouns, following English biases derived from a search of inflected nouns in the COCA corpus (Davies, 2008). Probabilities for subject-verb agreement come from an average over studies of American English with items in an $N_{s/p}$ $N_{s/p}$ form (as shown in Table 1, see Appendix A for full data).

To account for notional agreement patterns, we also allowed a plural headed noun phrase (NP_p) to expand to the terminals N_s N_p, reflecting a version of the right-headed analysis of notional agreement adopted by Smith et al. (2018) where the syntactic head of the phrase was the plural second noun and the noun phrase itself was marked as plural. The probability of the pseudopartitive was based on the reported rates of plural agreement for this type of noun phrase in American English (Haskell et al., 2010; see Appendix B for full calculations).

We acknowledge that pseudopartives are not available in English for all N_s N_p items. However, the parse is not strongly lexically restricted: for example, many nouns can be part of distributive referent (e.g. *The label on the bottles*) and take either singular or plural agreement. Our analysis – which relied solely on a non-lexicalized PCFG, excluding semantics – assumed that all nouns used in these preamble experiments have some degree of access to this parse. This is a clear oversimplification. We leave lexical-specificity and semantic contributions to processing of this construction to future work.

In addition to the complex noun phrase structure containing a PP, which represents the canonical 'agreement attraction' sentence preamble, the model grammar also included one NPC parse in which a verb agreed with a local noun, in order to represent the fact that in English, sometimes the second noun is the subject of, and so agrees with, the adjacent verb. This parse is a main-clause subject noun phrase containing an embedded reduced object-relative clause N_i V_i (e.g. *The key [the cabinets use] broke*), with probability based on a corpus analysis by Roland, Dick, and Elman (2007). Including this parse also allowed us to test how a model fitted to intervening attraction captured non-intervening attraction, which served as a test of how well PIPS generalizes.

(1) A probabilistic context-free grammar G yielding 3 sentence types: [N_i V_i]; [[N_i N_j] V_i]; [N_i [N_j V_j] V_i]. Subscripts denote the grammatical number of the associated symbol; pipes separate different expansions of the same symbol ranked from higher to lower probability. Note that there is no VP in this grammar, to simplify the problem space.

$S \rightarrow 0.44\ N_s\ V_s \mid 0.22\ N_p\ V_p \mid 0.208\ NPC_s\ V_s \mid 0.122\ NPC_p\ V_p$

$NPC_s \rightarrow 0.54\ N_s\ N_s \mid 0.27\ N_s\ N_p \mid 0.18\ N_s\ RC$

$NPC_p \rightarrow 0.47\ N_p\ N_s \mid 0.24\ N_p\ N_p \mid 0.18\ N_p\ RC \mid 0.10\ N_s\ N_p$

$RC \rightarrow 0.66\ N_s\ V_s \mid 0.33\ N_p\ V_p$

Based on the grammar G, 11 sentences of length 2 through 4 can be generated. Their target probabilities (p_t), observed probabilities in the trained model (p_o), terminals, and phrase structures are given in (2). Subscripts denote grammatical number. Sentences 2 through 5 are the target sentences for the main simulations, sentences 6, 7, 9 and 10 contain an embedded reduced object relative and sentence 8 is the pseudopartitive.

(2) Sentence 0: $p_t = 0.4344$, $p_o = 0.4042$ ([S_s N_s V_s])

Sentence 1: $p_t = 0.2270$, $p_o = 0.2154$ ([S_p N_p V_p])

Sentence 2: $p_t = 0.1192$, $p_o = 0.1138$ ([S_s [NPC_s N_s N_s] V_s])

Sentence 3: $p_t = 0.0572$, $p_o = 0.0487$ ([S_s [NPC_s N_s N_p] V_s])

Sentence 4: $p_t = 0.0570$, $p_o = 0.0425$ ([S_p [NPC_p N_p N_s] V_p])

Sentence 5: $p_t = 0.0304$, $p_o = 0.0240$ ([S_p [NPC_p N_p N_p] V_p])

Sentence 6: $p_t = 0.0286$, $p_o = 0.0274$ ([S_s [NPC_s N_s [RC_s N_s V_s]] V_s])

Sentence 7: $p_t = 0.0148$, $p_o = 0.0097$ ([S_p [NPC_p N_p [RC_s N_s V_s]] V_p])

Sentence 8: $p_t = 0.0126$, $p_o = 0.0269$ ([S_p [NPC_p N_s N_p] V_p])

Sentence 9: $p_t = 0.0112$, $p_o = 0.0081$ ([S_s [NPC_s N_s [RC_p N_p V_p]] V_s])

Sentence 10: $p_t = 0.0076$, $p_o = 0.0087$ ([S_p [NPC_p N_p [RC_p N_p V_p]] V_p])

2.2 Model structure and training

Two separate free parameters encoded the representational similarity (i) between fillers representing nouns differing only in number inflection (N_s and N_p; *'Noun Terminal Similarity'*) and (ii) between pairs of fillers representing different expansions of a structural element with the same number inflection (e.g., all of the symbols representing the NPC_s expansions in (1); *'Structural Constituent Similarity'*). These parameters reflect variations in lexical and structural similarity on a scale from 0.0 (fully orthogonal), where the two elements were as distinct as any randomly chosen pair of items, to 1.0 (identical), where there was no distinction between the two elements. Both parameters were varied across simulations so that the dot product of each relevant pair of vectors ranged from 0.2 (mostly orthogonal) and 0.7 (highly similar). The dot product of all other pairs — including the two terminal symbols representing verbs inflected as singular or plural — was set to 0 (perfectly orthogonal), following the simplifying assumption that what matters most for agreement is properties of the preamble, not the verb.

From these constraints, 29 filler vectors were randomly chosen as distributed vector encodings of the terminal and non-terminal symbols. Ten orthonormal role vectors were randomly chosen as vector encodings of the structural position of the symbols. As in Smolensky (1990), these filler and role vectors were composed by the outer product to generate 290 binding vectors, e.g. `Sentence ⊗ root`. The role vectors were structured using the "brick role" system proposed by Cho et al. (2020), with roles encoded based upon whether the position was the right or left child of its parent node, its syntactic depth in the tree, and its position from left to right. We selected to use these roles because they are useful for constraining interactions between different structural positions, allowing us to scale to a model of this size and complexity—see Cho et al. (2020) for details.

As discussed above, the network dynamics were structured so as to favor the selection of grammatical over ungrammatical structures, and to favor higher- over lower-probability grammatical structures, approximating the empirically-observed probability distribution across structures when run as a language generator. Two key mechanisms supported these dynamics: commitment strength (also referred to as *quantization* or *discreteness*; see Cho et al., 2018, 2020 for more details) and spreading activation. Commitment strength is unique to GSC. This parameter is what allows the model to allow flexibility at the outset of planning but end with full commitment to a single parse: it starts weak and grows in strength over the course of computation. Spreading activation was similar to other connectionist networks[4]. Activation spreading along weighted connections linking representational units connected in treelets served to push the system towards particular states. In PIPS, these weights pushed the system towards grammatical states based on their relative probabilities in training (Cho et al., 2018, 2020) so that higher frequency structures became more likely outputs. More details about the weight setting procedure follow below.

Sentence generation was modeled by initializing the system to a random point near its equilibrium state at commitment strength 0. Activation spread between representational units, with normally distributed random noise (standard deviation = 0.01) added to each unit's activation. Commitment strength was then increased up to a maximum of 15 (at a rate of 1 unit increase per unit of simulated time), pushing the system to select a discrete structure. This selection was probabilistic due to the random selection of the initial state and random noise in unit activations.

Weights were initialized to parameters used in discrete-state connectionist networks (Hale & Smolensky, 2006). An error-driven training procedure then updated these parameters. In this training, the sentence generation procedure was run 100 times to estimate the model's current probability distribution over discrete sentence structures. The weights were then adjusted so as to increase the probability of selecting frequent structures and to decrease the probability of less frequent or ungrammatical structures (Cho et al., 2020). After 10 epochs of training, the

---

4 The unit activation function was linear. Activation dynamics were constrained by a baseline constraint which pulled activations towards the center of the representational space (Cho et al., 2018, 2020). This is done to prevent activations from moving outside the desired representational space (a jet of air keeping the balloon inside the tent, in our earlier analogy).

model generated grammatical structures on more than 91% of trials; on these trials, the model approximated the grammar's probability distribution over full parse trees (see 2 above describing the grammar).

2.3 The preamble completion task

In the preamble completion task, a participant hears or reads a sentence fragment containing inflected nouns (*The key to the cabinets*) and uses it to form a complete sentence (e.g., *The key to the cabinets is shiny*). Performing this task requires participants to perceive the preamble, encode the words (*key*, *cabinets*) and an associated structural analysis (e.g., [NPC_s N_s N_p]) in memory, and then use this memory representation to drive production of a full sentence.

After training on context-free production, the model then was given inputs to drive production. As in the human version of the task, the model was presented with four preamble types varying in head and local number (N_s N_p, N_s N_s, N_p N_s, and N_p N_p), which correspond to elements that need to be encoded and then retrieved from memory. Inputs served to partially activate representational states corresponding to words (e.g., N_s) and structural constituents (e.g., NP_s); memory errors were deviations from the input. In the model, the N_s N_p input was always presented as a singular NP and paired with the appropriate singular higher-order structure. Having these inputs allowed us to characterize productions as errors (case where the output mismatched the input) even when the sentences produced were fully grammatical.

To examine the role of encoding lexical versus structural information for production, we varied the weighting of these inputs between an entirely lexicalist strategy (s0), in which the preamble is encoded as a sequence of terminals with no higher-order structure (e.g., N_s, N_p, V_s with no tree structure) , versus an entirely structuralist strategy (s1), in which the preamble is encoded only by its structural constituents with no activation of the terminals (e.g., [S_s[NPC_s _ _] _], where the blanks represent slots for terminal nodes). We varied the s1-to-s0 ratio as a free parameter across simulations, such that the weights of both always summed to 1; that is, if the s1 weight was 0, the s0 weight was 1. These activities scaled the input to terminals and non-terminal units. During the course of the simulation, the activity then decayed over the 15 time intervals of production (in arbitrary units) at a rate of 0.9.

For each combination of preamble type, noun terminal similarity, structural constituent similarity, and s1 weight, four separate sets of model runs were performed with different random seeds, with 1000 iterations of sentence generation in each set of runs. Model outcomes were coded as follows. *Correct*: both nouns repeated correctly and followed by a head-matching verb in the appropriate structure (e.g., $N_s\ N_p \rightarrow [S_s\ [NPC_s\ N_s\ N_p]\ V_s]$). *Verb error, non-pseudopartitive*: both nouns repeated correctly with a non-head-matching verb, with structure nodes for either sentence number and any NP other than the one in sentence 8 (e.g., $N_s\ N_p \rightarrow [S_{s/p}\ [NPC_{s/p}\ N_s\ N_p]\ V_p]$). *Verb error, pseudopartitive*: both nouns repeated correctly with a non-head-matching verb, with NP and sentence structure nodes from sentence 8 (e.g., $N_s\ N_p \rightarrow [S_p\ [NPC_p\ N_s\ N_p]\ V_p]$). *Head error*: change to the number of the head noun, with verb matching the modified head and the appropriate structure (e.g., $N_s\ N_p \rightarrow [S_p\ [NPC_p\ N_p\ N_p]\ V_p]$). *Local error*: change to the number of the local noun, with head-matching verb and the appropriate structure (e.g., $N_s\ N_p \rightarrow [S_s\ [NPC_s\ N_s\ N_s]\ V_s]$). *Other:* all other responses, including embedded RCs, incomplete responses, and responses containing both a change to head or local noun and a verb error. Note that while pseudopartitive verb errors for the Ns Np sentences are correct with respect to the model grammar, all verb errors are incorrect with respect to the input. Similarly, head and local errors are correct with respect to the grammar (they are grammatical sentences) but errors with respect to the input.

The primary diagnostic for model fit was which parameter settings best approximated the empirical effect of attraction (0.11 more verb errors for $N_s N_p$ than $N_s N_s$) and the mismatch asymmetry (0.09 more verb errors for $N_s N_p$ than $N_p N_s$; see Table 1). We did this by ranking all models by their distance to each empirical target, and then sorting based upon the sum of the two ranks, breaking ties as necessary by the total numerical distance to both metrics[5]. This provided a qualitative measure of model fit accounting for how settings behaved given different input.

3. Results

As outlined in the introduction, we began by fitting a model to verb errors and examining how conjunctive blend states contribute to verb error production. To evaluate the model, we then tested how these settings elicit non-local attraction and preamble errors. This demonstrated whether the same parameters fit multiple types of errors, which is predicted if all error types are consequences of the same model dynamics. We then explored the model space by manipulating one parameter at a time, disclosing, for each aspect of the model, how it affects model outcomes. Note that these patterns are correlational, not causal: establishing causality in complex dynamical systems is not straightforward (see e.g. Chattopadhyay, Manupriya, Sarkar, & Balasubramanian,

5. As a quantitative measure of model fit, we also assessed KL divergence between the distribution of the model's outputs and the empirical targets for all agreement errors, other errors (pooled together) and correct repetitions across preamble inputs. This tended to penalize one preamble type over all others (penalizing different types with differing parameter settings), which made fits to the empirical attraction and mismatch asymmetry effects relatively poor.

2019). However, observing what happens given changes to free parameters provides some support for possible factors contributing to the model's patterns of performance.

3.1 Grid search for best-fitting grammar

We began with a grid search for the parameter settings that best fit the empirical effects of attraction and the mismatch asymmetry. We trained a series of 275 models allowing noun terminal similarity and structural constituent similarity to vary orthogonally between 0.2 and 0.7, and allowing s1 weight to vary between 0 and 1. The impact of these parameters on the ability of the model to fit the empirical attraction and mismatch asymmetry effects is shown in Figure 2, where a perfect fit is represented by the intersection of the curve for each effect with the horizontal zero line. For these simulations, both types of verb errors (pseudopartitive and non-pseudopartitive) were pooled together, as the form of errors observed experimentally does not distinguish between the parses. A variety of parameter settings replicate the attraction and mismatch asymmetry effects separately (i.e., both curves cross the zero point somewhere in each panel). To fit both effects simultaneously — finding settings in which the four inputs generated appropriate ratios of verb errors in relation to each other — we identified the parameter values where both curves cross zero nearest to each other. To do this, we ranked each model for the absolute value of its distance from the target attraction and mismatch asymmetry effects. The distance rankings for the top five models are displayed in Table 3 along with the overall distance from the target correct and error proportions for each sentence.

The resulting model had the following parameter values: structural constituent similarity = 0.5, noun similarity = 0.7, and s1 weight= 0.5. In the model, the attraction effect was 14.6% (=

0.182–0.036) vs. an 11% (= 0.13 – 0.02) target, yielding a **Δ** of –0.036 (= 11% – 14.6%); the mismatch asymmetry effect was 9.0% (=0.182 – 0.092) (vs. a 9% =0.13 – 0.04 target), yielding a **Δ** of 0.000.

Across preamble types, verb error rates in the model and in experimental data were highest in the N_s N_p condition, followed by the N_p N_s condition. The N_s N_s and N_p N_p conditions both elicited low rates of verb errors – though in the model, the N_s N_s condition elicited slightly more errors, and in the experimental data, the N_p N_p condition elicited slightly more errors. Verb error rates were relatively close to the empirical targets laid out in Table 1, though the N_s N_p and N_p N_s error rates were further from the targets than N_s N_s and N_p N_p error rates. Subsetting by verb error type, 99% of the verb errors in the N_s N_p condition contained the pseudopartitive, compared to 0% of verb errors in the other conditions. The model under-produced correct completions in general, and did so especially for the N_s N_p preambles. Combined, these properties give the model a qualitative fit to the human data, not a quantitative fit: the model fits the relative error rates across preamble types, but does not precisely match experimental data.

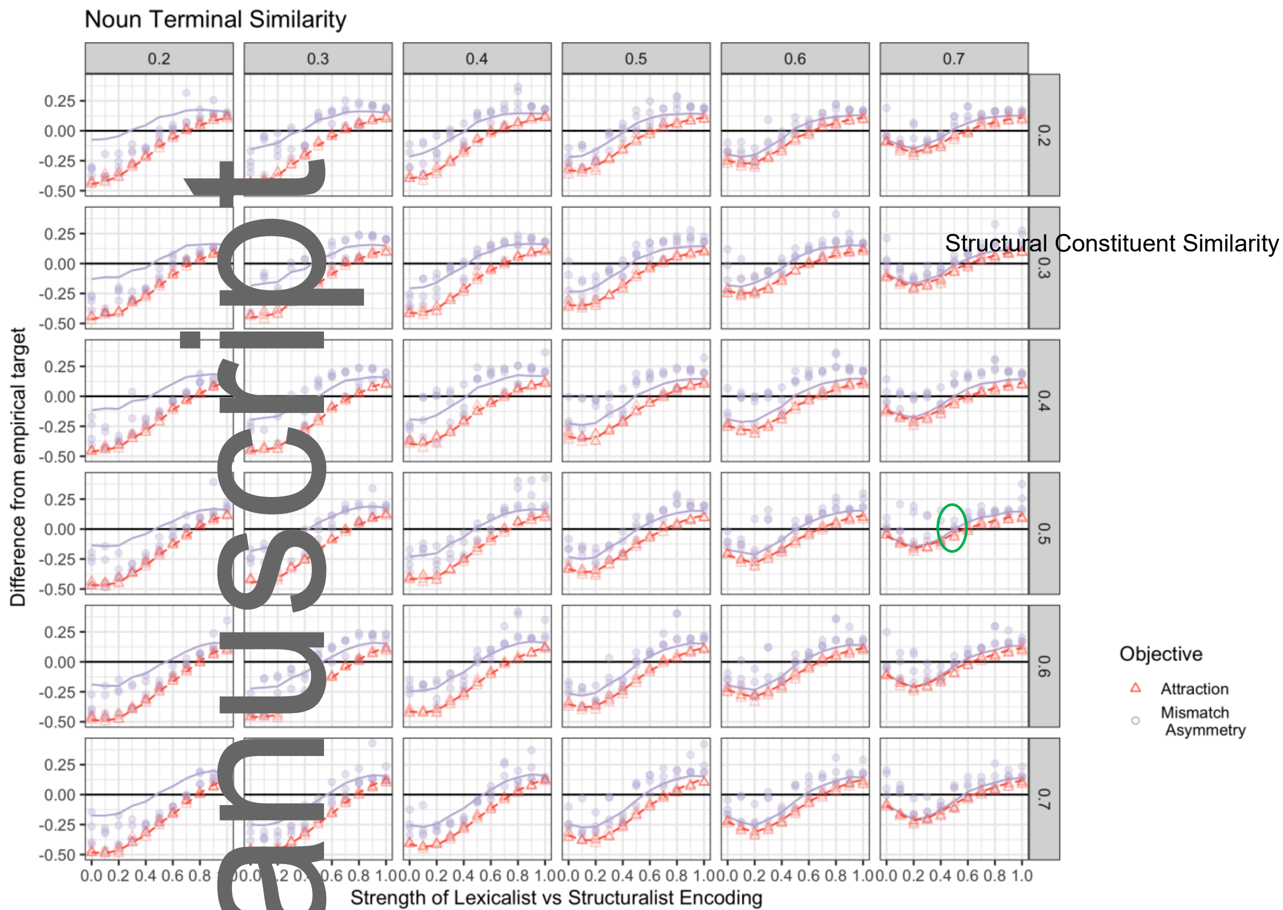

Figure 2. *Difference between model outcome and empirical target for attraction (red dashed line) and mismatch asymmetry effects (solid purple line) for a grid search by structural constituent similarity (similarity of symbols representing different types of singular/plural NP, RC, and S; panels by row), noun terminal similarity (similarity of singular and plural forms; panels by column), and s1 weight (lexicalist vs. structuralist encoding; x axis within each panel). Points reflect results by model run with varied random seeds. Optimum highlighted with green oval.*

Table 3. Model fits for top five simulations in the grid search of structural constituent similarity, noun similarity, and s1 weight, sorted by their combined attraction (Attr) and mismatch asymmetry (MA) ranking. **Δ** represents observed value subtracted from target, pC represents proportion correct completions in model, and pE represents proportion verb errors.

								N_sN_p		N_sN_s		N_pN_s		N_pN_p	
								pC	pE	pC	pE	pC	pE	pC	pE
Structure Sim	Noun Sim	s1 Wt	**Δ** Attr	Rank Attr	**Δ** MA	Rank MA	Sum Rank	*Target 0.74*	*Target 0.13*	*Target 0.89*	*Target 0.02*	*Target 0.82*	*Target 0.04*	*Target 0.80*	*Target 0.03*
0.5	**0.7**	**0.5**	**-0.036**	**45**	**0.000**	**1**	**46**	**0.450**	**0.182**	**0.849**	**0.036**	**0.621**	**0.092**	**0.702**	**0.013**
0.6	0.7	0.6	-0.009	14	0.031	32	46	0.595	0.149	0.905	0.030	0.713	0.090	0.787	0.022
0.7	0.7	0.6	-0.017	25	0.024	22	47	0.546	0.164	0.894	0.037	0.705	0.097	0.744	0.019
0.5	0.7	0.6	-0.005	8	0.038	41	49	0.650	0.141	0.900	0.026	0.749	0.089	0.770	0.021
0.4	0.7	0.5	-0.041	57	0.001	2	59	0.605	0.182	0.898	0.031	0.664	0.093	0.774	0.016

3.2. Transient blends in production

After fitting a model to verb errors, we then evaluated it with a deeper investigation of how constituent activation changed over time within model runs when the model was given N_s N_p input (grammar sentence 3). This allows us to clearly demonstrate the role of transient conjunctive blends in eliciting verb errors, which is a central premise of the model.

Figure 3 contrasts activation on pairs of fillers in the same role depending on whether the best-fitting model generated the correct verb (`[`S_s `[`NPC_s N_s N_p`]` V_s`]`) or made a verb error with a pseudopartitive structure (`[`S_p `[`NPC_p N_s N_p`]` V_p`]`). Panel A examines the activation of fillers dominating the first two terminal nodes `Ns` and `Np`. When a verb error was made, the incorrect pseudopartitive noun phrase ([NPC_p N_s N_p`]`) was activated instead of the correct singular noun phrase with the same terminal node expansion ([NPC_s N_s N_p`]`), and remained active even at the end of the model run. Panel B shows how a transient re-analysis of the local N_p as the head of an RC phrase, compared to its correct identity as a local noun in an NP, was correlated with the production of agreement errors: activation of an analysis of the local noun as an RC head during the midpoint of the computation was associated with errors, even though the activation typically died down by the end of the model run. Panel C shows how transient activation of a relative clause immediately dominating the local noun and following verb (`RCp`) instead of a singular verb (`Vs`) was associated with agreement errors, such that activation of the relative clause at the beginning of the model run decreased activation of the singular verb, allowing the plural verb to be selected instead. These show the crucial role of competition and blend states in how PIPS produces verb errors.

When the model made a non-pseudopartitive verb error, activation on all nodes was similar, aside from the pattern of activation on the pseudopartitive noun phrase itself. This increased at early stages of the computation and then decreased again by the end of the model run. This suggests that competition from the pseudopartitive parse increases attraction, even when the ultimate output does not use the pseudopartitive structure.

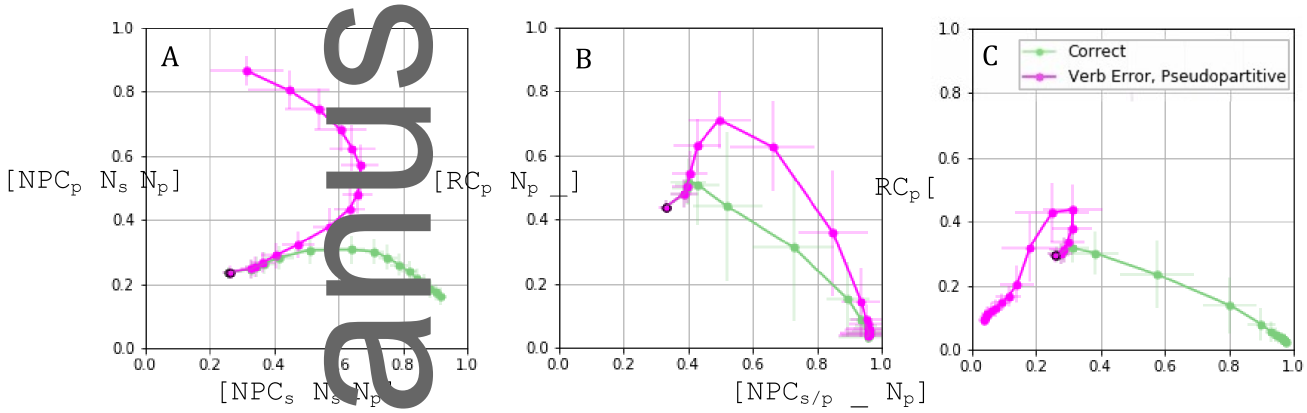

Figure 3. *Mean activation at each simulated time point on pairs of fillers in the same role based upon correct (S_s [NPC_s N_s N_p] V_s]) vs. pseudopartitive error ([S_p [NPC_p N_s N_p] V_p] outcome (error bars show standard error across model runs). At the initial time point (circled in black), fillers have similar, low (≲ .5) activation levels. Panel A: pseudopartitive noun phrase ([NPC_p N_s N_p]) vs. singular noun phrase with plural local noun ([NPC_s N_s N_p]). Panel B: plural noun as head of relative clause ([RC_p N_p _]) vs. as local noun in either a singular or plural noun phrase ([$\mathrm{NPC}_{s/p}$ _ N_p]). Panel C: relative clause containing plural verb (RC_p [ ]) vs. singular verb (V_s). In all panels, the element on the x-axis reflects the correct target and the element on the y-axis reflects a competing non-target element. See text for details.*

3.3 Non-intervening attraction

To see if PIPS can successfully generalize to other empirical data, we next assessed how the trained model completed embedded relative clause inputs. This examines whether PIPS produces non-intervening attraction with a mismatch asymmetry, which would be predicted if both arise as consequences of the same system. This is an important test because unlike intervening attraction, which can be grammatically correct in the N_s N_p case in PIPS because of the pseudopartitive, but which mismatches the input, non-intervening attraction is an error with respect to both the model grammar and the input.

The input to the model here was two nouns, as in the canonical attraction cases (N_s or N_p in the first two terminal positions) and the units comprising the embedded relative clause structure ([$S_{s/p}$ [$NPC_{s/p}$ _ [$RC_{s/p}$ _ _]] _]); all other parameters were as described in the methods section. Responses were coded following the guidelines in Bock and Miller (1991) based upon the terminals only. *Correct* trials reproduced the nouns in the input and used a correctly inflected verb in at least the first position (inside the relative clause). *Verb error* trials reproduced the nouns from the input and used an incorrectly inflected verb in at least the first position. All other responses were coded as *Other errors*. Results appear in Table 4. Similar to the human data (Bock & Miller, 1991; Staub, 2009, 2010), there was a mismatch asymmetry where more verb errors were produced when the first noun was plural and the second noun was singular than vice-versa. While few verb errors were elicited overall, the asymmetry was stable: the same pattern obtained in each of four separate sets of runs of the model with different randomization seeds. Most often, verb errors corresponded to fully grammatical re-parses of the input, matching the complex noun phrase structure (e.g. [S_s [NPC N_s N_p] V_s]); this shows that these errors typically correspond to a structural re-analysis of the input.

Table 4. Proportions of correct completions, verb errors, and all other errors by type of embedded relative clause input (non-intervening attraction) in from PIPS.

Input	Correct	Verb Error	Other Error
$[S_s[NPC_s\ N_s\ [RC_p\ N_p\ \_]]\_]$	0.921	0.002	0.078
$[S_s[NPC_s\ N_s\ [RC_s\ N_s\ \_]]\_]$	0.946	0.000	0.054
$[S_p[NPC_p\ N_p\ [RC_s\ N_s\ \_]]\_]$	0.867	0.004	0.130
$[S_p[NPC_p\ N_p\ [RC_p\ N_p\ \_]]\_]$	0.922	0.000	0.078

Note that base rates of verb errors were lower than earlier work: Bock and Miller (1991) elicited 6% verb errors for non-intervening $N_s\,N_p$ items with the structure $[S_p\ [NPC_p\ N_p\ [RC_s\ N_s\ \_]]\ \_]$, and 1% for non-intervening $N_p\,N_{s\ p}$ items with the structure $[S_s\ [NPC_s\ N_s\ [RC_p\ N_p\ \_]]\ \_]]$[6] . In experimental data, intervening and non-intervening attraction rates are found to be comparable; this is discussed further in the General Discussion.

3.4 Preamble error distribution

We next examined patterns of preamble errors in PIPS. This is a second test of how the model generalizes to other empirical data. In this model, preamble errors are fully grammatical sentences that are erroneous with respect to the input. The core phenomena this model should capture are displayed in Table 1 and Table 2; while Table 1 includes more data, Table 2 has a more specific coding of head and local preamble errors. Since we have only sparse human data on the rates of preamble errors, we focused primarily on evaluating the

[6] The sentence processing convention is to list the attractor (here, the first noun in the sentence) in the second position.

qualitative patterns identified in the introduction.

Figure 4 compares PIPS to the observed experimental data. The first observation, that local errors are more frequent than head errors, was matched in PIPS: the local error rate was 7.4%, versus a 3.3% head error rate (that is, the average value in the left column of plots is larger than the average value in the right). The second observation was that head and local error rates pattern in the same way with respect to head number, with plural heads associated with more head and more local errors. As shown in Figure 4, PIPS matched this pattern for head error rates (purple above green line), but showed the opposite pattern for local errors. Similarly, the interactions were reasonably matched by PIPS for head errors but not for local errors. This is clear by looking qualitatively at the simple effects of local number on head errors. In both PIPS and the empirical data, there is a large difference between the two head nouns paired with local singulars, and a small difference between the two head nouns paired with local plurals. For head errors, PIPS' plural-headed preambles showed more errors for singular vs. plural local nouns (n.b. the plural-headed effect was much stronger in PIPS' results than the experimental data). In contrast, PIPS' local errors showed the reverse pattern.

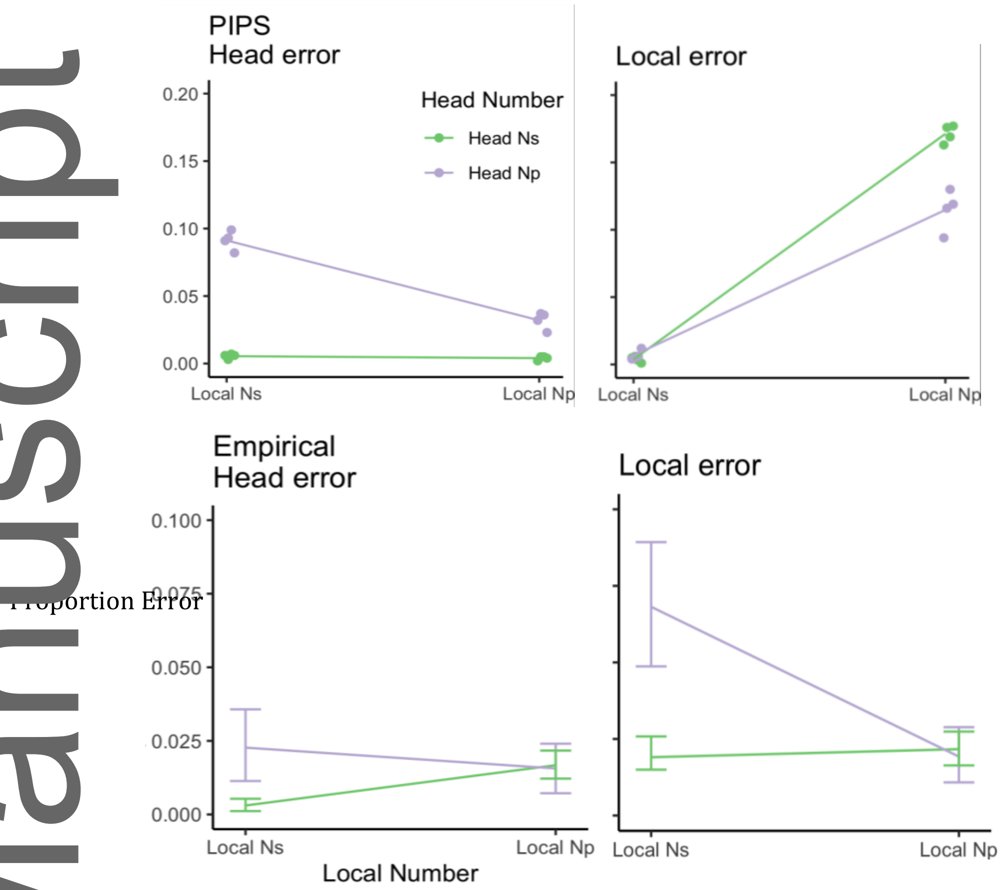

Figure 4. *Top row: Response probabilities in PIPS for the optimal model by preamble type. Points show results from individual model runs. Bottom row (repeated from Figure 1): empirical preamble error probabilities. 95% CIs generated by 5000 runs of a non-parametric bootstrap, sampling with replacement. Note that scales are different between rows to ease qualitative comparison.*

Combined, this means that PIPS elicited both preamble error types at reasonable rates relative to each other, with more local than head errors. Although PIPS was fit to verb errors alone, we have captured the critical main effect and interactions observed for head errors in experimental data, supporting our claim that verb errors and head errors are parallel consequences of competition in the same system. PIPS elicited local errors at a reasonable overall rate but elicited radically too few local errors in the $N_p N_s$ condition. This might be because the model was not configured correctly to elicit them, for example, because the planning constraints are different for head and local nouns. We return to this question in the Discussion.

3.5 Analysis of the components of PIPS

Having established that PIPS qualitatively matched the attraction effect and mismatch asymmetry for verb agreement errors and having tested its generalization to non-intervening attraction and preamble errors, we next explored how each parameter in PIPS contributes to error rates. We varied the type of memory encoding (lexical or structural), as well as representational similarity of lexical and structural elements. We also varied the model grammar. This section explores how these properties contributed to the overall pattern of behavior by manipulating each parameter in turn and examining how error probabilities qualitatively varied as these parameters shift. The goal is to show what each parameter contributes to the model, supporting development of a more accurate model.

3.5.1 Syntactic vs. lexical components of memory representations

Memory representations in PIPS encode information about the syntactic structure and lexical content of the preamble. To examine how these two aspects of memory may contribute to errors, we shifted the relative activations of the preamble's structural constituents and lexical items (i.e., the encoding strategy, reflected by parameter s1). Figure 5 shows how model outcomes for each preamble type changed as the encoding strategy shifted, varying from 0.0 (all lexicalist) to 1.0 (all structuralist) in steps of 0.1.[7]

[7] To ensure that we were accurately measuring trends in a stochastic distribution, we compared the model outputs reported in Figure 5, bottom left panel (NsNp input given to four sets of 1000 sentence generation iterations each with different random seeds, averaged) to a single set of 10000 iterations. This showed qualitatively similar results. The only reliable quantitative deviations

When following an entirely structuralist encoding strategy (high values of s1), pseudopartitive verb errors were infrequent for N_s N_p inputs. The corresponding weak activation of the lexical items was also associated with an increased rate of local errors, especially for N_s N_p inputs. This reveals an inverse relationship between local errors and agreement errors for N_s N_p inputs when the preamble structure has been encoded in memory with even moderate strength. It also suggests a possible tradeoff between local errors and attraction, since attraction is a difference score between the N_s N_p and N_s N_s inputs. This is consistent with the empirical results reported in Table 2 and has been remarked on previously by Thornton and MacDonald (2003).

When following a dominantly lexicalist encoding strategy (low values of s1), there were many pseudopartitive-containing verb errors produced for the N_s N_p preamble, with the peak of verb errors at s1 = 0.2. This suggests the importance of local lexical input in driving attraction, as in all other models of agreement (M&M, SOSP, and ACT-R). There was also a high rate of head errors, particularly for both types of mismatch sentences (N_s N_p and N_p N_s). This suggests that head errors might be made as a way of compensating for conflict between locally disagreeing elements. The local conflict that leads to attraction can be resolved by changing the head number to match the local noun; however, since this also requires changes to the number on the noun phrase and sentence, this type of change may be most feasible when the higher-order structure is weakly encoded. The implication of this pattern is that head errors should correlate with agreement errors. This has not been discussed in the literature, in part due to the low base rates of head error occurrence. The data summarized in Table 2 are not inconsistent with this pattern, but do not allow us to conclusively test it.

---

were found at s0 = 0.0, with relatively small (<.015) deviations in proportion of correct productions and verb errors. However, these deviations did not alter the overall trends noted in the section.

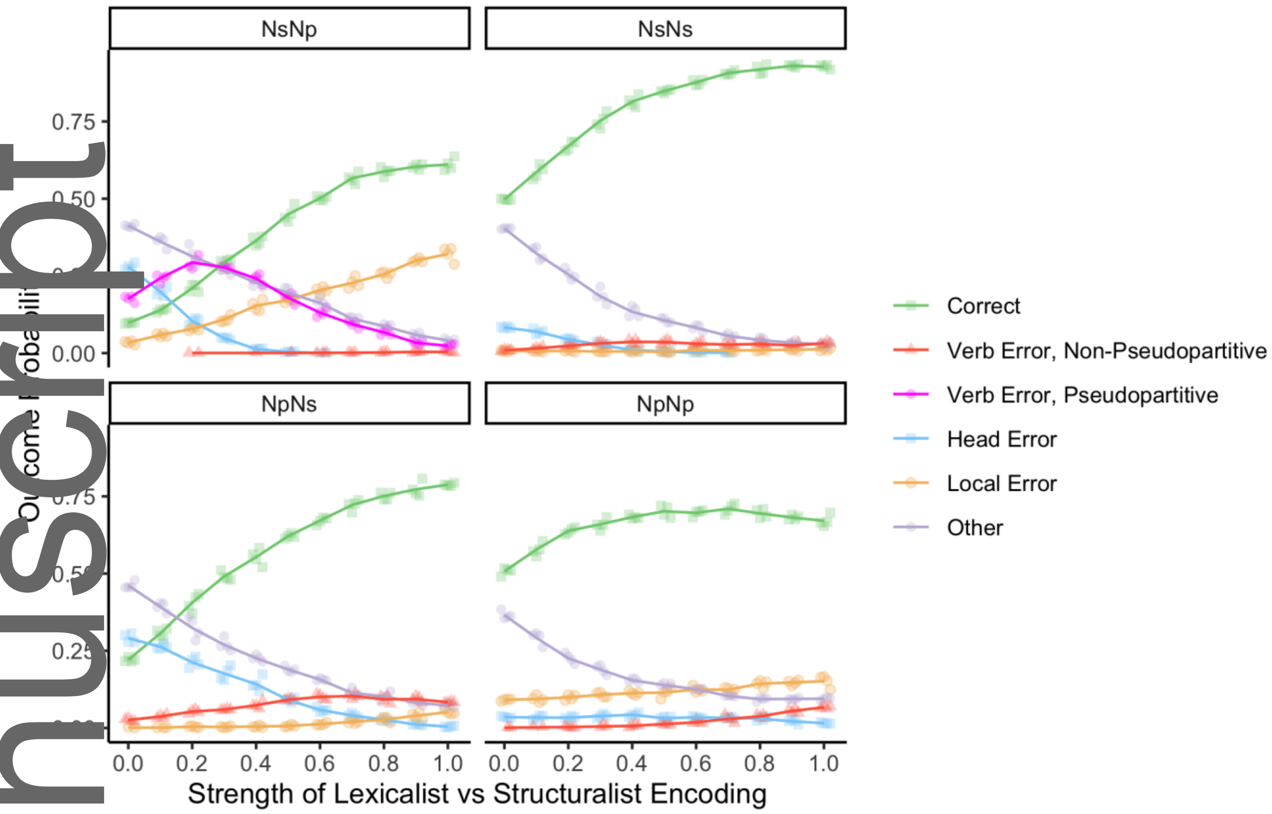

Figure 5. Response probabilities for the optimal model (structural constituent similarity = 0.5, noun similarity = 0.7, and s1 weight= 0.5) varied by encoding strategy (s1 weight) and preamble type (panels). Points correspond to individual model runs and lines correspond to averages across runs with matching settings.

3.5.2 Continuous representational similarity: Lexical items

Next, we assessed how changing the similarity of the singular and plural form of nouns impacted model outcomes. Since constituents are represented as vectors, we could manipulate the representational similarity of singular and plural nouns in a continuous fashion (here, varied at a step size of 0.1 between 0.2 and 0.7). As shown in Figure 6, verb errors for mismatch preambles ($N_s N_p$ and $N_p N_s$) decreased as noun similarity increased. This suggests that increasing the similarity of noun forms may decrease competition from alternate parses, reducing the verb error rate. On first glance, this may seem to contrast with the observation that high semantic similarity (e.g. Barker et al., 2001) or tight semantic integration (Solomon & Pearlmutter, 2004) increases verb errors; however, note that our manipulation is of noun forms differing in number, not semantics.

The differential effects of noun similarity for verb errors on match versus mismatch preambles means that noun similarity co-varied with attraction (the difference in all verb errors, shown in red and magenta, between N_s N_p and N_s N_s sentences), but not the mismatch asymmetry (the difference in all verb errors, shown in red and magenta, between N_s N_p and N_s N_s sentences). This is consistent with the hypothesized role for cue-based memory retrieval in eliciting agreement attraction (e.g. Badecker & Kuminiak, 2007; Lewis & Vasishth, 2005; Lorimor, et al. 2015, 2016).

Local errors also increased for preambles with plural local nouns (N_s N_p and N_p N_p), but not for preambles with singular local nouns (N_s N_s and N_p N_s), as noun terminal similarity increased. One possible explanation for this is that while increasing noun similarity may allow singular local nouns to replace plural ones more easily (and vice-versa), singular nouns' higher harmony (reflecting their higher frequency) supports accurate processing. Further tests would be necessary, however, to support this hypothesis, which we leave for future work.

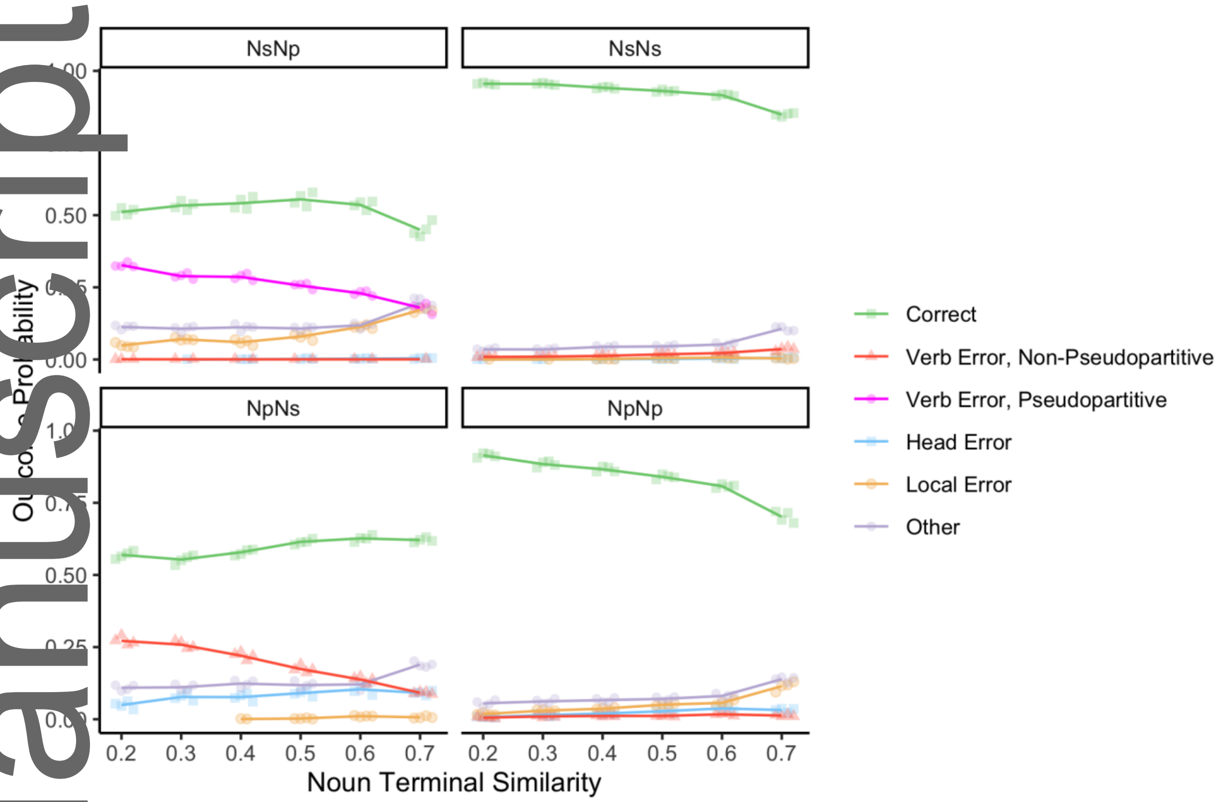

Figure 6. *Response probabilities for the optimal model (structural constituent similarity = 0.5, noun similarity = 0.7, and s1 weight= 0.5) varied by noun terminal similarity (similarity of singular and plural forms) and preamble type (panels). Points correspond to individual model runs and lines correspond to averages across runs with matching settings.*

3.5.3 Continuous representational similarity: Structural constituents

Finally, we examined the role of structural constituent similarity in model outcomes, varying this parameter at a step size of 0.1 from 0.2 to 0.7. A plot of model outcomes appears in Figure 7. Increasing structural constituent similarity tended to slightly decrease the correct rate for all preambles, and clearly decreased the correct rate for the $N_s N_p$ condition: increasing similarity between pairs of structures, like the various NP structures, may allow them to compete more for selection.

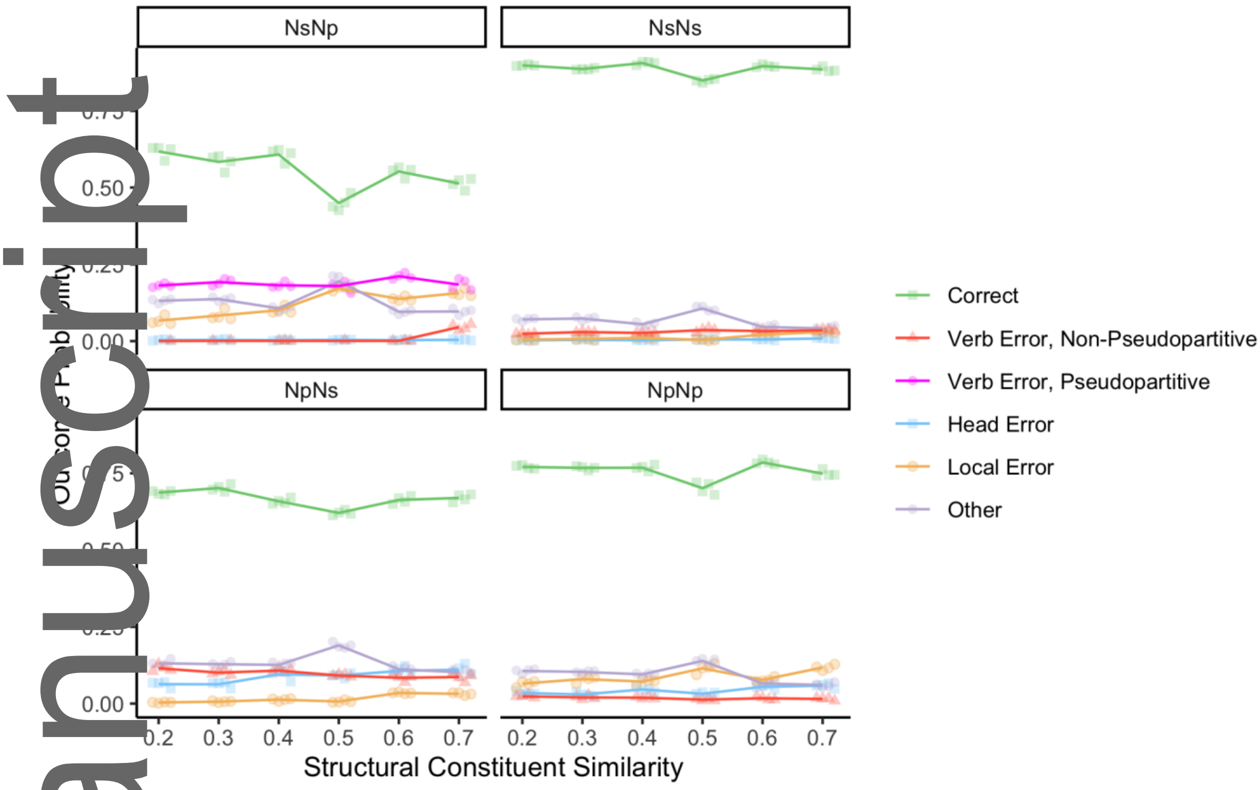

Figure 7 Response probabilities for the optimal model (structural constituent similarity = 0.5, noun similarity = 0.7, and s1 weight= 0.5) varied by structural constituent similarity (similarity of symbols representing different types of singular/plural NP, RC, and S) and preamble type (panels). Points correspond to individual model runs and lines correspond to averages across runs with matching settings.

3.6 Competition and experience

Core to PIPS is the claim that competition between structures drives errors. To further investigate the contribution of this component of the model, we altered the structure of the training data — lowering the frequency of the pseudopartitive parse so as to reduce competition while holding the free model parameters constant at the optimal values selected above. Comparison between the output of this model and the base model reveals how aspects of the model's training experience influenced its performance.

We reduced the frequency of sentence 8 (the pseudopartitive) in the training data to

25% of the original value and re-fit the model to these data. As competition from the pseudopartitive decreased, verb errors for N_s N_p preambles decreased from 18% to 16%. The shift in training data also led to increased verb errors for N_p N_s and N_p N_p preambles (from 9% to 12% and from 1% to 3%, respectively). The source of this is unclear; reducing the probability of the pseudopartitive may weaken the diversity of structures supporting plural verbs, leading to increased errors where singular verbs replace plural verbs. These combined effects reduce the mismatch asymmetry for plural vs. singular local nouns from 9% (in the original model) to 4%.

Follow-up simulations that entirely eliminated the pseudopartitive parse resulted in a *reversal* of the mismatch asymmetry effect: in these simulations, the N_p N_s Vs preamble (with a lower-frequency plural head) elicited more total verb errors than the N_s N_p Vp preamble (19% vs. 8%) and the difference between the N_p N_s Vs and N_p N_p Vs preambles was greater than the difference between the N_s N_p Vp and N_s N_s Vp preambles (13% vs. 5%). This suggests that the availability of the pseudopartitive parse is important for eliciting the mismatch asymmetry effect observed in English.

4. Discussion

We have presented PIPS, a model of sentence production. When applied to agreement production, PIPS accounts for verb agreement errors by relying upon domain-general principles. The model uses a grammar rooted in the phrase frequencies of American English, such that elements that appear more often are preferred by the model. To produce a sentence, the model uses spreading-activation rules to activate representational state vectors corresponding to lexical and structural constituents; in our modeled preamble completion task, the model's activated representational state vectors come from a combination of a partially-activated (fuzzily remembered) input and the model's grammatical knowledge.

The premise of this GSC-style model is that blend states consisting of multiple simultaneously-represented possible utterance plans play a role in language production. Initially, the model simultaneously activates many possible structures in the form of blends because of low commitment strength; over the course of processing, these structures compete until the model is pushed to select a single structure via increasing commitment strength. These transiently activated blends lead to competition between target and non-target structures, allowing PIPS to make human-like errors.

Experimental research on agreement production has focused on a few key phenomena, including agreement asymmetries based on grammatical number and the role of notional (semantic) number in agreement (e.g., Bock & Miller, 1991; Eberhard et al., 2005, Eberhard, 1999, Vigliocco et al., 1995), the role of exposure to various agreement configurations (e.g. Haskell et al. 2010; Lorimor et al., 2018) and the general constraints that memory places on agreement (e.g. Hartsuiker & Barkhuysen, 2006; Slevc & Martin, 2016). PIPS successfully captures these properties of agreement and takes a step towards covering other types of errors commonly elicited in the same experimental tasks, as we unpack below.

4.1 Accounting for verb errors

Critical in a model that captures subject-verb number agreement is how it can account for two agreement asymmetries that reflect the differential susceptibility of sentence types to agreement errors based upon the number configuration of the two nouns. These are attraction and the mismatch asymmetry; in PIPS, both arise from competition between similar constituents (lexical items and structures) in the grammar. In PIPS, the mismatch asymmetry arises because of a second observation in the empirical data: a grammatical construction associated with notional number is available for N_s N_p phrases only (e.g. Haskell et al., 2010).

Attraction is the pattern that noun phrase preambles containing a singular head and plural local noun (N_s N_p; *The key to the cabinets*) elicit more verb errors than those with a singular head and singular local noun (N_s N_s; *The key to the cabinet*). In PIPS, we believe this attraction comes from the transient influence of other sentences in which a verb agrees locally, such as a reduced relative clause (*The key [the cabinets use] broke*), and from the high lexical/syntactic similarity of all nouns in the model grammar regardless of their grammatical number. The implication is that attraction likely arises because of sometimes transient activation of similar structures and words, most of which are contained in grammatical constituents such as the pseudopartitive or reduced relative clauses.

The mismatch asymmetry is the pattern that noun phrase preambles containing a singular head and plural local noun (N_s N_p; *The key to the cabinets*) elicit more verb errors than those with a plural head and singular local noun (N_p N_s; *The keys to the cabinet*) despite the fact that the head and local noun mismatch in both preambles, and despite the fact that singular nouns have a higher overall frequency. In PIPS, the mismatch asymmetry comes from a grammatical parse that is typically associated with notional agreement. In American English, it is relatively common to have a pseudopartitive parse of N_s N_p phrases where the second noun becomes the agreement controller (e.g. *A lot of postcards are*). Following the high frequency of these phrases in American English, and inspired by earlier experimental (Haskell et al., 2010) and modeling work (SOSP; Smith et al., 2018), we implemented a pseudopartitive parse in PIPS where the nouns N_s N_p expand to a plural noun phrase. In a set of simulations reducing or removing the pseudopartitive parse, the mismatch asymmetry diminished or vanished, suggesting this non-target parse plays an important role in creating the asymmetry in error rates between N_s N_p and N_p N_s items. Note that since the plural-headed pseudopartitive parse is in the (non-lexicalized) PIPS grammar, activating it can be grammatically correct if the appropriate higher-order structure is also changed. Critically, while this parse is grammatical, it is still an error. This parse is in conflict with the input, where the higher-order structure specifies N_s N_p items as belonging to a singular NP. This makes all first-noun mismatching verbs errors with respect to the input, even if they are in some cases grammatically correct.

PIPS achieves the mismatch asymmetry despite lacking semantics: only one noun, and one verb, are instantiated in the grammar for each number inflection. This means that the probabilistic availability of the pseudopartitive in the grammar reflects inherent flexibility in the class of noun phrases as a whole, akin to the distributive and non-distributive readings of phrases like *The label on the bottles.* Haskell et al. (2010) suggested that the availability of a right-headed parse changes the production of verbs in general. Our model follows this assumption and shows that grammatically allowing for a construction associated with notionally-driven agreement creates the mismatch asymmetry. In other words, the mismatch asymmetry arises because of grammaticalization of exposure to notional agreement. The implication is therefore that languages without a pseudopartitive construction should not elicit a mismatch asymmetry effect. The size and presence of mismatch asymmetries varies across languages, especially for gender agreement (see e.g. Franck, 2018; Deutsch & Dank, 2011), and it has been argued that either morphological decomposition (Deutsch & Dank, 2011) or the relative markedness of the head noun (Franck, 2018) may contribute to these differences. Following Haskell et al., (2010) and Franck (2018), we suggest that the presence of notionally-agreeing constructions like pseudopartitives also contributes to the effect; this is a prediction to be tested in future research.

We also capture the mismatch asymmetry in so-called *non-intervening* or *remote* attraction cases (*The cabinets that the key *were used to open*), which is a strong test of the model. Earlier work in English (Bock & Miller, 1991; Staub, 2009, 2010) demonstrates an asymmetry such that *The cabinets that the key* elicits more attraction than *The cabinet that* *the keys*. However, non-intervening attraction also follows a different error reaction time profile than other attraction cases (Staub, 2010). Staub suggests that this means non-intervening attraction occurs because the structure of the sentence was mis-identified while intervening attraction occurs because of feature percolation; this left a mystery why the markedness asymmetry would be present.

PIPS provides one possible explanation. In PIPS, the mismatch asymmetry comes from the pseudopartitive boosting errors in N_s N_p items. In the case of intervening attraction, which is seeded with the input structure $[S_s\ [NPC_s\ N_s\ N_p]\ \_\ ]$, the elements comprising the pseudopartitive easily turn on because of the similar structure to the input, producing the output $[S_p\ [NPC_p\ N_s\ N_p]\ V_p]$). In the case of non-intervening attraction, the input structure is $[S_p\ [NPC_p\ N_p\ [RC_s\ N_s\ \_\ ]]\_]$, which differs from the pseudopartitive in its branching structure and in the type of NP node; this often elicits completions like $[S_s\ [NPC_s\ N_s\ N_p]\ V_s]$, a fully grammatical three-word sentence. Following Staub (2009, 2010), this means that while attraction happens in both cases by competition from a set of self-reinforcing symbols, the competition in non-intervening cases means entertaining a structural reanalysis. Note that this is a fundamentally different mechanism than the hierarchical planning account proposed by Franck and colleagues (2002, 2006). Their account does especially well in distinguishing which non-intervening configurations elicit attraction, and it would be worth future work to evaluate how well PIPS does in replicating these data.

One weakness of PIPS's account of non-intervening attraction is that attraction error rates were much lower for non-intervening attraction than for intervening attraction. This contradicts empirical data from human errors, where attraction was of similar magnitude in both cases (Bock & Miller, 1991; Staub, 2009, 2010). A possible reason for this is that that we initialized the model with an equal amount of activation at the terminal nodes, encoding the words of the preamble, and the non-terminal nodes, encoding its structure. Because the non-intervening attraction sentences are more complex than the canonical attraction sentences, there is more external activation supporting the structural analysis; this may protect it from errors. Equalizing the overall amount of activation (rather than the activation of individual elements) might provide for a better fit to the empirical data, and would be a worthy extension for future work.

The reliance on domain-general computational mechanisms such as spreading activation and competition means that PIPS implements agreement in a way that is radically different from the leading model of agreement production, Marking and Morphing (M&M, Eberhard et al., 2005). M&M was created to describe the influence of lexical, structural, and notional information in agreement. Specifically, it aimed to explain how lexical and semantic properties of nouns affect verbs and pronouns, and the way that grammatical and notional number are reconciled to allow both to contribute to agreement using a simple set of rules.

In PIPS, we focused on the first of these explananda. We tested the role of lexical factors including lexical and structural representational similarity and the strength with which words and phrases are encoded or retrieved. To elicit human-like agreement patterns, we needed high lexical similarity, moderately high structural constituent similarity and a memory representation that is balanced for lexical and structural information. This means that both M&M and PIPS agree that in order to account for agreement errors, the system must strike a balance between lexical (local) and structural (global) needs.

As noted above, PIPS has no semantics, which means that we account for notional number indirectly: by implementing a syntactic rule corresponding to a pseudopartitive noun phrase. This is a divergence from both SOSP (Smith et al., 2018) and M&M (Eberhard et al., 2005), which explicitly integrate semantic and syntactic aspects of agreement. This was a purposeful choice made to explicitly examine the effects of lexical and structural competition without needing to incorporate meaning.

However, an implementation of semantics might allow PIPS to go further in describing notional agreement, and in particular, in describing the lexical and semantic conditions that encourage activation of the pseudopartitive parse associated with notional plurality (e.g. Eberhard, 1999; Humphreys & Bock, 2015); this has been extensively explored in M&M (Eberhard et al., 2005). In PIPS, this could be instantiated by manipulating representational similarity in a larger grammar based upon shared semantic/ syntactic features between multiple word and sentence types. This would make PIPS more similar to SOSP in its treatment of treelets as parts of a semantic-syntactic representation. An incorporation of message-level representation in the PIPS model might provide further evidence for the role of message conceptualization in sentence production, which has received relatively little attention in the literature (see e.g. Konopka & Brown-Schmidt, 2014, for review).

Incorporating semantics in PIPS would also allow us to test how different types of similarity affect errors. While we show that increasing structural similarity between nouns decreases verb errors, because the local noun becomes less good as an attractor, earlier work shows that increasing semantic similarity increases verb errors because of interference at encoding (e.g. Barker et al., 2001) or because of differences in message-level representation (Solomon & Pearlmutter, 2004). This difference could result because the two types of similarity elicit different types of competition: future work would be needed to test this.

4.2 Extension to preamble errors

In agreement-error elicitation experiments, the most common mistakes are not agreement errors, but miscellaneous errors. Preamble errors — errors mis-recalling the head or local noun—are an important sub-type of these miscellaneous errors. Because the representational space in PIPS is gradient, it easily makes errors in reproducing its input because of competition from other items in the grammar. Effectively, noise in the system can cause the model to jump from one grammatical state, matching the input, to another: the preamble error. The pattern of head errors is captured fairly well in the model. By changing how the input is encoded, altering the degree and type of representational similarity between model constituents, and altering sentence frequencies in the model's grammar, we show that head errors and attraction errors are both consequences of the same model dynamics. Note that because the modeled data are sparse, this forms an empirical prediction to be validated in future experimental work. We urge individuals to report noun number errors in agreement elicitation experiments to further test these claims.

The fact that PIPS *can* account for multiple error types means that it takes a step beyond other models used to describe agreement production: M&M, in particular, has no way of incorporating preamble errors because it assumes that the preamble has been encoded accurately and influences number marking on the phrase. SOSP has also focused to date on modeling blend or coercion errors, not on modeling fully grammatical productions that mismatch the input, which means that SOSP also does not have a straightforward way of modeling preamble errors. However, more refinements need to be made in order for PIPS to better match human data. While PIPS succeeds at capturing the fact that local errors are more common than head errors, and approximates the overall pattern for head errors, it fails to capturing the effects of head and local number on local errors. In empirical data, N_pN_s prompts elicit the most local errors: this is not the case in PIPS. One possibility is that the empirical data are too sparse to disclose the true pattern with respect to head and local number; this seems unlikely, but is possible.

Another more interesting possibility is that PIPS fails to account for local preamble because the model does not produce incrementally: in PIPS, the entire noun phrase and verb are planned before production begins. Since local errors are linearly close to the verb, it follows that a more incremental scope of planning should boost local errors over head errors. This is consistent with earlier empirical work where local nouns that are closer to the head in linear planning scope induce more attraction (e.g. Gillespie & Pearlmutter, 2011); the prediction is that a narrower scope of planning might also influence local preamble errors. Developing an incremental version of PIPS would allow us to test this, and it would be a worthy modeling extension in its own right.

A final possibility is that $N_p N_s$ items are notably odd for semantic or pragmatic reasons. The same nouns are typically used in all four cells of most preamble completion studies; the $N_p N_s$ items could be less plausible than the rest. If this is the case, individuals will misrepeat the input in order to correct a perceived infelicity (see e.g. Brehm, Jackson, & Miller, in press, for data consistent with this hypothesis). This would suggest that local errors may in part be due to a secondary phenomenon not directly related to computing subject-verb agreement.

4.3 Effects of frequency in agreement

A novel aspect of PIPS relative to other models that explain agreement production is that it encodes a grammar trained on the relative frequency of structures in American English. This allowed us to explicitly test the role of frequency in the model, which is an important way that individual experience might affect agreement production. The role of frequency is clearly highlighted in the PIPS model, and underscores that in our model, agreement and preamble errors are consequences of the same dynamics.

Structure frequencies generate the crucial mismatch asymmetry in PIPS, providing a mechanism to explain why this asymmetry appears in English. In simulations where the frequency of pseudopartitive phrases is decreased in the training grammar, the mismatch asymmetry diminishes. This suggests that differences in whether languages have a pseudopartitive or other similar construction could be critically important in whether a given language elicits a mismatch asymmetry.

Recent and past experience matter for sentence production, as demonstrated in the large literature on syntactic priming (Bock, 1986; see Mahowald, James, Futrell, and Gibson, 2016 for meta-analysis) and experience-based changes to sentence biases over time (e.g. Chang, Dell, and Bock, 2006; Gennari & MacDonald, 2009; Ferreira & Schotter, 2013; Konopka, 2012). Correspondingly, the literature on subject-verb agreement (e.g. as highlighted in Haskell et al., 2010) suggests an important role of priming (short-term experience) and frequency (a contributor to long-term experience).

One could easily extend PIPS to account for short-term experience changes and to more deeply explore the consequences of structural frequencies. To account for short-term experience, one could update constituent weights between trials. This would allow an examination of how error production changes due to priming or learning (as discussed in Haskell et al., 2010). Exploring frequency more deeply would also be fruitful: this would allow an examination of how structural frequency differences within or between languages affect what types of errors are produced (as discussed in e.g., Bock et al., 2012 or Foote & Bock, 2012).

4.4 Memory encoding and memory retrieval

A final property we appeal to in PIPS is the role of memory in sentence production. As in experimental elicitation of agreement errors, PIPS is seeded with a preamble as input, and asked to repeat and complete it. Mismatches between the input and output correspond to memory errors. By only partially encoding words and structures, PIPS makes mistakes in a human-like way, resolving the competition it receives from other items in its grammar by producing an error on the head number, local number, or verb. To manipulate memory encoding, we use a pair of yoked parameters that tune whether lexical items or structure is encoded more veridically; future applications of PIPS could separate these parameters to test the role of memory encoding for whole structures, treelets, or lexical items in error production.

Further refinements of PIPS might also consider a more elaborated implementation of memory retrieval. For example, not only would target elements be weakly activated at encoding, allowing the activation of multiple different elements during production planning, but specific non-target elements might receive boosts in activation at the onset of production in order to represent mis-retrieval. This would allow a closer comparison to existing domain-general models like ACT-R (e.g. Lewis & Vasishth, 2005; as discussed in Lorimor et al., 2015, 2016) which capture errors in agreement production by appealing to mis-retrieval of words or features from memory. Directly contrasting the role of encoding versus retrieval dynamics in sentence production would also allow for deeper connections to be made with related effects in sentence comprehension (e.g. Villiata et al, 2018) and production (Barker et al., 2001), where semantic properties of items cause interference at encoding. This would allow us to test whether we observe the same semantically-driven interference pattern.

4.5 Domain generality allows for many extensions

Comparison between how PIPS and other models (M&M, SOSP) account for the empirical data from agreement production highlights a final key fact. The apparently domain-specific phenomena of attraction errors, mismatch asymmetry errors, and head preamble errors arise in PIPS out of domain-general properties. This is a consequence of PIPS being a dynamical systems model in which multiple elements compete during language production or parsing. When other seemingly-discrete phenomena are investigated using dynamical systems models like GSC (Cho et al., 2017), or SOSP (Smith & Tabor, 2018), previously unconnected types of errors arise as consequences of the same continuous constraint space.

The fact that PIPS generates errors of multiple types highlights that while PIPS models agreement attraction, it is not specifically a model of agreement attraction. This shows the utility of adopting a general model such as PIPS to explore the role of linguistic and cognitive principles in language production more broadly. As highlighted throughout this section, there are many extensions possible within PIPS to explore how the structure of the grammar and the general distributions of forms affect error production. There are still many disagreements within the field of agreement production to be answered. PIPS does not clearly capture the patterns observed for local preamble errors, which should be examined in future work. We also suggest that a deeper and wider exploration of how other types of preamble errors, verb errors, and notional agreement are inter-related would be fruitful, as would an exploration of agreement in richer inflectional paradigms where more candidates are available for production.

Within sentence production and comprehension, there are also many other phenomena in which an appeal to multiple simultaneously active elements is necessary, and these would also be good candidates for exploration using a model like PIPS where planning starts with conjunctive blends and ends with fully discrete outcomes. This includes the many diverse phenomena addressed by constraint-based theories of sentence comprehension (e.g. MacDonald, Pearlmutter, & Seidenberg, 1994) or questions about interpretation or binding in systems where more than one possibility is likely (scalar implicatures, e.g. Degen & Tanenhaus, 2015; pronoun resolution: see e.g. Arnold & Zerkle, 2019 for recent review). PIPS or a similar model might provide new leverage to explain how these occur in the human mind.

5. Conclusion

Preamble completion paradigms highlight two common production errors: subject-verb agreement errors and errors in repeating the sentence preamble prompt. Using PIPS, a computational model in the GSC framework, we describe both as consequences of competition during transient activation of similar competing sentences in a grammar that is trained on structure frequencies from American English. This provides computational evidence for the role of alternate production plans at a morphosyntactic and lexical level in explaining why speech errors happen without modeling semantics.

Acknowledgements

This research was funded by INSPIRE Grant BCS1344269 and the College of Literature, Science and Arts, University of Michigan. We would like to thank the GSC team, the MPI Simulation and Modeling User Group, and Antje Meyer for helpful feedback.

Conflict of Interest

The authors have no known conflicts of interest to disclose.

Data Avaliability Statement

All data and materials from this study are available at https://osf.io/3udb8/

References

Acuña-Fariña, J. C. The role of morphology in setting production biases in agreement: A cross-linguistic completion study. *Language Sciences* 66 (2018): 28-41.

Antón-Méndez, I., & Hartsuiker, R. J. (2010). Morphophonological and conceptual effects on Dutch subject-verb agreement. *Language and Cognitive Processes*, *25*(5), 728-748.

Arnold, J. E., & Zerkle, S.A. (2019) Why do people produce pronouns? Pragmatic selection vs. rational models, *Language, Cognition and Neuroscience*, 34:9, 1152-1175,

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*(1), 65-85.

Barker, J., Nicol, J., & Garrett, M. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research, 30(1),* 91-114.

Bergen, L., & Gibson, E. (2012). *Agreement errors as rational encoding errors*. Poster presented at the 25th Annual CUNY Conference on Human Sentence Processing, New York, NY.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355-387.

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*(1), 99-127.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, *23*(1), 45-93.

Bock, K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, *66*(1), 17-37.

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, *51*(2), 251-278.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, *43*(2), 83-128.

Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement?. *Cognition*, *128*(2), 149-169.

Brehm, L., & Bock, K. (2017). Referential and lexical forces in number agreement. *Language, Cognition and Neuroscience*, *32*(2), 129-146.

[dataset]Brehm, L., Goldrick, M., Cho, P. W., & Smolensky, P. (2021). PIPS: Parallelism in Planning Syntax. Retrieved from osf.io/3udb8

Brehm, L., Hussey, E., & Christianson, K. (2020). The role of word frequency and morpho-orthography in agreement processing. *Language, Cognition and Neuroscience*, *35*(1), 58-77.

Brehm, L., Jackson, C. N., & Miller, K. L. (in press). Probabilistic online processing of sentence anomalies. *Language, Cognition, and Neuroscience*.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234.

Chattopadhyay, A., Manupriya, P., Sarkar, A., & Balasubramanian, V. N. (2019, May). Neural network attributions: A causal perspective. In *International Conference on Machine Learning* (pp. 981-990). PMLR.

Cho, P. W., Goldrick, M., & Smolensky, P. (2017). Incremental parsing in a continuous dynamical system: Sentence processing in Gradient Symbolic Computation. *Linguistics*

Vanguard, 3, 20160105.

Cho, P. W., Goldrick, M., & Smolensky, P. (2020) Parallel parsing in a Gradient Symbolic Computation parser. https://doi.org/10.31234/osf.io/utcgv

Cho, P. W., Goldrick, M., Lewis, R., & Smolensky, P. (2018). Dynamic encoding of structural uncertainty in gradient symbols. Proceedings of *Cognitive Modeling and Computational Linguistics*

Coppock, E. (2010). Parallel grammatical encoding in sentence production: evidence from syntactic blends. *Language and Cognitive Processes, 25,* 38–49.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*(1), 57-71.

Davies, M. (2008). The corpus of contemporary American English (COCA): 520 million words, 1990-present.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint- based approach. *Cognitive Science*, *39*(4), 667-710.

Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language*, *60*(1), 112-143.

Deutsch, A., & Dank, M. (2011). Symmetric and asymmetric patterns of attraction errors in producing subject — predicate agreement in Hebrew: An issue of morphological structure. *Language and Cognitive Processes*, *26*(1), 24-46.

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject — verb agreement in English. *Journal of Memory and Language*, *41*(4), 560-578.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, *112*(3), 531.

Emmorey, K., Borinstein, H., Thompson, R., & Gollan, T. (2008). Bimodal bilingualism. *Bilingualism: Language and Cognition, 11*(1), 43-61.

Ferreira, V. S., & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors?. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1548-1571.

Foote, R., & Bock, K. (2012). The role of morphology in subject — verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes*, *27*(3), 429-461.

Franck, J. (2018). Syntactic encoding: Novel insights into the relationship between grammar and processing. *The Handbook of Psycholinguistics*. pp. 13-37.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, *101*(1), 173-216.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371-404.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*(3), 329-374.

Frazier, L., & Clifton Jr, C. (2015). Without his shirt off he saved the child from almost drowning: interpreting an uncertain input. *Language, Cognition and Neuroscience, 30(6),* 635-647.

Fyndanis, V., Arcara, G., Christidou, P., & Caplan, D. (2018). Morphosyntactic production and verbal working memory: Evidence from Greek aphasia and healthy aging. *Journal of Speech, Language, and Hearing Research*, *61*(5), 1171-1187.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1-23.

Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject–verb agreement production. *Cognition, 118(3)*, 377-397.

Goldrick, M., & Daland, R. (2009). Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology, 26*, 147-185.

Goldrick, M., Keshet, J., Gustafson, E., Heller, J., & Needle, J. (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition, 149*, 31-39.

Goldrick, M., Putnam, M., & Schwarz, L. (2016). Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition, 19(5),* 857–876. doi: 10.1017/S1366728915000802

Hale, J., & Smolensky, P. (2006). Harmonic Grammars and harmonic parsers for formal languages. *Smolensky and Legendre*, 393-416.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive psychology*, *110*, 70-104.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*(1-3), 181-204.

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & cognition*, *31*(8), 1316-1326.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject — verb agreement. *Journal of Memory and Language*, *48*(4), 760-778.

Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, *114*(2), 151-164.

Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, *12*(4), 689-695.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316-339.

Jescheniak, J.D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 1256-1274.

Keung, L. C., & Staub, A. (2018). Variable agreement with coordinate subjects is not a form of agreement attraction. *Journal of Memory and Language*, *103*, 1-18.

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology, 124,* 101359.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143-162.

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. *The Oxford handbook of language production*, 3-20.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73*, 1-40.

Kuchinsky, S. E., & Bock, K. (2010). From seeing to saying: Perceiving, planning, producing. *Paper presented at the 23rd CUNY human sentence processing conference, New York*.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133-149.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375 — 419.

Lorimor, H., Adams, N. C., & Middleton, E. L. (2018). Agreement with conjoined NPs reflects language experience. *Frontiers in psychology*, *9*, 489.

Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and cognitive processes*, *23*(6), 769-799.

Lorimor, H., Jackson, C. N., & Foote, R. (2015). How gender affects number: Cue-based retrieval in agreement production. *Language, Cognition and Neuroscience*, *30*(8), 947-954.

Lorimor, H., Jackson, C. N., Spalek, K., & van Hell, J. G. (2016). The impact of notional number and grammatical gender on number agreement with conjoined noun phrases. *Language, Cognition and Neuroscience*, *31*(5), 646-661.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676.

Mirković, J., & MacDonald, M. C. (2013). When singular and plural are both grammatical: Semantic and morphophonological effects in agreement. *Journal of memory and language*, *69*(3), 277-298.

Momma, S., & Ferreira, V. S. (2019). Beyond linear order: The role of argument structure in speaking. *Cognitive psychology, 114*, 101228.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555.

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology, 69(5),* 950-971.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348-379.

Sankoff, D., Poplack, S., & Vannianiarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change, 2,* 71–101.

Santesteban, M., Pickering, M. J., & Branigan, H. P. (2013). The effects of word order on subject-verb and objectverb agreement: Evidence from Basque. *Journal of Memory and Language, 68(2),* 160-179.

Slioussar, N., & Malko, A. (2016). Gender agreement attraction in Russian: production and comprehension evidence. *Frontiers in Psychology*, *7*, 1651.

Slevc, L. R., & Martin, R. C. (2016). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, *28*(7), 773-790.

Smith, G., Franck, J., & Tabor, W. (2018). A self- organizing approach to subject — verb number agreement. *Cognitive Science*, *42*, 1043-1074.

Smith, G., Franck, J., & Tabor, W. (2021). Encoding interference effects support self-organized sentence processing. *Cognitive Psychology, 124*, 101356.

Smith, G., & Tabor, W., (2018). Toward a theory of timing effects in self-organized sentence processing. Proceedings of the International Conference on Cognitive Modeling 2018, Madison, WI.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, *46*(1-2), 159-216.

Smolensky, P. (2006). Tensor product representations: Formal foundations. The harmonic mind: From neural computation to Optimality-Theoretic grammar, ed. by P. Smolensky & G. Legendre, 271-344.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science, 38*, 1102-1138.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49(1)*, 1-46.

Strang, B. M. H. (1966). Some features of S-V concord in present-day English. In I. Cellini & G. Melchiori (Eds.), English studies today: Fourth series (pp. 73-87). Rome, Italy: Edizioni di Storia e Letteratura.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language, 60(2),* 308-327.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition, 114(3),* 447-454.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195-215.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*(4), 740-759.

Veenstra, A., Antoniou, K., Katsos, N., & Kissine, M. (2018). Resisting attraction: Individual differences in executive control are associated with subject — verb agreement errors in production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1242.

Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, *45*(3), 368-390.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*(3), 261-298.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186-215.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, *9*, 2.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206-237.

Appendix A

Counts of responses from spoken preamble completion paradigms in American English using the structure NP_1 [PP P NP_2], where the head of NP_1 is the head noun and the head of NP_2 is the local noun. Ns = singular count noun; Np = plural count noun; Sg/Pl Coll = singular/plural collective noun; Sg/Pl Irr = irregular singular/plural noun, including pseudoplurals (e.g. *rose*), pluralia tanta (e.g. *tweezers*), mass nouns (e.g. *rice*), and nouns with irregular morphology (e.g. *mice*); NI = non-inflecting count noun (e.g. *fish*). A breakdown of 'Other' responses is included when reported in each paper; these include 'preamble' errors (also called 'miscellaneous errors' Pre; completions where preamble was not repeated veridically), 'uninflected' completions (Uninfl; completions using a lexical verb that does not require marking of inflection), and any missing trials (NR). These data are available in plain text format on the OSF.

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Bock et al., 2006	1	Sg Coll	Ns	Sg	327	435	390	390	0	0
Bock et al., 2006	2	Sg Coll	Ns	Sg	157	50	369	72	297	0
Bock et al., 2006	3	Sg Coll	Ns	Sg	567	7	146	55	91	0
Bock et al., 1999[8]	1	Sg Coll	Ns	Sg	273	557	466	.	.	.
Humphreys & Bock, 2005	1	Sg Coll	Ns	Sg	100	39	293	79	214	0
Bock et al., 2004	3	Sg Coll	Ns	Pl	170	79	135	56	79	0
Bock et al., 2006	1	Sg Coll	Np	Sg	220	497	435	435	0	0
Bock et al., 2006	2	Sg Coll	Np	Sg	103	112	361	86	275	0
Bock et al., 1999	1	Sg Coll	Np	Sg	153	571	572	.	.	.
Humphreys & Bock, 2005	1	Sg Coll	Np	Sg	45	92	295	97	198	0
Bock et al., 2004	3	Sg Coll	Np	Pl	105	109	170	103	67	0
Humphreys & Bock, 2005	1	Sg Coll	Np	Pl	41	120	271	65	206	0
Bock & Eberhard, 1993	4	Ns	Sg Coll	Sg	292	1	91	33	58	0
Bock et al., 2006	4	Ns	Sg Coll	Sg	110	0	40	14	26	0

8. For this paper, presumed number of 'other' trials calculated from total trial number.

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Bock et al., 2004	2	Ns	Sg Coll	Sg	323	2	59	34	25	0
Bock et al., 2001	3	Ns	Sg Coll	Sg	143	1	48	13	35	0
Bock & Eberhard, 1993	4	Ns	Pl Coll	Sg	178	88	118	50	68	0
Bock et al., 2006	4	Ns	Pl Coll	Sg	43	12	20	5	15	0
Bock et al., 2004	2	Ns	Pl Coll	Sg	185	79	120	83	37	0
Bock et al., 2001	3	Ns	Pl Coll	Sg	105	15	72	31	41	0
Bock & Eberhard, 1993	1	Ns	Ns Irr	Sg	234	0	66	13	53	0
Bock & Eberhard, 1993	2	Ns	Ns Irr	Sg	68	0	28	7	21	0
Bock & Eberhard, 1993	3	Ns	Ns Irr	Sg	76	0	20	4	16	0
Barker et al., 2001	1	Ns	Ns	Sg	614	1	153	119	34	0
Barker et al., 2001	2	Ns	Ns	Sg	654	21	93	91	2	0
Bock & Cutting, 1992	1	Ns	Ns	Sg	392	4	244	51	193	0
Bock & Cutting, 1992	2	Ns	Ns	Sg	954	10	316	159	157	0
Bock & Cutting, 1992	3	Ns	Ns	Sg	833	7	312	177	135	0
Bock & Eberhard, 1993	1	Ns	Ns	Sg	235	0	65	19	46	0
Bock & Eberhard, 1993	2	Ns	Ns	Sg	59	0	37	16	21	0
Bock & Eberhard, 1993	3	Ns	Ns	Sg	73	1	22	6	16	0
Bock & Eberhard, 1993	4	Ns	Ns	Sg	291	2	91	28	63	0

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Bock & Miller, 1991[9]	1	Ns	Ns	Sg	198	2	120	.	.	.
Bock & Miller, 1991	2	Ns	Ns	Sg	381	10	121	.	.	.
Bock & Miller, 1991	3	Ns	Ns	Sg	338	18	156	.	.	.
Bock & Miller, 1991	2-r	Ns	Ns	Sg	364	13	135	.	.	.
Bock et al., 2006	1	Ns	Ns	Sg	824	23	305	305	0	0
Bock et al., 2006	2	Ns	Ns	Sg	219	2	355	54	301	0
Bock et al., 2006	3	Ns	Ns	Sg	556	6	158	55	103	0
Bock et al., 2006	4	Ns	Ns	Sg	65	0	10	1	9	0
Bock et al., 2012	1	Ns	Ns	Sg	352	0	160	38	122	0
Bock et al., 2012	2	Ns	Ns	Sg	479	12	21	21	NA	0
Bock et al., 2004	1	Ns	Ns	Sg	773	16	235	68	167	0
Bock et al., 2004	2	Ns	Ns	Sg	318	0	66	31	35	0
Bock et al., 2004	3	Ns	Ns	Sg	251	9	124	60	64	0
Bock et al., 2004	4	Ns	Ns	Sg	301	3	80	36	44	0
Bock et al., 2004	5	Ns	Ns	Sg	327	0	87	32	55	0
Bock et al., 2001	1	Ns	Ns	Sg	416	2	158	30	128	0
Bock et al., 2001	2	Ns	Ns	Sg	260	3	151	12	139	0
Bock et al., 2001	3	Ns	Ns	Sg	138	1	53	16	37	0
Bock et al., 1999	1	Ns	Ns	Sg	856	25	415	.	.	.
Brehm & Bock, 2013	1	Ns	Ns	Sg	1755	42	3	0	NA	3
Brehm & Bock, 2013	2	Ns	Ns	Sg	785	2	413	90	304	19
Eberhard, 1997	1	Ns	Ns	Sg	425	11	140	20	120	0

9. For this paper, number of correct trials estimated from graphs; presumed number of ‘other’ trials estimated from total trial number.

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Eberhard, 1997	3	Ns	Ns	Sg	214	1	73	6	67	0
Eberhard, 1999	1	Ns	Ns	Sg	52	0	20	5	15	0
Eberhard, 1999	2	Ns	Ns	Sg	99	0	61	9	52	0
Eberhard, 1999	3	Ns	Ns	Sg	93	1	66	16	50	0
Foote & Bock 2012[10]	1	Ns	Ns	Sg	410	0	6	.	.	.
Gillespie & Pearlmutter, 2013	1	Ns	Ns	Sg	874	3	563	177	377	9
Gillespie & Pearlmutter, 2013	2	Ns	Ns	Sg	1258	5	741	327	407	7
Solomon & Pearlmutter, 2004	1	Ns	Ns	Sg	153	1	80	16	33	31
Solomon & Pearlmutter, 2004	2	Ns	Ns	Sg	67	0	193	30	148	15
Solomon & Pearlmutter, 2004	3	Ns	Ns	Sg	58	0	242	19	177	46
Solomon & Pearlmutter, 2004	4	Ns	Ns	Sg	253	4	495	107	310	78
Solomon & Pearlmutter, 2004	5	Ns	Ns	Sg	532	7	523	150	158	215
Thornton & MacDonald, 2003	1	Ns	Ns	Sg	352	7	41	37	4	0
Thornton & MacDonald 2003[11]	2	Ns	Ns	Sg	586	1	55	35	20	0
Vigliocco & Nicol, 1998	1	Ns	Ns	Sg	240	1	47	47	0	0
Bock et al., 2012	2	Ns	Ns	Pl	480	6	26	26	NA	0
Bock et al., 2004	4	Ns	Ns	Pl	293	2	89	32	57	0

10. For this paper, number of correct trials estimated from proportion agreement errors out of valid trials and counts of agreement errors; presumed number of ‘other’ trials estimated from total trial number.

11. There is one extra trial reported in this cell.

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Bock et al., 2004	5	Ns	Ns	Pl	314	5	95	39	56	0
Brehm & Bock, 2013	1	Ns	Ns	Pl	1698	99	3	2	NA	1
Brehm & Bock, 2013	2	Ns	Ns	Pl	701	74	425	98	302	25
Eberhard, 1999	1	Ns	Ns	Pl	54	0	18	3	15	0
Eberhard, 1999	2	Ns	Ns	Pl	92	0	68	7	61	0
Eberhard, 1999	3	Ns	Ns	Pl	91	3	66	7	59	0
Solomon & Pearlmutter, 2004	1	Ns	Ns	Pl	123	0	111	41	31	39
Solomon & Pearlmutter, 2004	2	Ns	Ns	Pl	62	2	196	31	147	18
Solomon & Pearlmutter, 2004	3	Ns	Ns	Pl	31	1	268	28	173	67
Solomon & Pearlmutter, 2004	4	Ns	Ns	Pl	116	10	156	28	68	60
Bock & Eberhard, 1993	3	Ns	Np Irr	Sg	53	9	34	9	25	0
Bock et al., 2006	4	Ns	Np Irr	Sg	55	3	17	2	15	0
Bock et al., 2001	2	Ns	Np Irr	Sg	205	29	180	39	141	0
Barker et al., 2001	1	Ns	Np	Sg	516	61	191	163	28	0
Barker et al., 2001	2	Ns	Np	Sg	487	132	149	149	0	0
Bock & Cutting, 1992	1	Ns	Np	Sg	307	51	282	108	174	0
Bock & Cutting, 1992	2	Ns	Np	Sg	869	44	367	210	157	0
Bock & Cutting, 1992	3	Ns	Np	Sg	788	54	310	174	136	0
Bock & Eberhard, 1993	1	Ns	Np	Sg	198	31	71	26	45	0
Bock & Eberhard, 1993	2	Ns	Np	Sg	44	21	31	12	19	0
Bock & Eberhard, 1993	3	Ns	Np	Sg	57	8	31	15	16	0

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Bock & Eberhard, 1993	4	Ns	Np	Sg	208	67	109	39	70	0
Bock & Miller, 1991	1	Ns	Np	Sg	154	50	116	.	.	.
Bock & Miller, 1991	2	Ns	Np	Sg	295	29	188	.	.	.
Bock & Miller, 1991	3	Ns	Np	Sg	249	56	207	.	.	.
Bock & Miller, 1991	2-r	Ns	Np	Sg	320	25	167	.	.	.
Bock et al., 2006	1	Ns	Np	Sg	541	125	486	486	0	0
Bock et al., 2006	2	Ns	Np	Sg	173	34	369	110	259	0
Bock et al., 2006	3	Ns	Np	Sg	517	56	147	45	102	0
Bock et al., 2006	4	Ns	Np	Sg	58	4	13	4	9	0
Bock et al., 2012	1	Ns	Np	Sg	312	37	163	64	99	0
Bock et al., 2012	2	Ns	Np	Sg	438	45	29	29	NA	0
Bock et al., 2004	1	Ns	Np	Sg	508	104	412	291	121	0
Bock et al., 2004	2	Ns	Np	Sg	239	58	87	44	43	0
Bock et al., 2004	3	Ns	Np	Sg	148	42	194	128	66	0
Bock et al., 2004	4	Ns	Np	Sg	176	68	140	108	32	0
Bock et al., 2004	5	Ns	Np	Sg	265	24	125	62	63	0
Bock et al., 2001	1	Ns	Np	Sg	256	130	190	89	101	0
Bock et al., 2001	2	Ns	Np	Sg	216	56	142	28	114	0
Bock et al., 2001	3	Ns	Np	Sg	112	17	63	21	42	0
Bock et al., 1999	1	Ns	Np	Sg	524	98	674	.	.	.
Brehm & Bock, 2013	1	Ns	Np	Sg	1645	155	0	0	NA	0
Brehm & Bock, 2013	2	Ns	Np	Sg	756	77	367	86	257	24
Eberhard, 1997	1	Ns	Np	Sg	271	112	193	81	112	0
Eberhard, 1997	3	Ns	Np	Sg	275	124	177	56	121	0
Eberhard, 1999	1	Ns	Np	Sg	48	7	17	5	12	0

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Eberhard, 1999	2	Ns	Np	Sg	79	17	64	6	58	0
Eberhard, 1999	3	Ns	Np	Sg	80	17	63	16	47	0
Foote & Bock 2012	1	Ns	Np	Sg	199	6	3	.	.	.
Gillespie & Pearlmutter, 2013	1	Ns	Np	Sg	787	65	588	240	342	6
Gillespie & Pearlmutter, 2013	2	Ns	Np	Sg	1137	96	771	383	374	14
Solomon & Pearlmutter, 2004	1	Ns	Np	Sg	118	32	84	24	29	31
Solomon & Pearlmutter, 2004	2	Ns	Np	Sg	67	16	177	25	140	12
Solomon & Pearlmutter, 2004	3	Ns	Np	Sg	35	13	252	36	161	55
Solomon & Pearlmutter, 2004	4	Ns	Np	Sg	188	47	517	155	255	107
Solomon & Pearlmutter, 2004	5	Ns	Np	Sg	456	54	552	199	147	206
Thornton & MacDonald, 2003	1	Ns	Np	Sg	287	64	49	42	7	0
Thornton & MacDonald, 2003	2	Ns	Np	Sg	504	60	77	52	25	0
Vigliocco & Nicol, 1998	1	Ns	Np	Sg	196	36	56	56	0	0
Vigliocco et al., 1996	3	Ns	Np	Sg	270	36	126	33	93	0
Vigliocco et al., 1996	4	Ns	Np	Sg	199	25	64	64	0	0
Bock et al., 2012	2	Ns	Np	Pl	388	92	32	32	NA	0
Bock et al., 2004	4	Ns	Np	Pl	157	68	159	120	39	0
Bock et al., 2004	5	Ns	Np	Pl	191	95	128	81	47	0
Brehm & Bock, 2013	1	Ns	Np	Pl	1576	218	6	2	NA	4
Brehm & Bock, 2013	2	Ns	Np	Pl	687	107	406	116	266	24
Eberhard, 1999	1	Ns	Np	Pl	33	22	17	1	16	0

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Eberhard, 1999	2	Ns	Np	Pl	62	30	68	15	53	0
Eberhard, 1999	3	Ns	Np	Pl	71	11	78	23	55	0
Foote & Bock 2012	1	Ns	Np	Pl	149	55	4	.	.	.
Solomon & Pearlmutter, 2004	1	Ns	Np	Pl	115	9	110	35	32	43
Solomon & Pearlmutter, 2004	2	Ns	Np	Pl	44	5	211	36	146	29
Solomon & Pearlmutter, 2004	3	Ns	Np	Pl	29	5	266	39	164	63
Solomon & Pearlmutter, 2004	4	Ns	Np	Pl	125	22	135	27	52	56
Vigliocco et al., 1996	3	Ns	Np	Pl	292	34	106	28	78	0
Vigliocco et al., 1996	4	Ns	Np	Pl	193	20	75	75	0	0
Bock et al., 2004	5	Ns	NI	Sg	181	12	14	14	0	0
Bock et al., 2004	5	Ns	NI	Pl	130	38	39	39	0	0
Bock & Eberhard, 1993	3	Np	NI	Pl	5	69	22	4	18	0
Bock & Cutting, 1992	1	Np	Ns	Pl	11	357	272	95	177	0
Bock & Eberhard, 1993	3	Np	Ns	Pl	3	70	23	6	17	0
Bock & Miller, 1991	1	Np	Ns	Pl	7	189	124	.	.	.
Bock & Miller, 1991	2	Np	Ns	Pl	8	235	269	.	.	.
Bock & Miller, 1991	3	Np	Ns	Pl	36	300	176	.	.	.
Bock & Miller, 1991	2-r	Np	Ns	Pl	8	270	234	.	.	.
Bock et al., 2006	1	Np	Ns	Pl	29	771	352	352	0	0
Bock et al., 2006	2	Np	Ns	Pl	2	208	366	102	264	0
Bock et al., 2012	1	Np	Ns	Pl	16	318	178	82	96	0

Paper	*Expt*	*Head*	*Local*	*Notional Num*	*Singular verb*	*Plural verb*	*Other*	*[Pre*	*Uninfl*	*NR]*
Bock et al., 2012	2	Np	Ns	Pl	88	887	49	49	NA	0
Bock et al., 2004	1	Np	Ns	Pl	54	663	307	186	121	0
Bock et al., 1999	1	Np	Ns	Pl	21	790	485	.	.	.
Eberhard, 1997	2	Np	Ns	Pl	34	1573	481	151	330	0
Humphreys & Bock, 2005	1	Np	Ns	Pl	4	131	297	93	204	0
Thornton & MacDonald, 2003	1	Np	Ns	Pl	15	297	88	80	8	0
Vigliocco & Nicol, 1998	1	Np	Ns	Pl	15	217	56	54	0	2
Bock & Cutting, 1992	1	Np	Np	Pl	9	370	261	91	170	0
Bock & Miller, 1991	1	Np	Np	Pl	4	196	120	.	.	.
Bock & Miller, 1991	2	Np	Np	Pl	1	325	186	.	.	.
Bock & Miller, 1991	3	Np	Np	Pl	3	297	212	.	.	.
Bock et al., 2006	1	Np	Np	Pl	11	713	428	428	0	0
Bock et al., 2006	2	Np	Np	Pl	3	192	381	137	244	0
Bock et al., 2012	1	Np	Np	Pl	18	305	189	97	92	0
Bock et al., 2012	2	Np	Np	Pl	51	933	40	40	NA	0
Bock et al., 2004	1	Np	Np	Pl	39	691	294	187	107	0
Bock et al., 1999	1	Np	Np	Pl	8	708	580	.	.	.
Humphreys & Bock, 2005	1	Np	Np	Pl	1	292	571	159	412	0
Thornton & MacDonald, 2003	1	Np	Np	Pl	21	340	39	35	4	0
Vigliocco & Nicol, 1998	1	Np	Np	Pl	12	197	79	79	0	0

Appendix B

Calculation of PCFG grammar probabilities.

First, we estimated the frequencies of various NP structures in extant corpora:.

0.02 NP → [N [$_{PP}$ P N]] (Biber, Grieve, & Iberri-Shea, 2009, Figure 9.2)

0.005 NP → [N [$_{RC}$ N V]] (Roland, Dick & Elman, 2007, also Biber et al. Figure 9.2)

0.05 NP → [D N] (Google Books: https://tinyurl.com/DET-NOUN)

Renormalizing, this gives us the following probabilities for structures in the grammar (with the preposition and determiner elided):

0.27 NP → [N [$_{PP}$ N]]

0.06 NP → [N [$_{RC}$ N V]]

0.67 NP → [N]

Our grammar distinguishes complex NPs (two words) vs. simplex NPs (one word). Combining the first two NP expansions above (and retaining the third) gives the following probabilities:

0.33 NP → [NPC]

0.82 NPC → [N [$_{PP}$ P N]]

0.18 NPC → [N [$_{RC}$ N V]]

Estimates from COCA (Davies, 2008) suggests that ⅔ of English nouns are singular (i.e., the singular-to-plural ratio is 2:1). In simple noun phrases and relative clauses, assume that grammatical number follows this pattern.

0.44 S → [N_s V_s]

0.22 S → [N_p V_p]

0.22 S → [NPC_s V_s]

0.11 S → [NPC_s V_p]

0.66 RC → [N_s V_s]

0.33 RC → [N_p V_p]

Following Haskell et al. (2010), assume that 20% of the Ns Np items take plural agreement. Re-analyze these as plural headed phrases (NP_p). The overall probability of Ns Np is 0.059

(0.27 x 0.22). To reduce this by 20%, decrease its probability by 0.012 and increase the probability of NPCp by 0.012.

$0.208\ S \rightarrow [NPC_s\ V_s]$

$0.122\ S \rightarrow [NPC_p\ V_p]$

Add a new rule to accommodate plural agreement with overall probability of 0.012 (.12/.10).

$0.10\ NPC_p \rightarrow [N_s\ N_p]$

Combined, this leads to the following grammar:

$0.44\ S \rightarrow [N_s\ V_s]$

$0.22\ S \rightarrow [N_p\ V_p\ ]$

$0.208\ S \rightarrow [NPC_s\ V_s\ ]$

$0.122\ S \rightarrow [NPC_p\ V_p]$

$0.54\ NPCs \rightarrow [N_s\ N_s]$

$0.27\ NPCs \rightarrow [N_s\ N_p]$

$0.18\ NPCs \rightarrow [N_s\ RC]$

$0.47\ NPCp \rightarrow [N_p\ N_s]$

$0.24\ NPCp \rightarrow [N_p\ N_p]$

$0.10\ NPCp \rightarrow [N_s\ N_p]$

$0.18\ NPCp \rightarrow [N_p\ RC]$

$0.66\ RC \rightarrow [N_s\ V_s\ ]$

$0.33\ RC \rightarrow [N_p\ V_p]$